House of Commons

Children, Schools and Families Committee

Testing and Assessment

Third Report of Session 2007–08

Volume II

Oral and written evidence

Ordered by The House of Commons to be printed 7 May 2008

HC 169-II
Published on 13 May 2008
by authority of the House of Commons
London: The Stationery Office Limited
£0.00

The Children, Schools and Families Committee

The Children, Schools and Families Committee is appointed by the House of Commons to examine the expenditure, administration and policy of the Department for Children, Schools and Families and its associated public bodies.

Membership at time Report agreed

Mr Barry Sheerman MP (*Labour, Huddersfield*) (Chairman)
Adam Afriyie MP (*Conservative, Windsor*)
Annette Brooke MP (*Liberal Democrat, Mid Dorset & Poole North*)
Ms Dawn Butler MP (*Labour, Brent South*)
Mr Douglas Carswell MP (*Conservative, Harwich*)
Mr David Chaytor MP (*Labour, Bury North*)
Mr John Heppell MP (*Labour, Nottingham East*)
Mrs Sharon Hodgson MP (*Labour, Gateshead East & Washington West*)
Paul Holmes MP (*Liberal Democrat, Chesterfield*)
Fiona Mactaggart MP (*Labour, Slough*)
Mr Andrew Pelling MP (*Conservative, Croydon Central*)
Mr Andy Slaughter MP (*Labour, Ealing, Acton & Shepherd's Bush*)
Mr Graham Stuart MP (*Conservative, Beverley & Holderness*)
Lynda Waltho MP (Labour, Stourbridge)

Powers

The Committee is one of the departmental select committees, the powers of which are set out in House of Commons Standing Orders, principally in SO No 152. These are available on the Internet via www.parliament.uk.

Publications

The Reports and evidence of the Committee are published by The Stationery Office by Order of the House. All publications of the Committee (including press notices) are on the Internet at www.parliament.uk/csf/

Committee staff

The current staff of the Committee are David Lloyd (Clerk), Sarah Thatcher, (Second Clerk), Emma Wisby (Committee Specialist), Judith Boyce (Committee Specialist), Katie Phelan (Committee Assistant), Susan Ramsay (Committee Secretary), and John Kittle (Senior Office Clerk).

Contacts

All correspondence should be addressed to the Clerk of the Children, Schools and Families Committee, House of Commons, 7 Millbank, London SW1P 3JA. The telephone number for general enquiries is 020 7219 6181; the Committee's e-mail address is csfcom@parliament.uk

Witnesses

List of written evidence printed

List of unprinted written evidence

Written evidence has been published on our website at:
http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/memo/169/contents.htm

Written evidence has also been reported to the House. Copies have been placed in the House of Commons Library, where they may be inspected by Members. Other copies are in the Parliamentary Archives, and are available to the public for inspection. Requests for inspection should be addressed to The Parliamentary Archives, Houses of Parliament, London SW1A 0PW (tel. 020 7219 3074). Opening hours are from 9.30 am to 5.00 pm on Mondays to Fridays.

TA02 Heading for Inclusion, Alliance for Inclusive Education

TA05 Portway Infant School, Hampshire

TA06	Purbrook Junior School, Waterlooville
TA07	Epilepsy Action
TA08	Doug French, University of Hull
TA10	LexiaUK
TA14	Advisory Committee on Mathematics Education (ACME)
TA15	Mathematics in Engineering and Industry (MEI)
TA17	RNIB
TA19	Institute of Physics
TA20	Association for Achievement and Improvement through Assessment (AAIA)
TA21	Campaign for Science and Education (CaSE)
TA22	Dr A Gardiner, University of Birmingham
TA25	Barbara J Cook, Headteacher, Guillemont Junior School, Farnborough, Hampshire
TA26	Alliance for Inclusive Education
TA27	The Royal Society
TA28A	Wellcome Trust, Annex B
TA30	Ms Janet Graham, Director, Supporting Professionalism in Admissions Programme (SPA)
TA31	Qualification and Curriculum Authority, Annexes 3-5
TA33	Lorraine Smith, Headteacher, Western Church of England Primary School, Winchester, Hampshire
TA34	Association of Science Education (ASE)
TA35	The Mathematical Association
TA37	Richard Cooper
TA38	S Forrest, Teacher of Mathematics, Wokingham Authority
TA39	H Nickels, Headteacher, Silverton CE Primary School, Devon
TA42	SCORE—Science Community Partnership Supporting Education
TA45	University of Cambridge International Examinations
TA48	Warwick Mansell

Oral evidence

Taken before the Children, Schools and Families Committee

on Monday 10 December 2007

Members present:

Mr Barry Sheerman, in the Chair

Ms Dawn Butler	Mr Andy Slaughter
Mr David Chaytor	Lynda Waltho
Mrs Sharon Hodgson	Stephen Williams
Fiona Mactaggart	

Witnesses: **Professor Sir Michael Barber,** Expert Partner, Global Public Sector Practice, McKinsey and Company, and **Professor Peter Tymms,** Director of Curriculum, Evaluation and Management, School of Education, Durham University

Q1 Chairman: May I welcome Professor Sir Michael Barber and Professor Peter Tymms to the first evidence session of the new Committee? We have been busily building the team, seminaring and deciding our priorities for investigation, but this is our first proper session, so thank you very much for being able to appear before us at reasonably short notice. Both of you will know that our predecessor Committee started an inquiry into testing and assessment. It was a quite different Committee, but with its interest in schools, it decided to embark on a serious investigation into testing and assessment. It managed to tie up with a nice little bow almost every other area through 11 different reports in the previous Parliament, but it could not conclude this one. It troubled people to the extent that copious volumes of written evidence had come to the Committee, and it would seem wrong if we did not make such an important issue our first topic, pick up that written evidence, slightly modify and expand the terms of reference and get on with it. So, thank you very much for being here. You are key people in this inquiry: first, Michael, because of your association with testing and assessment, through which many of us have known you for a long time, right back to your National Union of Teachers days; and secondly, Professor Tymms, through your career in a number of institutions, where we have known you, and known and admired your work. We generally give witnesses a couple of minutes to make some introductory remarks. You know what you have been invited to talk about. If you would like to have a couple of minutes—not too long, although a couple of minutes is probably a bit short—to get us started, then I shall start the questioning. Peter, you were here first, so we shall take you first.

Professor Tymms: I am director of a centre at the University of Durham which monitors the progress of children in order to give schools—not anybody else—good information. It provides us with a tremendous database from which to view other issues, meaning that I have taken an interest in all the different assessments—key stage and so on. They have concluded that standards in reading have stayed constant for a long time, but that in mathematics, they have risen since about 1995. Those are the headlines on testing. On the introduction of new policies, I am keen to say—I might return to this—that there is a need for good trials. If we try something new, we should get it working before we move it out to the rest of the public. I am very keen for new ways of operating to be properly evaluated before they are rolled out, and then to be tracked effectively. We have been missing that.

Chairman: Thank you.

Sir Michael Barber: Thank you very much for your invitation, Chairman. I shall comment on the story of standards in primary schools, which I see in four phases. The first came between 1988 and 1996, when the then Conservative Government put in place the national curriculum, national assessment, Ofsted inspections, league tables and the devolution of resources to schools. There were lots of ups and downs in that story, but nevertheless that framework was established. Secondly, there was the phase with which I was associated—Government policy under David Blunkett who was the then Secretary of State for Education and Employment—during which there was a focus on what we called standards, rather than on structures. A big investment in teachers' skills, through the national literacy and numeracy strategies, led to rises in the national test results. I have always accepted that some of that was down to teaching to the tests, but a lot of it was down to real improvements evidenced by Ofsted data and international comparisons. In the third phase, between 2000 and 2005, the Government were focused largely on long-term, underpinning and structural reforms, including of the teaching profession, of secondary education and the introduction of the children's agenda, at which stage results plateaued. Things got harder, too, because we had picked the low-hanging fruit, as it were. I think that we should have stayed much more focused on literacy and numeracy, in addition to the others things that we did. That was my error. Now there is an opportunity to make real progress on literacy and numeracy as a result of the Rose review last year and the new emphasis on phonics. By the way, I

completely agree with Peter on the pilots and progression. If all those things are put together, I could envisage a fourth stage, during which we can begin to make progress. In summary, we have gone from being below average, on international comparisons, to above average—we are above France, Scotland and the EU average. However, we have a long way to go and significant improvements to make. If we want to be world class, we must do more.

Q2 Chairman: Thank you for those introductory remarks. I remember taking the Committee to New Zealand where people wanted to be able to assess more carefully the progress of students and were looking at what we had done. I recall their horror when it was suggested that they might adopt our system. They said, "We want to know how our young people are doing, but we do not want to go to the extent that you are of testing at so many ages." Are you sympathetic to that point of view? Do you think that we over-test?

Sir Michael Barber: Personally, I do not think that we over-test in primary schools—if that is what you are talking about. Primary school children take literacy and numeracy tests aged seven and externally-set and marked literacy, numeracy and science tests aged 11. That is a relatively small number of tests during a six-year primary school career. The information provided by the tests is fundamental to understanding how the system is working and to looking for strategies for future improvements. I do not think that we over-test at all.

Q3 Chairman: Even if that adds up to ages seven, 11, 14, 16, 17 and 18?

Sir Michael Barber: I focused my answer on primary schools. There is a separate debate to be had about secondary examinations and tests at ages 14, 16, 17 and 18. However, at primary level, we conduct the bare minimum of testing if we want to give parents, the system, schools and teachers the information that they need, at different levels, in order to drive through future improvements. One of the benefits of 10 years, or so, of national assessments is that this system has better information with which to make decisions than many others around the world.

Professor Tymms: I do not think that testing at seven and 11 is too much testing. However, if you have a system in which you take those tests, put them into league tables and send Ofsted inspectors in to hold people accountable, schools will test a lot more. So we probably do have too much testing in the top end of primary schools, but that is not statutory testing. It is the preparation for the statutory testing, so it is a consequence of what is happening. Of course, we do need the kind of information that those tests were designed to get at. You mentioned the need to know what our children are doing and their levels. If we wanted to know the reading standards of 11-year-olds in this country, we could probably find out by assessing 2,000 pupils picked at random. We do not have to assess 600,000 pupils. One purpose is to know what the levels are, which could be done with a sampling procedure, with the same tests every year, which would be secret and run by professionals going out and getting the data. There is another kind of information, for teachers about their pupils, which they could get by their own internal tests or other tests if they wanted, and another kind of information for parents. There is an interface: how do they get that information? Do they go to the schools, or do they read it in their newspapers? Do they know about their own pupils? Those layers of information, and how to get them, provide the complex background to the answer to your question. There is too much testing, but not because of a single test at 11—for goodness' sake, children can do that. I think that I was tested every two weeks when I was about eight years old, and I quite enjoyed them. Not all children do, but the possibility of that exists. We need good information in the system for parents, teachers and Parliament, and we need to know it nationally, but we do not necessarily have to do the sort of testing that we currently have to get that information. There are different purposes and reasons for doing it. I guess that I can expand on that as you need.

Q4 Chairman: But Michael is known to believe—I am not setting you against each other—in the notion that testing would drive up standards. It was the "engine", was it not? I am not misquoting you, am I?

Sir Michael Barber: It is not a misquote, but it is not a complete view of what I believe. I believe that, in order to drive up standards, we need a combination of challenge and support. Assessment and Ofsted inspection provide the challenge in the system, and then we need serious investment in teachers and their skills, pay and conditions. I am in favour of assessment, being able to benchmark schools and the information that that provides to heads, teachers and parents. I agree with Peter that there may in addition be an advantage to sampling techniques, probably linked with the international benchmarks to assess the performance of the whole system.

Q5 Chairman: I have slightly misquoted you: testing was "the engine to drive performance", I think you said.

Sir Michael Barber: But I am saying that the accountability system on its own is not enough. You need investment in teachers' skills, which is what the national literacy and numeracy strategies did. They gave teachers the skills and wherewithal to understand how to teach reading, writing and mathematics. The evidence of that is powerful. Only recently, the effective pre-school and primary education research programme, which Pam Sammons and others run, has shown clearly the benefits in student outcomes if teachers teach the last part of the literacy hour well—the plenary. Detailed pedagogical skills need to be developed by teachers, which needs an investment. Obviously, you also need to pay teachers well, ensure that the system is recruiting enough teachers and devolve money to the schools. I am strongly in favour of the challenge that comes from an accountability system, along with the wherewithal for heads and teachers to get the job done in schools—not one or the other, but both.

Q6 Chairman: Any comment on that, Peter?
Professor Tymms: There is an assumption here that standards have risen and that the national literacy strategy made a difference. In fact, over those years, reading hardly shifted at all. I perhaps need to back that up, because there are a lot of different sets of data. Somebody can claim one thing, somebody can claim another and so on. Is this an appropriate moment to go into that?
Chairman: Yes, indeed.
Professor Tymms: Okay. From 1995 to 2000, we saw a massive rise in the statutory test data at the end of primary school. They were below 50% and got up towards 80%. From about 2000 onwards, they were pretty flat. That looks like a massive rise in standards, and then it was too difficult because we had got to the top end, all our efforts had gone and so on. In fact, in 1998 or thereabouts, I was looking at our test data—we use the same test every year with the same groups of pupils—and did not see any shift in reading standards. The key stage assessments use a new test every year, and one must decide what mark corresponds to Level 4. That is harder. Test scores rose year on year as a percentage of Level 4 with a new test, but did not rise with a static test, and that raised a question. At the same time, Hawker was working at the Qualifications and Curriculum Authority, and said in *The Times Educational Supplement* that if results continued to rise, we would need an independent investigation. Around that time, QCA decided internally that it would investigate further. It commissioned Cambridge Assessment under Massey to take the tests from 1996 and 1999, and to go to a place that had not been practising the tests—Northern Ireland. It took equivalent samples of pupils and gave the 1996 and 1999 tests to them. If those tests were measuring a Level 4 of the same standard, the same proportion should have got Level 4, but they did not. Far more got Level 4 with the later test, so the standards were not equivalent, and that was fully supported in the Massey study. Massey did a follow-up study in which he compared the 2000 and 1996 tests, and found rises in maths, which were not as big as the tests suggested, but nevertheless were rises. He found that writing scores had increased, but called the rise in reading skills illusory. Additionally, several local education authorities collected independent data on reading, using the same test across the whole LA year after year, and there was practically no shift in reading scores, but there was a rise in maths scores. I was able to look at 11 separate studies, which all told the same story: over that period there was probably a slight to nothing rise—about one 10th of a standard deviation—which might have been achieved if children had practised tests, but there was no underlying rise. In maths, there was an underlying rise. There are two things going on. One is that children get better at tests if they practise them. Prior to national testing, they were doing practically no tests—it was necessary to go back to the time of the 11-plus for that. We saw a rise because of practising tests, and we saw an additional rise because standards were not being set correctly by the School Curriculum and Assessment Authority and then QCA between 1995 and 2000. Then there was teaching to the test. After 2000, QCA got its act together and set standards correctly. It now has a proper system in place, and standards are flat. There are small rises, and we must treat them with interest, but with a pinch of salt. Let us suppose that it is decided in committee that Level 4 is anything above 30 marks. If it were decided that it was one mark higher than that, the Level 4 percentage might go up by 2% or 3%, and that would make national headlines, but that would be due to errors of measurement. The discussion in the Committee is about three or four points around that point. The accuracy in one year, although there may be 600,000 pupils, is dependent on the cut mark, which is clear and was set incorrectly between 1995 and 2000. The assumption that standards were going up because we were introducing accountability, because we had testing, because we had Ofsted, and because we had the 500 initiatives that the Labour party put in place without evaluation shortly after coming to office, was based on a misjudgment about standards. Maths, yes; reading, no; writing, yes.
Sir Michael Barber: This is, as evidenced by Peter's comments, a complicated area, and I accept that completely. First, the national literacy and numeracy strategies are effectively a major investment in teachers' skills and their capacity to teach in classrooms. That is a long-term investment; it is not just about this year's, next year's or last year's test results. It is a long-term investment in the teaching profession's capacity, and it is well worth making because for decades before that primary school teachers were criticised for not teaching reading, writing and maths properly, but no one had invested in their skills and understanding of best practices. Secondly, there is a debate about extent, but we seem to be in agreement on maths and writing. When I was in the delivery unit after I left the Department for Education and Employment, I learned that it is dangerous to rely on one set of data. When looking at reading standards, it is right to look at several sets of data. One is the national curriculum test results, which tell an important story. Of course, there is an element of teaching to the test, but an element of teaching to a good test is not necessarily a bad thing, although overdoing it is. I always accepted that in debate with head teachers and teachers during that time. The second thing is that Ofsted records a very significant improvement in teachers' skills over that period of time. If teachers improve their skills in teaching reading, writing and mathematics, you would expect the results to go up. The third data set that I would put in that linked argument is that international comparisons—most importantly, the progress in international reading literacy study, or PIRLS[1]—showed that England in 2001 did very well up on international comparisons in reading. In 1999 came the first accusations that the test results were not real. Jim Rose led a review involving representatives of all the parties represented on this Committee, which found no evidence whatever of any tampering with the tests.

[1] Progress in International Reading Literacy Study

In addition, people in other countries have taken the kinds of things we did in that phase of the reform and replicated, adapted or built on them—Ontario being the best example—and they, too, have had improvements in reading, writing and maths. To summarise, although we might disagree about the extent of improvement, I think we agree that there has been significant improvement in maths and writing, which are very important. We are debating whether there has been improvement in reading. I think the combination of data sets that I have just set out suggests that there has been significant improvement in reading. I would be the first to say that it is not enough and that we have further to go in all three areas; nevertheless, we have made real progress. My final point is that over that period, there has, as far as I can make out, been no significant change in reading and writing in Scotland, where there was no literacy strategy. The results in international comparisons indicate that Scotland ticks along roughly at the same position.

Q7 Chairman: There has been a sharp drop in recent PIRLS. Does that mean we are going backwards?
Sir Michael Barber: Actually, I think it means that other countries have improved faster over that period. As I said in my opening statement, between 2001 and 2005, the Government were focused on some serious, long-term, underpinning reforms—most importantly, in my view, for the long run, solving the teacher recruitment shortage and bringing some very good new people into the teaching profession. That will have benefits for decades to come, but there was a loss of focus on literacy and numeracy at that point. Personally, I wish I had pressed harder on that at the time, but that is what you are seeing—the PIRLS data follows the same patterns as the national curriculum tests.

Q8 Chairman: I want to shift on because colleagues will get restless, but Peter was shaking his head, so I shall have to ask you to comment, Peter.
Professor Tymms: I must comment on several of those points. Take PIRLS, for starters, in 2001, and in 2006, when it apparently went back. Michael's comment was that we did not look good the second time because other countries went better than us. Certainly, some countries went better, but, in fact, PIRLS is standardised and uses Rasch models to get the same marks meaning the same thing, and our marks dropped back there. It was not just other people getting better; we actually got worse. But I want to persuade you that PIRLS in 2001 got it wrong and made us look better than we were and that the level has remained static. The reason for that is that for those international tests to work properly, the students who are tested must be a representative sample of the country. The PIRLS committee defines how to collect those pupils. We went out, in this country, to collect the pupils to do it and asked the schools to do the tests, but about half of the schools did not want to do it and refused to play ball. The second wave of schools were asked and only some of them complied, and then a third wave were asked. If you look at the 2001 PIRLS data, you will see two asterisks by England, because our sampling procedure was not right. If you are the head of a school and you are asked to do the tests, but your kids are not reading too well that year, you will say no, whereas if they are doing really well, you will say, "Oh yes, I'll go for it." So we had a bias in the data. We got people who really wanted to play ball, and it made us look better than we were. The next year, when schools were paid to do the tests—some held out and got quite a lot of money—we got a proper representative sample and found our proper place, which shows that our standards are just, sort of, in the middle for reading. The blip previously, which was crowed about a lot, was a mistake in the data.

Q9 Chairman: So, it was quite an awkward mistake in some ways, if it was a mistake. It is interesting that under PIRLS—we will shift on, before I get a rebellion here—most of the big countries like us, such as Germany and France, are about the same. Okay, Finland and some smaller countries such as Taiwan and Korea will always be high up there, but countries with big populations—in Europe, places such as France and Germany that are, in a sense, like Great Britain—are at around the same position.
Professor Tymms: I would point to a different pattern in the data which relates not to size but to the language that is chosen. Translating the results of reading tests in other languages is problematic to begin with. Can one say that reading levels are the same? You pay when you take your choice. But a long tail of underachievement in reading, will also be found in all the other countries where English is spoken. You will find it in Australia and even in Singapore, which is largely a Chinese population but reading in English, and in Canada and America. That is because English is a difficult language to learn to read, whereas Finnish is much more regular in the way that it is written on to the page. If you are going to be born dyslexic, do not be born in a country where people speak English, because it will really be a problem. Be born in another country such as Germany or Italy. I make that general point.
Sir Michael Barber: Peter has made an important point. I would like to add two other things. First, other European countries look at our reforms in education over the past 10 years and are impressed by them. I have had conversations with people from several of the countries that we have talked about, and on this set of PIRLS we were actually significantly above the EU average. We were above France and just behind Germany. The long tail of underachievement is a real issue. Personally, I think that the places to look for English-speaking populations that do really well on reading, writing and, indeed, generally are the Canadian provinces. Some of their practices are very impressive. That is one place I would urge you to look if you are thinking about the future.
Chairman: Thank you for those opening responses.

Q10 Fiona Mactaggart: You talk a lot about whether our assessment system accurately assesses standards over time, but that is only one purpose of assessment. I wonder whether our national

assessment system is fit for purpose as a tool for assessment for learning. I am concerned about the fact that we have examinations at seven. I am not sure that they help teachers as much as they should. Could you give your views on whether Standard Assessment Tests—SATs—in primary and secondary education help teachers use assessment for learning?

Professor Tymms: They were not designed to do that. A test taken at the end of primary school is clearly not meant to help children in primary schools because they are about to leave and go to secondary schools, which often ignore the information and do their own tests as soon as students come in because they do not believe what the primary schools say they have done. Unfortunately, that is the way of the world. It happens when children who have A-levels in mathematics go to university. They are immediately tested in mathematics. Even if you take pre-school, all the information passed from the pre-school to the reception teacher is often ignored, as the reception teacher does their own assessment. The tests are certainly not being used as assessment for learning, other than that the practice for the tests and other tests that might be used leading up to a test might be used in that way. They might be used as assessment for learning a little bit at age seven, but an infant school certainly would not use them in that way because it would be passing its kids on to the junior school. The tests are not intended to do that kind of thing, so they cannot be and are not used in that way. They are meant to hold schools to account and in order to produce information for parents. If we want assessment for learning, we must do something different. Many schools and teachers do that kind of thing off their own bat. There are other ways to assess. For example, there are diagnostic and confirmatory assessments. We could go into that kind of thing, but they are not assessments for learning.

Sir Michael Barber: You made an aside about tests or exams at seven. It is important for the system and, indeed, teachers in schools, to know early on whether children are learning to read and write and do mathematics, because if intervention is needed to support a child in getting on track with their cohort, the sooner you know that they have a problem, the easier it is to fix it. One purpose of national curriculum tests is to provide accountability and to provide information for parents, as Peter rightly said, and it is absolutely right that that should be the case. However, in addition to that, over a period of time the tests have taught teachers what the levels are. The basis of assessment for learning is for the teacher and, obviously, the student or pupil to be able to understand what level they are working at and what they need to do next to get to the next level. If it had not been for the national curriculum and the national tests, I doubt very much whether the quality of those conversations would be as good as they are. The key to assessment for learning is investment in teachers' skills to do that, so that they are constantly focused—not just individually, but in teams with their colleagues—on improving the quality of their teaching, working out what they must do to get the next child up to the next level and therefore constantly improving their pedagogy, which is the essence of the whole issue.

Q11 Fiona Mactaggart: The interesting thing is that your view, Peter, is that the real function of those tests is to hold schools to account, rather than as assessments for learning. I was speaking to a head teacher on Friday, who said to me, "Fiona, I just wish all primary schools were all through, because then we wouldn't have inflated test results for 7-year-olds coming out of infant schools." Her analysis was that in infant schools, for which Key Stage 1 SATs were summative results, there was a tendency towards grade inflation, which undermines your point, Michael. I agree that you need to know to intervene early, but if the accountability function militates against accuracy of assessment for learning, how do you square it?

Sir Michael Barber: First, the Key Stage 1 results are not under the same accountability pressures as those for Key Stages 2 or 4. Secondly, I would not have moved away from externally set and marked tests for Key Stage 1, because if you consider the evidence in the work of Pam Sammons and others, objective tests marked externally to the school are more likely than teacher-assessed tests in the school to provide a drive for equity. If that had been done, I doubt that the issue you just raised would have occurred.

Professor Tymms: The assessment for learning is really interesting. The evidence is that if we give back to pupils information on how to get better, but we do not give them grades, they are likely to get better. Putting in the grades, marks or levels and feeding back countermands—undermines—the feedback. That is very clear in the randomised trials and in the meta-analysis by Black and Wiliam in *Inside the Black Box*. The feedback to pupils on how to get better is vital, but it is undermined in other ways. The other point that Michael raised about identifying special needs early is also crucial. The key stage assessments will not identify special needs or identify them early; they are too late and not precise enough. If, for example, a child is likely to have trouble reading, they can exhibit it when they are 5 or 4-years-old through a phonological problem, which can be assessed diagnostically at an early stage. A child later on, who has, for example, a decoding or a word-recognition problem, or perhaps they can do both but they do not understand or make sense of the text despite being able to bark the words, can also be diagnosed. Diagnostic assessments can be put in place, but they are different from the summative assessments at the key stages. There are horses for courses, and we must be careful about how we aim to use them.

Q12 Fiona Mactaggart: So, if the assessments do not necessarily do what we want, how else could we assess the impact of national policies on schools? How can we test what the Government policies, national curriculum or improvements in teacher training do? How do we know?

Professor Tymms: We need a series of different systems; we should not have a one-size-fits-all test. We need an independent body, charged with monitoring standards over time, which would use a sampling procedure in the same way as the NAEP does in the United States, as the APU used to in England and as other governments do in their countries. The procedure would become impervious to small changes in the curriculum, because it would have a bank of data against which it would check issues over time, so that we might track them and receive regular information about a variety of them, including not only attainment but attitudes, aspirations, vocabulary and so on. I would ensure that teachers had available to them good diagnostic assessments of the type that I described. I would also ensure that there was a full understanding of assessment for learning among the pupils, and I would continue to have national tests at the age of 11, but I would not put the results in league tables. In fact, I would ensure that there were laws to prevent that sort of thing from happening.

Q13 Fiona Mactaggart: Would you have to keep them secret from parents?

Professor Tymms: No. Parents would be allowed to go to a school and ask for the results, but I would not make the results the subject of newspaper reports, with everyone looking at them in a sort of voyeuristic way. There are real problems with those tables, which are actually undermining the quality and the good impact that assessment data can have. We are forcing teachers to be unprofessional. League tables are an enemy of improvement in our educational system, but good data is not. We need good data. We need to know the standards and variations across time, but we do not need a voyeuristic way of operating and pressure that makes teachers behave unprofessionally.

Sir Michael Barber: At the risk of ruining Peter's reputation, I agree with a lot of that, and I want to say a few things about it. First, as I understand it, a new regulator is due to be set up. An announcement was made a couple of months ago by Ed Balls: I am not sure where that has got to, but the announcement was made in precise response to the issues that Peter has raised. Personally, I have no doubt about the professionalism of the QCA in the past decade. It has done a good job, but it is important that standards are not just maintained but seen to be maintained. The new regulator will help with that once it is up and running. Secondly, on monitoring standards over time, as I said earlier, particularly now that international benchmarking has become so important not just here but around the world, I would like the regulator to use samples connected with those benchmarks and help to solve the problems of getting schools to participate in samples, which Peter mentioned. That would be extremely helpful. I agree completely with Peter about investing in teachers' skills and giving them the diagnostic skills to make them expert in assessment for learning. When I debate the programme for international student assessment results with Andreas Schleicher, who runs PISA—he is an outstanding person and it may be worth your interviewing him—he says that virtually no country in the world implements more of the policies that would be expected to work according to the PISA data than England, but that that has not yet translated into consistent quality, classroom by classroom. That is the big challenge, and what Peter recommended would help to achieve it. Like Peter, I would keep tests at 11. On league tables, the issue—and I have this debate with head teachers a lot—is that unless a law is passed, which I do not see as terribly likely, there are only two options for the schools system. One is that the Government, in consultation with stakeholders, designs and publishes league tables. The other is that one of the newspapers does it for them. That is what happened in Holland and it is happening, too, in Toronto and in Finland. It happens with universities. If you talk to university vice-chancellors, you find that they are in despair because various newspapers and organisations are publishing league tables of university performance over which they have no leverage. The data will be out there—this is an era of freedom of information, so there is a choice between the Government doing it or somebody else doing it for them. If I were a head teacher, I would rather have the Government do it—at least you can have a debate with them—than have the *Daily Mail* or another newspaper publish my league tables for me.

Professor Tymms: Can I pick up on that? I wish to make two points about league tables. First, we publish the percentage of children who attain a Level 4 and above, so if a school wants to go up the league tables it puts its effort into the pupils who might just get a Level 4 or a Level 3. It puts its efforts into the borderline pupils, and it does not worry about the child who may go to Cambridge one day and has been reading for years, or the child with special needs who is nowhere near Level 4. That is not going to show up on the indicator, so we are using a corrupting indicator in our league tables. Secondly, if you look at the positions of primary and secondary schools in the league tables, you will find that secondary schools are pretty solid in their positions year on year, but primary schools jump up and down. That is not because of varying teachers but because of varying statistics. If a school has only 11 pupils and one gets a Level 4 instead of a Level 3, the school is suddenly up by almost 10% and jumps massively. There is a massive fluctuation, because we produce league tables for tiny numbers of pupils. We can include only children who are there from Key Stage 1 to Key Stage 2 on the value added, which often means there is turbulence in a school. We should not publish for tiny numbers. The Royal Statistical Society recommends always quoting a measure of uncertainty for error, which is never done in those tables. We have 20,000 primary schools, and if the Government did not produce tables that the newspapers could just pick up and put in, it would require a pretty hard-working journalist to persuade them to give the press their data. It would be possible to make laws saying that you cannot publish tables. Parliament makes laws saying

that you should not have your expenses scrutinised, so why can we not produce a law that says that schools' results should not be scrutinised?

Q14 Mr Slaughter: You said a few moments ago, Sir Michael, that one of the purposes of national testing at seven and 11 was to identify children who are in difficulties. That sounds counter-intuitive. Would you not expect teachers to know that anyway? If testing has a role, is it not in assessing the needs of individual children, just as testing is used, for example, to assess the needs of people with a hearing problem? Otherwise, it is likely to lead to buck passing? If we test everybody, it almost becomes the responsibility of the state or someone else to ensure that everyone reaches a higher level. Given the length of time that we have had testing, how far has that become true? Stories in newspapers report the reverse, and say that a substantial minority of children still move onto secondary school without those skills.

Sir Michael Barber: I am not arguing that national curriculum tests alone will solve every child's problems. I agree strongly with what Peter said about teachers developing the diagnostic skills to diagnose such things. We want all teachers—I shall focus on primary schools—to be able to teach reading, writing, mathematics, and some other things, well, and then develop over time the skills needed to deal with individuals who fall behind. It is very good to see Government initiatives, such as the Every Child a Reader initiative, that pick up children who fall behind. I am in favour of all that. You need good diagnosis, which incidentally is one of the features of the Finnish education system that makes it so good—they diagnose these things early. The national curriculum tests have spread understanding among teachers of what the levels are and of what being good at reading, writing and mathematics looks like. They also enable the system to identify that among not just individual students, but among groups of students who have fallen behind. The system has great data about particular groups of students or schools that are falling behind, which enables it to make informed decisions about where to target efforts. My point is not just about individual students, therefore, but about groups of students or variations within the cohort. I shall comment on the point about league tables. In the end, the data will out—this is an era of freedom of information. We can have a perfectly valid debate about whether Level 4 is the right indicator. However, the percentage achieving Level 5 went up very rapidly during the early phase of the national literacy strategy, which suggests that good teaching is good teaching is good teaching. That was a result of the combination of the accountability system and the big investment in teachers' skills.

Q15 Lynda Waltho: In evidence so far, we have heard that the testing regime serves a large number of purposes—specifically, end of key stage, school accountability, assuring standards over time and assessment for learning. I am getting the feeling that there is not a lot of confidence that at least two of those are being achieved. What about the others? Can the system fulfil any of those purposes? Is it working? Is it fit for purpose? I do not have the impression that it is. As a former teacher and a parent, I found the regime useful in all of those areas at some point, but what is your assessment of its capabilities across that range?

Professor Tymms: I do not think that it is being used at all for assessment for learning. And I do not think that it can be, except where it is used incidentally. It provides a level against which teachers can set their pupils. If a teacher in a high-achieving school could judge her pupils, she would probably underestimate them because she would base her judgment on those she knows. The reverse would probably happen in a low-achieving school. Standardised levels for national tests give the firm ground on which a teacher can make a judgment. That is a good thing. It is there and it is being used. It gets information to parents, but it has its downsides. I do not think that testing is good at monitoring standards over time. We are saying, "Take this test, and we will hold you to account for the results and put them in league tables. We will send in an Ofsted inspector and ask you to assess your pupils and send us the results". That is an inherently problematic system. It is a little difficult. Another inherently problematic thing is having qualifications and curriculum in the same body—the QCA. Somebody should design the curriculum and somebody should assess it, but they should be separate bodies. That is an unhealthy way to operate a system. If we want to know what standards are over time, we are far better off with an independent body. If we change the curriculum—we read in *The Times* that that will happen, and we hear it regularly—and introduce an oral test, suddenly Level 4 will not mean the same thing, because a different curriculum will be assessed. We cannot monitor standards over time, but by having an independent body charged with monitoring standards not just against the national curriculum but against an international concept of mathematics or reading, we can track things over time. We must do different things. I come back to the need to understand the special needs of the child and pick out the child who already has a serious problem. Teachers can assess their children pretty well, but they cannot be expert in all the special needs—varieties of dyslexia, dyscalculia, attention-deficit hyperactivity disorder and so on—nor should they be expected to be. However, they might spot a problem with a child who needs to be assessed in different ways, so tools to help the teacher help the child and identify special needs and things falling back or not going quite right to begin with would make sense. Computerised diagnostic assessments with bespoke tests in which the child uses headphones to listen to the computer and is asked questions according to how they respond is to be the way of the future, but it cannot be the way of the future for statutory assessments, which require a new test every year to maintain security.

Q16 Lynda Waltho: There would be more tests then.

Professor Tymms: Different types, and probably less testing. We have more testing if we have league tables. It is the league tables that are our enemy.
Sir Michael Barber: I think that, on the whole, the national curriculum tests are beneficial. I have a lot of confidence in them, and I am always cautious in advising anybody or any education system to move too rapidly in changing assessment or qualifications, as that involves a lot of risk. Nevertheless, one should not stick with things for all time. I think that they have been good tests and that they have been good for accountability purposes. Along with the supports that I mentioned earlier, they have helped to drive improvement in the system. I agree with Peter about the need for an independent body to monitor standards over time—that is absolutely right. The proposal that is currently being piloted in 400 or 500 schools—progression pilots in which children are tested when they are ready for level tests—is very promising, but it is all in the detail. If that works, it could be beneficial in making sure that children at all stages and ages are making progress. The data show that, at present, there is a bit of drop-off in progress for years 3 and 4, but we would be able to move away from that if we had testing-when-ready tests. There is a lot of promise in them, but, as with any shift in the testing and assessment system, it is all about getting the detail right.

Q17 Chairman: We can come back to your last point. You mentioned a comment by Professor Schleicher.
Sir Michael Barber: I do not think that Andreas Schleicher is a professor, but he would be a very worthy one.

Q18 Chairman: Can you guide us to what you were quoting from?
Sir Michael Barber: I was quoting from a conversation with him. Before using his comments in the Committee, I checked that he was happy to be quoted on the record. You can put the quote on the record. He is quite happy to be quoted along the lines that I gave.

Q19 Lynda Waltho: You both discussed whether league tables were an enemy or a friend. It seems that you have completely different ideas. I agree with you, Sir Michael. I think that it is likely that the newspapers will develop their own league tables. If they do league tables about what we spend on our breakfast at the House of Commons, they will do league tables for school results, believe me. Would it not be better if the Government set out explicitly the full range of purposes for league tables; in effect, if they explained the results better? Would that make a difference, or am I just being a bit naive?
Professor Tymms: It would be interesting to try, but I do not know. If I buy something, I never bother reading the instructions until I get stuck. I would guess that most people would just look down the league tables and read the small print and headlines to find out who is at the top and who is at the bottom. When the league tables come out every year, the major headlines that we see are whether boys have done better than girls, or *vice versa*, or that one type of school has come top. It is the same old thing time and again, despite great efforts to steer journalists in a different direction. I despair of league tables, but it would certainly be worth trying providing more information. I think that the Royal Statistical Society's recommendation not to give out numbers unless we include the uncertainties around them is a very proper thing to do, but it is probably a bit late. The cat is out of the bag, and people are looking at the league tables. Even if there is more information, people will concentrate on the headline figures.
Sir Michael Barber: You can always look at how you can improve a data system like that and explain it better. I agree about that. I have been a strong advocate of league tables—and not only in relation to schools—because they put issues out in public and force the system to address those problems. League tables, not just in education, have had that benefit. Going back some time, I remember lots of conversations with people running local education authorities. They would know that a school was poor, and it would drift along being poor. That was known behind closed doors, but nothing was done about it. Once you put the data out in public, you have to focus the system on solving those problems. One reason why we have made real progress as a system, in the past 10 to 15 years, in dealing with school failure—going back well before 1997—is that data are out in the open. That forces the system to address those problems.
Professor Tymms: Why has it not got better then?
Sir Michael Barber: It has got significantly better. We have far fewer seriously underperforming schools than we had before.
Chairman: We do not usually allow one witness to question another, but never mind. You can bat it back.
Sir Michael Barber: It was a fair question.

Q20 Mr Chaytor: Looking at tables and accountability, may I ask you a question, Michael? In response to a remark from Peter, you said that it is important not to rely on a single data set, but is not that exactly the flaw of our system of league tables? Whatever the level, whether in primary or secondary school, the headline is the single data set. Is there any other public institution or system of accountability for public services in Britain that relies on a single data set, other than that which we have in schools? Do we use a single data set for hospitals, police authorities or primary care trusts?
Sir Michael Barber: My remark about not relying on a single data set was in reference to measuring progress over time. That is why I referred to several sets when we debated what had happened to literacy in the past decade or more. That is what I meant. You would triangulate the data sets. I think that league tables based on national tests are perfectly respectable and fit for that purpose. As I said in answer to Lynda Waltho, it is not the case that you cannot improve them; you can have a debate about how to improve them. In the schools system, we do not rely purely on tests and league tables to assess the

quality of schools. We also have Ofsted inspection, which considers the leadership and management of schools, the ethos within them and the quality of teaching as well as the standards that are achieved. That is important because it creates a more rounded picture of what schools are for.

Q21 Mr Chaytor: But in terms of accountability to parents, which is the most significant—the 5 A to Cs score, the percentage at Level 4 score or the Ofsted report? The report is a broader document, but it is also dominated by results—perhaps increasingly?
Sir Michael Barber: It takes account of results, but it does not add anything new to them. However, it looks at what is going on inside the school that delivers those results. Some of the things that I mentioned, such as quality of leadership and management are lead indicators of what will happen to results. With stronger leadership and better-quality teaching, in time the results will improve. I strongly support Ofsted inspection for that reason. There are things that you can do to improve it all the time. That is part of the task of the new chief inspector, whom I understand you will interview soon. You can debate that with her. As I understand it—and you will know from your constituents—parents consider performance in published test results, but they also examine Ofsted reports and take great interest in them when they come round. Of course, they appear only once every three years as opposed to every year.

Q22 Mr Chaytor: May I ask both of you, but perhaps Peter first, what is the relationship between the single data set of test results and pupil intake? We can all agree that the quality of teaching is essential to improvement, but is there received wisdom that such-and-such a percentage of the outcome is determined by the input?
Professor Tymms: A league table position is largely determined by the intake of pupils to that school. It might vary depending on how you analyse it, but if you had measures of pupils on intake, that would certainly explain more than 50% of the variants in the results, and maybe up to 70% The amount that is due to the quality of teaching is typically quoted as being about 10 to 15% of the variants in secondary schools, after intake is taken into account, which means that we are down to about 5 to 7% of the variation in the league tables being due to the quality of the school—maybe less, once everything is taken into account. In primary schools it is slightly more, but it is still dominated by the intake.
What we see in the league table is dominated by the intake, so we talk about a school at the bottom end of the league, but if we put all the schools in the table, a lot of schools at the bottom would be special schools, as they have children with severe learning problems. We need to know what the intake is and the progress made, and therefore the value added, in order to make sense of the figures. A lot of mistakes were made through judgments that schools at the bottom of league tables were bad, because that was not taken into account. It is quite difficult to take that into account, but we are moving forward. That is why the earlier measures are so important. Of course, once there is teacher judgment, you can no longer rely on outcome measures, as they are not objective tests and teachers might do things to boost their positions. The data become suspect.

Q23 Mr Chaytor: Would you accept that figure of 50 to 70%?
Sir Michael Barber: It varies from one system to another, but home background is clearly a major influence on outcomes. Nobody is debating that. We recently published a report having examined some of the best-performing systems in the world, which get much higher consistency in the quality of teaching and therefore the quality of outcomes than ours. They seem to be better at overcoming the disadvantage that children bring into a school. It is important stuff—what do those systems do? I am summarising a substantial report, but first, they select great people into teaching. Even in the 21st century, when young people have many options, they are still getting great people into teaching. We have done reasonably well on that in the past decade, but nobody can be complacent. Secondly, they train them really well, focusing on the quality of classroom teaching. Thirdly, they do the sort of things that Peter and I have been talking about—they ensure that the processes in the schools, assessment for learning and others, mean that each teacher constantly improves their skills and their ability to deliver great lessons for their students. Fourthly, they have systems that do not write off any student, as we were talking about earlier. They care, they spot early when children are falling behind and they pick them up and catch them up.
We could do all that. If we did—some schools do it brilliantly—we would reduce the impact of home background on the outcomes that students achieve. That is what we must do, and publishing the data puts that issue on the agenda in a way that nothing else would.

Q24 Mr Chaytor: If there is a general consensus that the relationship between home background and pupil intake is the dominant explanation of a score in the league table, is there not a dynamic built into the system that there will always be failing schools? From day one of the league tables, a certain number of schools were at the bottom of the pile. The existence of the league table reinforces the sense of failure in those schools and there is almost a spiral of decline. Is that not an inevitable consequence of a league table system based on a single data set?
Professor Tymms: Yes, I think that you are quite right. For example, you will find that fewer people apply for headships in schools at the bottom of the league table. Such schools have great difficulty appointing heads—they might have to appoint ordinary teachers—whereas there are enormous numbers of applications to schools at the top of the league table. Those schools have the pick of the bunch which provides a positive reinforcement, while others get worse and worse. It is the Matthew effect in operation—"For whosoever hath, to him shall be given". That is a real concern. On the

international differences between schools, it is right to say that some countries have enormous variations between schools and that others have very little variation. In our country, there is a large variation—we have private schools and some very tough schools. However, if you go to the United States or to China—bizarrely—you will find much greater variations, largely because their schools are funded by local taxes, which means that if you live in a poor area, you have a poor school and poorly-paid teachers. We have that a bit in this country owing to the private system. A nice league table came out in the *Educational Researcher* looking at qualifications of teachers in schools according to affluence and deprivation. In this country, you will typically find that the more affluent the school, the higher the qualifications and greater the experience of the teachers. That trend is much more dramatic in some countries, but in others it is actually reversed—they put their apparently better teachers into tougher schools in order to reverse that situation. We do not do that kind of thing here; we do not even think that that is possible. We have a serious discrepancy, however, between those at the top and those at the bottom. We know about that on an individual pupil basis, but it is on a school basis as well, which is reflected in the league tables.

Sir Michael Barber: I agree with what Peter said about the US. You might suppose that schools would enter a spiral of decline, but that is not what happens or what the data show. The number of schools achieving less than 30% five As to Cs has dropped dramatically from more than 600 to about 50—I cannot remember the data exactly, but they are available. By putting the data in the open, resources have been targeted to those schools, so programmes such as the Excellence in Cities programme, have helped struggling schools to improve. We have seen bigger improvements in some of those areas than in other parts of the country. You could reinforce that further. I am interested in what they have done in New York city recently with their new accountabilities system, under which a school gets double value for moving forward a student in the bottom third of the performance distribution. You could provide greater incentives to moving forward students in the bottom third. Programmes such as the Teach First initiative and the Excellence in Cities programme have got good teachers and head teachers into disadvantaged schools. One of the reasons for that has been the fact that the data are out in the open.

Professor Tymms: I cannot let that go. The advice that we are hearing on payment by results is so misguided. If teachers can get more money for their schools according to the number of pupils, we have a problem. We have a system in which teachers have been paid according to their pupils' progress. That is an unhealthy system to advocate. That system advocates schools and gives them more money because they push more pupils forward, but they are the ones producing those results. Again, you strain professionality by going down that route.

Sir Michael Barber: May I correct that? With the allocation of resources, you need to do that in order to bring equity. I am not advocating anything other than that. The Excellence in Cities programme gives money to schools and areas because they suffer from disadvantages compared with the average. The resources are to bring greater equity. I am not sure what Peter was commenting on, but I was not making the point that he disagreed with.

Q25 Chairman: Peter, would you not want to reward specialist teachers, even if they are charged and do better with the most difficult students?

Professor Tymms: It is a very difficult problem. It would be attractive to say that people doing better should be paid more and promoted. However, schools have promotion systems already that reward those teachers. We should not pay them according to their year's results or tell them, "If your pupils get Level 4s we will give you more money." They are the very teachers invigilating those pupils. They are the ones opening those papers and giving out the results. Making that direct link would strain professionality too much. Furthermore, we are talking about one or two pupils getting an extra result in one year compared with the previous year. That is too close to the bone. It is not the way to go. We need to distance ourselves from that direct link with pupils' marks on papers and from rewarding head teachers for moving up the league tables. Let us consider the percentage of five As to Cs in secondary schools. Of course, many more students have achieved that and many more schools do that, but students are just entered for a few more tests. That is largely what happened, and largely what caused the improvement. The underlying quality of the improvement is not there to be shown. Many students who would not previously have been entered for GCSEs now are, but that does not mean that standards have changed. We must be careful how we define schools that are doing badly and those that are doing well.

Q26 Ms Butler: On that point, do you think that the contextual value added data play a role in how we weight pupils who have done better after coming in at the lower end of the spectrum?

Sir Michael Barber: I think that contextual value added data is important, because it helps us to understand the system in a way that cannot be done without it, so I am strongly in favour of it. The quality of the data in our system is now better than it has ever been, and compares very well internationally. The ability to do value added analysis on individual pupil level data, which we now have in the national system, is a huge benefit. We need contextual value added data as well as raw data, because when students reach the age of 16, they may go into the labour market with everyone else, so it is not enough to take account just of value added. People need to reach a basic standard that gives them access, hopefully, to higher education or to work. I am in favour of the raw results being used and thought about to drive action, but I am also in favour of contextual value added data being available so that we can understand what impact

policies and schools are having on the system. It is helpful to understand the system, but it is not enough on its own to drive equity in outcomes.

Professor Tymms: Yes, value added is vital and helps us to understand, but the way in which it is calculated is important. Contextual value added is one way of calculating it, but we must be careful. For example, when looking at the progress made by children in maths and reading at Key Stage 1 to Key Stage 2, and value added, we ask what children normally get given those Level 1 results, and what did they get at Level 2? If they did better than most children with the same starting point, that is essentially the value added, but in a broader value added system, we might take account of children's home background, ethnicity, age and so on. There we must be careful. For example, in the system children from a poor background do not do well, so if such children fall by the wayside and do less well on average when progressing from Key Stage 1 to Key Stage 2, our value added system, which takes that into account, assumes that that is all right. In fact, it may be the system that is making them fall by the wayside, because we are excusing bad performance. Contextual value added, which tries to take everything into account, brushes that under the carpet, and we must expose it and see what is happening. There are different ways of looking at value added, and in Durham we always give schools different ways of looking at that, so that they can see it is one way or another. That is important. In the United States, a couple of great researchers, Doug Willms and Steve Raudenbush, talk about two types of value added: type A and type B. Parents want to know how their child will progress at a school. They want to know pupils' scores at the beginning and later, so that they know what is likely to happen in that school. That is type A value added. An administrator might ask how well the school is doing, given its circumstances. We know that pupils progress less well in schools in tough areas, so various schools should be looked at to see how well they are doing. Those are different types of value added. A system that says there is one type of value added—core contextual value added—is misleading, because we need much more information. We can get that information, and it can improve the system. Good information for parents, for administrators and for the country is vital.

Sir Michael Barber: For the record, I agree totally. That is one reason why national curriculum assessment for all students is an important part of being able to generate such data.

Q27 Mr Chaytor: May I pursue one more issue? On the choice and setting of targets at Key Stage 2, Level 4 is seen as the point below which children have failed. However, am I not right in thinking that when the key stage system was established in 1988, Level 4 was chosen as the average level of performance? My question is twofold. First, will there come a point at which the failure threshold will have to move up to Level 5? Secondly, what does the research suggest about the impact on children's enjoyment of learning and on their motivation when they start their secondary school career knowing that they have failed and that they have been labelled by the local newspaper as having failed? What is the link between targets and enjoyment and motivation?

Professor Tymms: They are really good questions, so I shall do my best to answer them. First, on the targets, of course we have had a shift in standards so that Level 4 is not the Level 4 with which we started. That does not make too much sense. Further, we should think about targets in terms of the value-added approach: you see where the children were and where they are likely to go and not in terms that Level 4 is good and below Level 4 is bad. For some pupils, Level 3 is a great result and a real success; for others, Level 4 is a dreadful fallback from where they were. So, when thinking about where we expect to go, we must think in those terms—about progress, rather than about absolute levels. A teacher or a school should be held to account only for the progress that their children make, not for the level that they attain. We must keep that in mind. The targets that are imposed are not the best ones; we should use targets that come from within. In the research into targets and whether if I set myself a target I do better, it is clear that targets really work on relatively simple tasks—such as chopping down trees and washing dishes. On complex targets, such as teaching and running a school, targets do not work, and that is where ownership comes in. We have got ourselves in a bit of a tizz over the targets. The research into fear of failure and so on is a complicated area. It is clear that young children, as they go through life, are predestined to fail in some things and succeed in others. In a sense, they expect that to happen and then to "Try harder and I'll do better." They are resilient in terms of a little failure and a little success. However, we do not want to slap down children who have done remarkably well to get to a Level 3 from where they started. It is an error to label them as failures, and it is also problematic to label their school as a failure, because they feel that in themselves. I have not seen research into the exact issue that you described, but I reviewed research into the feelings of children towards reading over the years. In our data, we saw that they stayed fairly constant over time, but other data suggest that children are less positive towards books than they used to be. We know that when they get older, they get less positive, which is a feature of education in general, and we know that boys more than girls become less positive as they get older, so by the time primary school finishes, there is a set of disaffected boys moving on to secondary school. They do not like school. If asked "Do you like school?", they say no. "Do you look forward to school?" "No." "Do you like your teachers?" "No". They then go on to a secondary school that has to start with the kids from where they are, and that is a pretty tough job. We must worry about these things, and any national monitoring system should examine attitudes, self-esteem, welfare and physical growth—all the issues coming out of *Every Child Matters.* We do not have that yet.

Q28 Chairman: May I take you back to the first part of David's question and to the question before that? We pushed you on why you are so resistant to

payments by results—for getting good achievement out of young people who are less easy to teach. We have had a system for years whereby, as I understand it, if you were the high mistress of St. Paul's in the City or of King Edward's boys or girls school, you had a wonderful group of highly motivated kids who had passed all sorts of examinations to get in. If you did not get wonderful results out of them, serious questions would be asked. The people teaching such groups have always received the best pay, but you are making anti-Freud—I mean David Freud—points. You would not incentivise somebody who did a really good job of taking the most difficult youngsters and bringing them up further than you would expect. Why are you so resistant to that?

Professor Tymms: I am resistant to the direct link between the marks of those kids and the pay of their teachers. I am not against reward, and I am not against paying teachers for good results and I am not against getting good teachers in and rewarding them or paying teachers more if they are working in tough circumstances and doing a good job. But a broader decision needs to be made by the head, or perhaps by others, to say, "This teacher is doing well and is worthy of good pay." It is the direct link to the marks that I worry about. That is where the devil lies.

Sir Michael Barber: I shall come to David's question. However, I think that, within the framework set for national pay and conditions, head teachers should make the decisions about who to reward. I think that for the system to do that from outside for individual teachers is complicated and likely to be damaging. However—I think I am agreeing with Peter here—whole-school rewards for making real progress, particularly in disadvantaged areas, would be wholly positive. Obviously, you have to get the detail right of how that works. On David's question, I agree with Peter that the system should get into measuring some of these wider outcomes, including enjoyment, motivation, and so on. I think that that is something that Ofsted inspection could do better in future. Ofsted inspection has been beneficial, but you could do more of that and use it to get into some of those issues, as indeed some systems are now thinking about—for example, in Victoria, Australia. I have written a book about Government targets called, *Instruction to Deliver*. You could look at the arguments for and against and the mistakes that were made, but you could also look at the benefits from really good targets that focus on the essence of the business. So I will not go into that. A good target can inject real ambition into a system. However, I should really like to address the Level 4 question. When I look at the 21st century, I see a labour market that is going to demand very high skills, not just in terms of reading, writing and mathematics, but in respect of rounded human beings able to work in teams and so on. I see a very demanding labour market for the young people coming through. The rest of their lives, too, will be very demanding: there are a lot of challenges in the 21st century. It is absolutely right that we are demanding more of our system than when the levels in the national curriculum were founded in 1988. Level 4 was chosen for the end of primary school because it is for reading and writing well, not just for basic reading and writing. A child who gets Level 3 can read perfectly well if you put a book in front of them, but reading and writing well is what gives you access to the secondary curriculum and that is what we have got to keep focused on. Sometimes I have the feeling that people believe—I know that some teachers and heads feel like this, because we have had this debate—that the Government imposed all these targets. However, the truth is that the targets, in effect, are the demands placed by the 21st century: the Government are a mediator of those and sometimes they get it right and sometimes they get it wrong. But we would be betraying our young people if we did not set out for them the demands of the future that they are going into. Therefore, we should be trying to get a school system that can match up to and meet those standards.

Q29 Mr Chaytor: Looking at Key Stage 4, is Warwick Mansell, in his book on testing and assessment, right to be scandalised by the extent of teacher intervention in the production of GCSE coursework?

Professor Tymms: I do not know enough about this.

Sir Michael Barber: I have not read Warwick Mansell's book.

Chairman: We always like it when witnesses say, "I don't know." It is the people who give us an opinion on everything, even if they do not know it, that we do not like. We are grateful for that. Stephen wants to do a postscript on this section and move on to the next section.

Q30 Stephen Williams: Perhaps our witnesses could never be politicians. Just a quick supplementary to David's line of questions, particularly to Sir Michael, who seems to be the main enthusiast for league tables. Just to be clear, is it Sir Michael's preference that, if league tables are going to exist, it would be better if the Government designed them, included all the variables on the tables, and published them like that? Is that basically what you would recommend?

Sir Michael Barber: If I have understood the question correctly—

Stephen Williams: At the moment, newspapers create league tables. The *Evening Standard* printed a league table, which I read on Thursday morning in London, and the Bristol *Evening Post*, which I saw when I got home in the afternoon, had a completely different league table, which was much better because it included free school meals, special educational needs students, the number of people entered and was measuring Level 4 rather than Level 5, which is what the *Evening Standard* seemed to be concerned about. So we had two completely different league tables at either end of the railway line. Would it better if the Government said that they were the league tables and that is what should be published?

Sir Michael Barber: I apologise for my misunderstanding. The Government should put the data out in formats that vary over time, and that is what has been happening. When the data is out there, individual newspapers can vary it. I was warning against the Government saying that they would not publish league tables at all, but the data getting out there and newspapers making up a set of league tables as happens in some countries and, indeed, in relation to higher education now. The fact that the Government are debating what should be in the league tables, which is after all public information that sets the standard for the system and gives parents information along with the various stakeholders, is right. Once the information is out there, newspapers can do what they choose.

Q31 Stephen Williams: I went to school in South Wales and, even though I do not like league tables, my natural curiosity leads me to want to know how Mountain Ash comprehensive school does in the league tables but I cannot find out. There are no league tables in Wales, so even though pupils sit the same public examinations as in England, there are no league tables. Does it necessarily follow that newspapers will create them if the data are not published?
Sir Michael Barber: Obviously, we shall see over time, but that is what has been happening around the world. One of the things that the Programme for International Student Assessment report says is that there is trend towards published public information about school performance. Indeed, that is associated with positive things in the PISA results.
Chairman: Let us look at grade inflation.

Q32 Stephen Williams: Every August, we go through the season of the three sets of SATs. Key stage results are published, as are A-levels and GCSEs. Different sections of the national media and commentators bemoan the declining standards compared with the time when they sat their examinations and so on. Is it the opinion of either of you that there really has been grade inflation at GCSE and A-level?
Professor Tymms: I shall respond by quoting the analysis of Dr Robert Coe of the data, which I can provide for the Committee, if necessary. We used our data in the Curriculum, Management and Evaluation Centre to examine matters. The way in which we analysed matters was to take data based on general developed ability, say, two years before GCSE and then look at the grades that the student gained at GCSE.

Q33 Stephen Williams: Key Stage 3 through to GCSE.
Professor Tymms: It was two years before. There is an assessment at he beginning of year 10 and then we look at the grades that were achieved. We can do that over many years. We take pupils with a particular level of ability and see what grades they get. Generally, we find pretty flat lines at GCSE. Standards appear to have been maintained at GCSE over several years. There is a little fluctuation according to some subjects, some of which apparently get easier while some apparently get a bit harder. However, the headline is pretty well standard. A2-level tells us quite a different story. If we use the same ability test, at the beginning of A2-level, and look at the grades, we find that pupils of a particular ability are getting higher and higher grades and have been for many years. In fact, if we went back some years, a D in mathematics might be the equivalent of getting a B now. That is quite a big jump. The biggest change is in mathematics, but it is less in others and there is a big difference in different subjects. It is complicated subject, but we were talking about fit for purpose. If we consider the purpose of A-level and selection for university, we see that Durham University's law department is inundated by students with straight As. The position is similar at Oxford and Cambridge, so to distinguish between them we create a market for producing tests for the selection of more students. The A-levels should have been doing that. We have a problem with the levels at A-level. So many students are getting As that we now need to distinguish between them.

Q34 Chairman: But only 20,000 students get three straight As out of all the people who take A-level. That must put matters into perspective.
Professor Tymms: Yes, but if you went back you would find that 30% used to fail A-level and get below an E. Now the number is down to just a few per cent. with straight failed A-levels. There has been a dramatic shift.
Stephen Williams: The 20,000 straight As would be enough to fill up all the departments at the top universities in the country.
Chairman: I am sorry, but it depends on what you call top universities.

Q35 Stephen Williams: Professor Tymms is saying that he accepts that there is grade inflation at A-level. How many people got a 2.1 at Durham 20 years ago compared with how many people get a 2.1 now?
Professor Tymms: There has been grade inflation there, but I do not know specifically about Durham University. I know about Harvard University.

Q36 Stephen Williams: Universities moan about the entry standards at A-level, but when I looked at it, lo and behold, I saw that the number of people getting 2.1 and firsts has gone up, because no one wants a 2.2 any more.
Professor Tymms: I am not going to defend that.
Sir Michael Barber: Peter probably knows better than me the data on A-levels. I just want to make one general point at the beginning. I believe that the kids coming out of our schools now are the best educated generation in history, and that owes a lot to the reforms and investment of the past 10 to 20 years. The kids do not get the credit that they deserve for that. They get run down a lot in the media, and that is a big problem. I very strongly believe that today's kids are the best educated generation in history. However, that is not to say that that is good or

equitable enough; I would like it to be better. I talked about the challenges of the 21st century, but I am very pleased that this generation is the best educated in history because of the problems facing not just this country but the planet generally over the next 10 to 20 years. That requires a well educated generation. My second point goes back to what we were saying before. Having a new independent exams regulator, as proposed by Ed Balls, will really help in this area. I hope that that will come to pass. Thirdly, the arrangements for doing A-level exams—retaking modules and so on—enable more young people to succeed. That may be one of the factors why Peter—and he may want to comment on this—sees what he is seeing. On GCSEs, I am glad to hear what Peter has to say. I believe—and I got into trouble for this in my first few months in the Department in 1997—that in the very early years of GCSEs, between 1988 and 1990, there was an element of grade inflation. There is an account of this debate in my book. The progressive changes in the QCA since then have tightened it up and held the standard rather well.

Q37 Stephen Williams: I was going to ask about the variables. I am sure that the National Union of Teachers and other teaching unions would say that we have the best qualified teaching profession that we have ever had, and that the quality of teaching is very high. However, is it also because the structure of the exams has changed? The modular system has been mentioned and the fact that you can retake modules. Therefore, can we really compare results now with those 10, 15 or 20 years ago, which the newspapers tend to do?
Professor Tymms: I recommend that the Committee talks to Dr Robert Coe, who has specifically studied the subject. I can just talk in general about it. There are several factors why that might have happened. Lots of things have changed here, so a direct comparison is not straightforward. However, modular has happened and there are more students. If you have more students, you want to aim your grades at the students in front of you; that is a natural thing to do. Yes, we wanted more people to go to university, so we have had to lower A-level standards in order to get them there. So there is a natural logic to this. I worry about the standards of mathematics and physics for students at the top end. I would look at the quality of syllabuses that are being covered and talk to mathematicians, physicists and chemists about what is actually happening. We need more scientists, and more scientists at a very high level. We need more people motivated to study science. There is a tendency to think that if we make those exams and give more grades, we will get more people studying it. Actually, some of the bright kids are challenged by really hard subjects and to make them easier is not helpful. It is a complicated situation, and attracting more people to science is perhaps outside our scope here.

Q38 Stephen Williams: Given that grades have gone up, and that is a given fact, does that mean that the standards themselves have been debased?
Professor Tymms: No, it does not automatically mean that. You need to look at this in more detail in order to check that. I am telling you that students with the same ability are getting higher grades, so you could argue that there has been better teaching between now and then, and that might indeed be the case, but we need to look at the standard setting and see what we mean by equivalent standards. This is a complicated area which evolves. No doubt the Committee will have heard of the Flynn effect. If you take non-verbal ability measures across the western world for the past 25 to 50 years, you will see that they have been rising steadily. People appear to be getting taller and cleverer. They are more able to do things that they have never done before. The same is not true for verbal skills. We also have the anti-Flynn effect. You will see a decrease in Piagetian levels of children just finishing primary school—Michael Shayer's work on that is very important. Why has that happened? Is it because we are taking away the Piagetian work in the early parts of primary schools that are now not focusing on that early development through play and so on? It is difficult to know that, but these are general patterns that we are seeing across the western world.
Sir Michael Barber: I can definitely say that my memory is not improving over time, but I just want to raise three general points. One is that I think that the quality of teaching and the quality of the teachers that we are recruiting have improved significantly. I think that young people are more motivated than they were 20 or 30 years ago. A lot of people in those days expected to get jobs in unskilled and semi-skilled work forces and did not need to try hard in school. This is the challenge for the future—we need to think about how we as a culture prepare ourselves for the 21st century as I described. There is an element in our culture that assumes that, if more children are passing exams, standards must have got worse. We must guard against that. We need a culture from business, universities, parents and the school system saying that more and more children can achieve high standards. That is what we need, and that is what we want to see in the 21st century.

Q39 Stephen Williams: One final question. Is it the Flynn or the Finn effect?
Professor Tymms: Flynn.

Q40 Stephen Williams: I heard about it on *Start the Week* this morning, and someone was pouring cold water on it, saying that factored backwards, it implies the Victorians were stupid, when clearly they were not. If grades have been inflated, and if it is accepted that roughly 90% of those who pass A-levels now go to university rather than straight into work, as was the case when I took them, are A-levels fit for purpose?
Professor Tymms: You really need to ask what the purpose is. If the purpose is straight selection to university, there is a problem at the top end with that differentiation. We need more differentiation, and if we do not get that right, other systems will come in—people will produce their own American SATs for

selection to university, or a new law test. That will undermine the purpose of A-levels, which have been a very good motivator in our colleges and sixth forms. There are some great teachers working in that area, and it would undermine that. There is another question about whether A-levels are fit for purpose. Do they prepare students well for their next stage of study? Again, it is quite complicated. AQA's research committee has been investigating whether that is the case. It has gone to the law departments and psychology departments to find out whether they believe that law and psychology A-levels and so on are useful. There is another issue out there. There are never straightforward answers, but we need to ask the questions. Are the students going on to university actually able to do those kinds of thing? People are always complaining about maths and reading, so we see four-year courses instead of three-year courses because students apparently have not done enough maths. If you are just asking straight whether they are fit for purpose, I do not think that they are fit for purpose at the top end for selection, but for the rest they do pretty well. I should add one other thing about A-level standards. It has to do with the setting of standards over time. I talked earlier about setting standards for key stage assessments over time. The way that it is done for Key Stage 2, for example, is multifarious. There are lots of ways to maintain the standards over time, but one way is to take the students who do the key stage assessment this year and give a proportion of them next year's test secretly to see how they do—pre-testing it with the next people and seeing what level they were given last year. It is not a perfect system, but it is an interesting way to do it. A-levels and GCSEs get no pre-testing. All the standard-setting is done afterwards on the basis of statistical relationships and judgments. No items used last year are used this year. In something like the programme for international student assessment, they do the tests, release half the items and keep some so they can be used next year to standardise next year's test. It is the same with the progress in international reading literacy study. A-levels and GCSEs do not have any pre-testing, which may be an issue that needs to be faced up. Most of the systems in the world have pre-testing.

Chairman: I am aware that we have two sections to complete this evening, and some of us want to hear Ed Balls in another place later. Sir Michael.

Sir Michael Barber: I will be brief. In an era when we are moving towards everybody staying compulsorily in full-time or part-time education until 18, which I believe to be absolutely right, A-levels are clearly not the whole answer to the challenge. To pick up on the point about fitness for purpose, we need to get apprenticeships working well. I spent Friday afternoon with some apprentices at the Rolls-Royce plant in Derby—a fascinating conversation. We need to get the new Diplomas to work well. We should make the international baccalaureate available. I am in favour of developing a range of possible qualifications for young people, so that we can have qualifications fit for the whole cohort, all of them have something to aim for and all of them go into the labour market with qualifications that have real value.

Q41 Chairman: If we want young people to stay on until 18, the natural school leaving age for learning and skills progression, what is the point of having a major exam at 16? Is it not becoming redundant?

Sir Michael Barber: When the full 14–19 programme is working well, the debate will change. I do not think that we are there yet, but I agree that that might well be part of the debate, absolutely.

Q42 Mrs Hodgson: I would like to move on to models of assessment, but I have a bit of a cold, so you must excuse my deep voice. I understand that, at the moment, the Government are doing about 500 pilots in schools on Making Good Progress. I understand that currently the main purposes of assessment are listed as points one to four. I just wanted to say something about point four: assessment for learning, improving both learning and teaching. I know that this Committee has heard my views on the personalised teaching agenda and I know that it is making good progress, emphasising more informal teacher assessment and personalisation in teaching. Regarding personalisation of teaching, should it not be specialisation in teaching? I say that because it touches on one of the things that I am concerned about, as the Chairman is well aware. Earlier, Sir Michael, you said, "The sooner you know the problem, the easier it is to fix it." So you probably can guess where I am going. I wonder why, when you were advising the Department for Education and Employment on the literacy hour and numeracy hour, you did not suggest that, when children are identified with, say, dyslexia, there should be specialist dyslexia teachers in every school to work with those children? So, getting back to the models of assessment and bearing my particular interest in mind, do you think that the current Key Stage tests remain the appropriate model of assessment and, if they are not, what alternatives would you suggest?

Sir Michael Barber: First of all, by the way, when I worked in the Department for Education and Employment on the literacy and numeracy hours and all of that, I had detailed conversations with the Dyslexia Institute and the British Dyslexia Association. Ken Follett, who is very actively involved in that world, was somebody whom I talked to often, and incidentally I still do talk to him. I think that what you say is right, that once you get really good teaching consistently across the cohort of literacy, most children will make progress, and then the ones that have a problem, whether it is dyslexia or something else, will be easier to identify. I think that the problem, if you go back before the literacy and numeracy strategies, was that children who had a problem got muddled up in the cohort, because nobody had invested in the teacher's skills to teach reading, writing and mathematics in the way that they are now generally able to do. So I completely agree with your point. Whether you use the word "personalisation" or "specialisation", I

believe very strongly that, as soon as a child is identified as having a problem such as dyslexia, there needs to be specialist people available to advise and help. Importantly, they need to advise the child on how to catch up with the cohort and not sink further behind the cohort. That is really important. I think that the progression pilots that you referred to, which the Government are running now, will effectively involve testing when ready; when the teacher thinks that a child is ready to go to the next level, they will use a single level test. That system has a lot of potential and we talked about it earlier in the Committee. I have been an advocate of just-in-time testing since the mid-1990s, when I published a book called *The Learning Game*, but they have to get the detail right. That is why I think that it is important that this type of testing is being piloted.

Professor Tymms: I have talked about the present system, so I will not add to what I have said about that. Let me just pick up on the teacher judgment and the single level test, because I read about that in *The Times* today and I had read some previous material in tender documents finalising the test data. I just wonder if I have got it right. Apparently, under this system the teachers will make judgments, then the pupils will do the tests and that information will be used to feed in to the information going in to league tables and so on. However, now we have cut off the test, which is security, and we are relying on the teacher judgment, but the teachers will be judged by their judgments. Surely that cannot be the way that the system will operate. That is one thing that puzzles me here. The second thing is that, if we are going to have a single test to do that, we know that, at the moment, the tests, say at Key Stage 2, which I regard as good, reliable, valid tests, have pretty big margins of error when it comes to assessing a particular level of a child. Therefore, by focusing on a single level, they will be less accurate than that. That will be worrying about the quality of the data, so I would be keen to see the results of the trials that are being done and whether that system is viable and produces good, reliable data on those students. I also noted that it suggests two tests a year for a pupil, rather than one, which seems a strange route to take. Thinking more broadly about the personalised and specialised learning, I have some sympathy with what you are saying about the specialised learning, but I also have sympathy for the personalised learning. With regard to the assessment that we use currently for children just starting school, there are some children whose vocabulary levels are extremely low, most are pretty good for children of that age and some children at the top are quite exceptional—some of them start school with higher language levels than some of the 11-year-olds leaving primary school. The teacher of such a class has to deal with that group year on year with that phenomenal range in mathematics, language and reading, and that is mixed-ability teaching, which means that you have to do something different with different children in the same class. There are other models: I mentioned the computerised diagnostic assessment earlier. In fact, in Northern Ireland, from this term, all 900 primary schools will not do SATs, but will do computerised diagnostic assessments that will give information to the teacher on the strengths and weaknesses of individual children so that they can improve that with the feedback. Therefore, there is a different model operating there, and we could look at how those things are operating differently.

Q43 Mrs Hodgson: With regard to what alternative you would suggest, what jumped out at me was that *Making Good Progress* has been called a one-way ratchet because the teacher will decide when the child is ready for that test. A child might consistently get bad tests, but if they are re-tested on a good day the ratchet will go up. There is never a chance, however, for the child to be levelled down, so it could just be that they have a good test on a good day. It therefore produces high levels of certainty so that misclassification is minimised, or re-testing of doubtful cases does not happen. I have not got the full details of *Making Good Progress*, but I do not know if there are any alternatives available instead of the new single level tests.

Professor Tymms: Yes, within our centre we run the Performance Indicators in Primary Schools project for schools. Many schools do the test with the children every year, and we look at year on year progress. They get flat graphs on that, or computer diagnostic assessments would do that—there are plenty of systems out there. This is just one system, and I really think that we need to look at the trials and the statistics on that to see how well they look. We need to monitor the progress of children and spot them when they are falling by the wayside.

Sir Michael Barber: Clearly, there are alternative systems. The technical details of the progression pilots need to be worked through to ensure that the problems that you and Peter have drawn attention to do not occur. I think that there is a lot of promise in them, but the detail will be crucial, as I have said consistently. I know that Committees are criticised for travelling, so maybe you could do this by reading papers or by video conference, but if I were you, I would look at what is being done in New York City, Hong Kong, where the secondary curriculum is being completely reorganised, and Ontario, where the literacy and numeracy programme, which was originally modelled on ours, is being built on and taken forward. These examples all have implications.

Q44 Mrs Hodgson: You mentioned personalised learning. I went on a delegation to Sweden that looked at the free school model that is used there, and I was very interested in how they really do focus on personalised learning, as they stream the children according to ability, not age. You might have one nine-year-old who was in with 11-year-olds for numeracy, but in with seven-year-olds for literacy. The children are mixed up according to their ability, which is very interesting.

Professor Tymms: In Bob Slavin's *Success for All* programme, he points to the good research evidence for bringing together children with the same reading age some time in the week. So that is an interesting way forward.

Sir Michael Barber: I agree with that.
Chairman: Dawn has waited extremely patiently to ask about the unintended consequences of testing.

Q45 Ms Butler: Sir Michael, you mentioned our basically being future-proof, and I completely agree: we have to make sure that we teach young people for the future, and the Government are right still to focus on maths, English and science as the core subjects. My first question is about testing. Professor Tymms, you said that it was not the testing, but the pre-testing that was the problem for the younger kids. You then said that there was no pre-testing for GCSEs and A-levels. What are the effects of that amount of testing on children, teachers and schools?
Professor Tymms: I am using "pre-testing" with two different meanings, so I must clarify that. What I meant in relation to setting standards was that the exam-awarding bodies did not pre-test the GSCE tests before they gave them out for real. What I meant in relation to primary schools was that the schools themselves take past papers and get their kids to redo them. Of course, that happens at GCSE as well—pupils will have mocks and practise this and that. The teachers do lots of previous work, but the pre-test is done at key stage assessments by QCA or whoever is employed to do it; it does not happen at A-level and the rest in the standard setting. That just clarifies the point.

Q46 Ms Butler: Wonderful. So what do you think the effects of that amount of testing are on children, teachers and schools?
Professor Tymms: They are multifarious. When you set up a system, you never quite know what is going to happen, and there are lots of unexpected consequences. We have to worry about the focus and the narrowing of the curriculum. Of course, we want to get reading, writing and maths right, but we also want drama and physical activity—we want to keep the children physically active—and there is evidence that that has decreased. In fact, in 2002, with Andy Wiggins, I did a survey comparing Scottish schools and English schools and found evidence of the narrowing of the curriculum, a blame culture in the classroom and so on. We need to watch such things to see what is happening—we need to track and monitor the monitoring. There are unintended consequences, including a focus on borderline children, which is an unhealthy thing. There is a focus on the ones who are likely to get the 4 A*s to C or the children who are not going to get Level 4. Little clubs are therefore set up to work on the borderline children, rather than the child with special needs. Lots of peculiar things go on as a result.
Sir Michael Barber: When I worked in the delivery unit, we looked at a lot of targets and data sets, and people predicted perverse or unintended consequences. We used to say, "Obviously, you should just predict the ones you think will happen and then we'll check." If you focused on street crime, for example, the police would predict that other crimes would get worse. In fact, that is not what happened, but it is always worth checking those things. On the level boundaries, we found that although the target was about Level 4, the percentage achieving Level 5 rose very rapidly, even though that was not the borderline at stake. Good teaching is good teaching, just as good policing is good policing. I would like to say two other things. Literacy and numeracy underpin the whole curriculum, and unless you get them right in primary school, young people will be held back in all kinds of ways, including in drama and all the other things that really matter. The second thing that I want to say is that, on the whole, the schools that do best academically also do best in a wider set of outcomes, because they are well-run institutions teaching well and doing everything properly. That is not a perfect fit, but it is generally the case. It is absolutely right to focus on literacy and numeracy, but of course you also want the wider curriculum for young people.

Q47 Ms Butler: That leads me to my next question. Would the performance and so on of schools be improved if we used a separate mechanism, such as reforming Ofsted inspections? You talked about Ofsted looking at all the different variations such as the leadership of schools and so on. Would improving Ofsted inspections improve schools and their overall performance?
Sir Michael Barber: Peter may want to come in, because he has had strong views for many years on Ofsted, but I think that Ofsted should constantly keep its inspection process under review. Since Ofsted was set up in its current form, it has been a positive influence on the schools system over the past 10 to 15 years, but it can always get better. As implied in your question, it should be the institution that looks at those wider things, including the ethos of the school, which matters so much, and its comments on them should get you in, beneath, below and around the data from the tests. Ofsted should constantly keep its processes under review. My view is that all processes, including leadership training, professional development for teachers and Ofsted, should focus in the next decade on achieving a consistent quality of classroom teaching. I quoted Andreas Schleicher, who said we are doing more of the right things than any other system in the world in England, but we have not yet had the impact on consistent classroom quality, so I should like to see Ofsted, professional development and leadership development all focusing on that, because it is the central challenge for our schools system.
Professor Tymms: Just before Ofsted changed to its present system, a paper was published by Newcastle university—by Shaw, Doug Newton and others—in which the authors compared the GCSE results of a school shortly after an Ofsted inspection with what it normally achieved. They showed that immediately after the inspection, their results were worse, which is interesting, considering the amount of money that was spent just to frighten the teachers. After that, Doug Newton was called in by Gordon Brown for an interview, and shortly afterwards the money for Ofsted was reduced and we went to the cheaper form of inspection. We need a thorough examination of Ofsted's impact on schools. What is it actually

doing? That is exactly your question, but rather than give an opinion, we should deliberately examine it to see what the impact is by looking at schools before and after they have inspections, and tracking them statistically across the country, because it is not clear that inspections are improving schools, although they might be. Neither is it clear that they are damaging schools, but they might be. We need to see that kind of evidence. It is a lot of money and there is a particular theory behind it. Another point that links into that is the view of what matters in the educational system. Michael has been saying that teachers matter, and I agree absolutely. He has also emphasised the importance of heads, but it is not so clear to me that heads are key with regard to reading and maths. In fact, what we have in schools are loosely coupled organisations: the head must influence this or that, and there is the teacher in the classroom. When I undertook a recent examination of 600 secondary schools and 600 primary schools, and looked at their value-addeds and how they changed when the head changed, I could find no evidence for such change at all. Actually, the teacher is the key. The head is vital for other things, such as the morale of staff, the building of new buildings and the design of the curriculum—appointing good staff is one vital thing that the head does—but we need to think about structure. We need to monitor things continuously and always ask what is the impact of what we are paying our money for. Ofsted is one of those things.

Sir Michael Barber: We can get caught up in metaphors, but the way I see it is that the head teacher's role is like the conductor of an orchestra. They do not play a single instrument, but if they do their bit, everybody else plays better. That is probably what we are trying to do with head teachers, particularly in our devolved system in which heads are given a lot of discretion.

Q48 Chairman: You have both been in this game for quite some time. A week is a long time in politics, and 10 years is an awfully long time in politics. If you could go back to when you started, what would you do differently, not only to drive up standards—one of you said that the standards are in the heart, rather than just the head—but to increase the ability of children to excel within themselves?

Sir Michael Barber: In the book I mentioned earlier, *Instruction to Deliver*, which was published in the summer, I own up to a whole range of mistakes. One reason for my looking slightly quizzical when you asked that question, is that I was thinking, "How long have you got?" I could spend the next hour or so talking about this, but I know that you have other things to do.

Chairman: We have the book to refer to.

Sir Michael Barber: First, something in which I was personally involved that I would see as a mistake took place in 2000. After the big jumps in numeracy and literacy that we have been debating, there was a general tendency, of which I was a part, to consider that primary school improvement had happened and that it was then all about secondary schools. That took the focus off, but we were really only at the beginning of seeing that improvement through. Secondly—this is a detail, but it is important, looking back—in the 2000 spending review, we set a new target for primary school literacy, aiming to raise it from 80 to 85%. I think that that was a mistake because we had not reached the 80% target. It was demoralising. I, personally, regret not negotiating more vigorously at the time. If you look in my book you will find a whole list of things that I got wrong. Overall, I am very proud of the contribution that I have been able to make to improving the education system over the last decade. While we could have been bolder and we could have achieved more, I am absolutely confident—I think the data confirm this—that we have the best-educated generation in history. There is much more to do to prepare for the 21st century, but it has been a great experience.

Q49 Chairman: Something quite interesting that you said earlier was that it is not we who are making these demands—it is the world. It is the competitive global economy and so on. Many countries seem to be responding to that task, not by using testing and assessment and the path that you or the Government have chosen, but by choosing very different ways. People tell the Committee that the curriculum is too narrow, that people teach to the test and that children no longer get the chance to explore a whole range of activities and subjects as they used to do. What do you say to people who say that?

Sir Michael Barber: Two things. One is that I am certainly not arguing, and that may now be my fate in history, that testing and assessment are the single lever to drive improving standards. They are part of a whole system. The crucial elements are combining the challenge that comes from the testing and accountability system with serious support, investment in teachers' skills, and, as Peter said, giving teachers the capacity to do the job. It is the combination that I believe in. Systems that have pressure without support generally do not succeed and systems that have support without pressure do not succeed either. It is getting the combination right that is the key, particularly when you want to change things. Some systems—Finland is an example—recruit good people into teaching, as they have a high standard among their graduate distribution, and they train them well. Their standards have been established, and have got into teachers' heads so they need less testing as they are already established at the top of the world league tables. If you are going to try to change things, the combination of challenge and support is most likely to get you there.

Q50 Chairman: Peter, what should they have done that they did not do?

Professor Tymms: First, they should have taken notice of the research evidence of what works. I do not mean the survey, or what is associated with what works, but what changes were made and where we saw the difference. In particular, I would go for randomised control trials. In reading, for example,

there is a wealth of knowledge. We know more about reading and how to help children with reading. That knowledge was more or less ignored when we were making changes, so evidence is importance, and light of that I would go to the experts. When the School Curriculum and Assessment Authority and its precursor, the School Examinations and Assessment Council, were set up, that was done without any test experts at all. It is only now, after the QCA has been put in place, that people are available who really knew about tests and the way forward. Now, the standard has been set properly. When it was done earlier, they would buy some people in and reckon that it could be sorted out. We need experts. When Estelle Morris spoke to the British Educational Research Association meeting a little while ago, she said that while she was Secretary of State she took almost no notice of the research that was around. I find that extremely worrying. We need to take notice of the research, rather than surveys and statements such as "This person is doing better," or "My father said this and therefore it is good for me." We should look at what has been done in randomised controlled trials that have been shown to work. Before we put in new systems we need to trial them and check that they work. When the national literacy strategy was going to be rolled out, a trial was running, which was stopped before the strategy was ready. Everybody had to do something that had not been trialled. Later, an evaluation was made *post hoc*, when everybody was doing the same thing and it was too late. We need to compare this and compare that. That is really important. There is real knowledge out there. We can evaluate things, and when we put in new systems, we need to track them over time. We need, too, to get good experts. Above all, we need good teachers. I absolutely agree: we need good teachers and we need to trust them. Perhaps we need to free up the curriculum, and perhaps teachers should experiment with it. To find new ways of working, we have to go outside England. Why cannot we allow in people to look at new ways of working, assessment and so on? They are pretty good people, those teachers. We absolutely rely on them and we should rely on them more.

Q51 Chairman: When the previous Committee looked at the issue of teaching children to read, we came up with two major recommendations. We tried to check evidence-based policy, and the evidence suggests that if you take any systematic way of teaching children to read, it works. We also said that it was to do with the quality of the teachers. We found that there is very little evidence that anyone ever trained our teachers to teach children to read on any basis at all. The Government then rushed off—influenced by a former member of this Committee, I believe—to set up a Committee that recommended synthetic phonics, which had been trialled only in Clackmannanshire. We were a little disappointed that our recommendations were not fully taken on board.

Sir Michael Barber: Chairman, I cannot help noticing the imbalance in your questions. You asked me what mistakes I have made and then asked Peter what mistakes I have made as well. I wish that you had asked him what mistakes he has made, but since you did not—

Q52 Chairman: What mistakes has he made?
Sir Michael Barber: You should ask him. However, since I have managed to get the floor, I think that basing policy on evidence is very important. I talk a lot about evidence-informed policy, and I believe that the programmes that we have been talking about are among the most evidence-informed policies ever, and we have had better evidence on which to base them. Another question that arises when you are involved in government is "how long you have got?" Looking at the data that we had on primary reading standards prior to 1996 and looking at the challenges of the 21st century—Peter and I are broadly in agreement about this—something had to be done urgently. We took the evidence that was available. There is a great report by Professor Roger Beard—he is now at the Institute of Education—which summarises the evidence base for the literacy strategy. We worked very hard to take all the evidence into account. I have been honest about mistakes that I made, but overall it was one of the most evidence-informed policies ever. Its replications around the world demonstrate that it can be replicated with variations with the same results.

Q53 Mrs Hodgson: On the point about good teachers, I have recently returned from Singapore where, as in your example of Finland, teachers are recruited from the top 10% of the cohort of university graduates. The Government offer whatever incentives they have to. They also headhunt teachers—they spot them. The education officers monitor graduates. They go up to them and say, "Have you thought about becoming a teacher?" The teaching profession is held in much higher regard, and is revered as it was here 50 or 60 years ago. The pay reflects that. Teachers are paid a lot better. There is an incentive, because if students are bright and go into teaching, they might be sent to the UK, where their teaching is funded. They then go back and teach in Singapore. It is interesting that we are not at that stage.
Sir Michael Barber: That is one of the examples that we use in our recently published report, *How the World's Best-Performing School Systems Come Out on Top*. We looked at systems on several continents, including the one in Singapore. What you say is absolutely right, with the exception that they do not pay teachers more than here. However, they pay them reasonably well. If you talk to the Singaporean Education Minister, as perhaps you did, you find that they are constantly looking for ways to motivate young people to go into teaching in the future. We have done reasonably well on that over the last few years, but we have a long way to go and can never be complacent about ensuring that we secure really good entrants into the teaching

profession, both out of university, and among mature people who have gone into other lines of work and change to teaching.

Q54 Chairman: Thank you, Sir Michael and Professor Tymms. It has been a really good sitting—a marathon sitting. I am sorry that we have kept you so long, but it has been so absorbing and interesting: we have enjoyed it immensely. I am sorry that we were not an all-party Committee today. It is a great pity that you did not have a slightly broader range of questions, but you did have a fair range. It was two-party, but not all-party. Will you remain in contact with us? If we want to come back and ask you some other questions about the evidence that you have given, will you be accessible?

Sir Michael Barber: Absolutely.

Professor Tymms: Sure.

Chairman: I am glad that we are not paying the full consultancy fee for today. Thank you very much for coming.

Monday 17 December 2007

Members present:

Mr. Barry Sheerman, in the Chair

Annette Brooke
Ms Dawn Butler
Mr David Chaytor
Mr John Heppell
Fiona Mactaggart
Lynda Waltho
Stephen Williams

Letter to the Education and Skills Committee from Dr Ken Boston, Chief Executive, Qualifications and Curriculum Authority (QCA)

The Qualifications and Curriculum Authority welcomes the opportunity to present this submission to the Select Committee.

The QCA is the statutory national authority for testing and assessment in England. It is responsible for development and delivery of the national curriculum tests, and for provision of the national results to Government. It regulates the market for delivery of nationally-accredited general qualifications and vocational qualifications by awarding bodies. It is responsible for the maintenance of assessment standards year on year. It is leading and managing current and projected reforms in the delivery and modernisation of tests and examinations.

The QCA is thus at the fulcrum of the national testing and assessment programme. It delivers, promotes and defends testing and assessment as a means for securing better teaching and learning, and for measuring and reporting change in educational outcomes at individual, institutional and national level. This point is important: the QCA is the guardian of standards, and its public contributions to the discussion of assessment reform are entirely from that perspective.

It is also important to acknowledge the strengths of the current assessment system, and in particular those of the national curriculum tests. Each national curriculum test is the product of a developmental process extending over more than two years, during which the test items and mark schemes are rigorously pre-tested, trialled and refined, and then pre-tested, trialled and refined again. The quality of these tests stands comparison with any similar tests developed internationally. Further, the techniques developed by the DfES to analyse change in educational performance at individual, school and local authority level over time, and to plan and deliver strategic interventions in response, have now reached a level of sophistication and practical utility which is world class. Such strengths are the product of well-managed and steady evolution, which must be the process by which further development continues to occur.

Assessment is integral to good teaching and learning: if teachers understand assessment better, performance will rise. Effective classroom assessment today will improve teaching and learning tomorrow. Timely and effective assessment, which measures and supports their learning, should be an entitlement for all young people. At the same time, the Government must have the most accurate and best possible measure of educational performance at school, local authority and national level. The introduction of the new secondary curriculum provides a timely opportunity to reflect on the best curriculum assessment arrangements from 2011 onwards.

This submission consists of five papers.

Paper 1, *Evaluating assessment systems*, has been prepared to assist the Select Committee to identify and consider the many complex questions that will arise during the course of the inquiry. It focuses on the key issues of validity, reliability and purpose, which will be at the heart of the Select Committee's deliberations.

Paper 2 is a summary of observations about the national curriculum testing programme, drawn from systematic and formal consultation with schools over a long period. As with any testing system, there is room for further development and extension: the present arrangements provide a foundation on which this can occur.

Within the very foreseeable future, it will be possible for traditional pencil-and-paper assessment largely to be replaced by on-line and even on-demand testing, should that be the desired policy direction.

Paper 3, *Testing and assessment: the use of electronic media*, describes the current status of these developments and projected future directions.[1]

Paper 4[2], which has been published on the QCA website, sets out the regulatory regime to support the development of e-assessment by awarding bodies, within a national framework which guarantees both standards and security.

[1] Not printed. Available on Committee website: http://www.publications.parliament:uk/pa/cm200708/cmselect/cmchilsch/memo/169/contents.htm
[2] ibid

Paper 5[3] is a comparative analysis of testing and assessment systems within a range of other countries. Some of these are above us, and others below us, in terms of international indicators of educational and economic performance.

This submission focuses largely on assessment in the primary and early secondary years of schooling, although much is also relevant to the examinations for the GCSE and GCE qualifications. This has been in response to the broad scope of the terms of reference of the Inquiry, and taking into account the recent attention given by the Select Committee to 14–19 education and the adult skills agenda. There is much to be said however about the assessment and reporting of practical competences and skills in the workplace, and even remotely by the use of technology. Should this be an area the Select Committee wishes to explore, QCA would be very willing to provide a further submission.

We look forward to offering oral evidence in support of the Select Committee Inquiry, and would value the opportunity to respond to matters raised by other contributors.

June 2007

Paper 1

EVALUATING ASSESSMENT SYSTEMS

Executive Summary

The following report highlights five common confusions related to the evaluation of educational assessment systems:[i]

1. the difference between validation and evaluation, where validation (concerning the accuracy of inferences from results) is merely one component of evaluation;
2. the meaning of "purpose" in "fitness-for-purpose", which can be interpreted in a variety of different ways, all of which are (differently) relevant to evaluation;
3. the number of purposes which can be identified, which is much higher than tends to be appreciated (for example, national curriculum test results are probably used for at least 14 different purposes);
4. why it matters when results are used for many different purposes, which is because different uses require that different kinds of inference be drawn from results, so results that warrant accurate inferences for one purpose may not warrant accurate inferences for another; and
5. the many components of a rigorous evaluation, which include analysis from the perspectives of technical accuracy, moral defensibility, social defensibility, legal acceptability, economic manageability and political viability.

The report emphasises the importance of: distinguishing the different meanings of similar terms; distinguishing logically separable evaluation questions; and distinguishing the many alternative perspectives on evaluation.

Evaluating Assessment Systems

1. *Introduction*

1.1 This report explores the concept of evaluation, as it applies to educational assessment systems; and then presents a framework for evaluating them.[ii] It is intended as a tool for helping the Education and Skills Select Committee to grapple with the many questions which comprise its New Inquiry into Testing and Assessment, particularly the very broad ones, like:

— is the testing and assessment in "summative" tests (for example, GCSE, AS, A2) fit for purpose?

— should the system of national tests be changed? If so, should the tests be modified or abolished?

1.2 This report does not offer a view on the legitimacy of our present national assessment systems. Instead, it offers a way of organising evidence and argument in order to reach such a view. It helps to identify what makes for a good evaluation question and what makes for a good evaluation conclusion: issues which can seem deceptively straightforward at first glance. In particular, it aims to expose a number of common confusions which can mislead the unwary inquirer.

1.3 This report offers generic insights, which apply in the same way across the spectrum of educational assessment systems (occupational, vocational, general; tests, examinations, teacher assessments; on paper, on-screen, online; etc.).

[3] ibid

2. *What does evaluation entail?*

2.1 The first confusion to confront is the nature of evaluation itself. An easy mistake to make is to reduce the big programme of evaluation to the smaller programme of validation.[iii] The central question at the heart of validation is this: are the inferences that we draw from our assessment results sufficiently accurate (for the uses to which they will be put)? Or, less formally: are our results accurate or not? Although this is a necessary and fundamental component of evaluation, it is still only one component. Evaluation requires the inquirer to consider any question that might bear upon the legitimacy of the assessment system, such as:

- might the way in which test results are reported have positive or negative impacts (eg, is it better simply to rank students or to tell them how much of a programme of study they have "mastered")?
- might the fact of testing itself have positive or negative impacts (eg, does the inevitable "washback" support or detract from good teaching and learning)?

2.2 Evaluation entails marshalling as much relevant evidence and argument as possible, to judge whether systems work as they are intended to and in the best interests of participants, stakeholders and society. The central question at the heart of evaluation is this: are our assessment systems fit-for-purpose?

3. *What does "fit-for-purpose" mean?*

3.1 A second confusion to confront is the meaning of fitness-for-purpose. Before exploring the concept of "fitness" we need to work out what we mean by "purpose". This is not as straightforward as it might sound. Consider the following three interpretations.

1. The purpose of assessment is to generate a particular kind of result. For example, students sit an exam in GCSE science to rank them in terms of their end-of-course level of attainment.
2. The purpose of assessment is to enable a particular kind of decision. For example, students sit an exam in GCSE science so that we can decide whether they have learned enough of the basic material to allow them to enrol on an A level science course.
3. The purpose of assessment is to bring about a particular kind of educational or social impact. For example, students sit an exam in GCSE science to force them to learn the subject properly, and to force their teachers to align their teaching of science with the national curriculum.

3.2 Obviously, to judge whether a system is fit-for-purpose, an evaluator needs to begin by identifying the purpose, or purposes, for which the system is supposed to be fit. However, if the evaluator is confused by the different possible meanings of "purpose", no satisfactory conclusion will be reached. This is why it is essential to distinguish these different interpretations; all of which are perfectly reasonable; and all of which need to be considered in their own right when mounting an evaluation.

3.3 As it happens, there is yet another interpretation to be wary of:

4. The purpose of the qualification is to bring about a particular kind of educational or social impact. For example, students study GCSE science to support progression to a higher level of study (for those who wish to), and to equip all students with sufficient scientific literacy to function adequately as 21st century citizens.

Again, this is a perfectly reasonable interpretation of "purpose". However, strictly speaking, it is not within the scope of an evaluation into the legitimacy of an assessment system. Instead, it implies a broader evaluation remit, into the legitimacy of an educational programme. It would be perfectly possible to have a legitimate educational programme with an illegitimate assessment system; and vice versa. The two evaluation foci need to be kept quite separate.

4. *How many purposes are there?*

4.1 A third confusion concerns the number of purposes which need to be considered when evaluating an assessment system. This is best illustrated by considering the uses to which assessment results are put (interpretation 2 above). In England, we have become familiar with classification schemes such as that presented in 1988 by the Task Group on Assessment and Testing:

- formative uses (assessment for learning)
- summative uses (assessment of learning)
- evaluative uses (assessment for accountability)
- diagnostic uses (assessment for special intervention).

4.2 Although this kind of scheme is useful, it fails to convey the full complexity of the situation. In fact, there are many more categories of use to which educational assessment results might be put. Figure 1 illustrates 22. The categories presented in Figure 1 are quite loose—and occasionally shade into each other—but the point isn't to present a definitive taxonomy, merely to illustrate just how many possible kinds of use there are. In fact, distinctions can often be made within categories, between uses which might recommend quite differently designed assessment systems (eg, long-, medium- and short-term system monitoring).

Figure 1

SOME EXAMPLES OF THE MANY KINDS OF USE TO WHICH ASSESSMENT RESULTS CAN BE PUT

1. social evaluation (to judge the social or personal value of students' achievements)

2. formative (to identify students' proximal learning needs, guiding subsequent teaching)

3. student monitoring (to decide whether students are making sufficient progress in attainment in relation to expectations or targets; and, potentially, to allocate rewards or sanctions)

4. diagnosis (to clarify the type and extent of students' learning difficulties in light of well-established criteria, for intervention)

5. provision eligibility (to determine whether students meet eligibility criteria for special educational provision)

6. screening (to identify students who differ significantly from their peers, for further assessment)

7. segregation (to segregate students into homogeneous groups, on the basis of aptitudes or attainments, to make the instructional process more straightforward)

8. guidance (to identify the most suitable courses, or vocations for students to pursue, given their aptitudes)

9. transfer (to identify the general educational needs of students who transfer to new schools)

10. placement (to locate students with respect to their position in a specified learning sequence, to identify the level of course which most closely reflects it)

11. qualification (to decide whether students are sufficiently qualified for a job, course or role in life—that is, whether they are equipped to succeed in it—and whether to enrol them or to appoint them to it)

12. selection (to predict which students—all of whom might, in principle, be sufficiently qualified—will be the most successful in a job, course or role in life, and to select between them)

13. licensing (to provide legal evidence—the licence—of minimum competence to practice a specialist activity, to warrant stakeholder trust in the practitioner)

14. certification (to provide evidence—the certificate—of higher competence to practise a specialist activity, or subset thereof, to warrant stakeholder trust in the practitioner)

15. school choice (to identify the most desirable school for a child to attend)

16. institution monitoring (to decide whether institutional performance—relating to individual teachers, classes or schools—is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions)

17. resource allocation (to identify institutional needs and, consequently, to allocate resources)

18. organisational intervention (to identify institutional failure and, consequently, to justify intervention)

19. programme evaluation (to evaluate the success of educational programmes or initiatives, nationally or locally)

20. system monitoring (to decide whether system performance—relating to individual regions or the nation—is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions)

21. comparability (to guide decisions on comparability of examination standards for later assessments on the basis of cohort performance in earlier ones)

22. national accounting (to "quality adjust" education output indicators)

5. *Why does the large number of purposes matter?*

5.1 Confusion number four concerns why it matters that results can, and often are, used for multiple purposes. Surely, some would claim, as long as assessment results are accurate, then we ought to be able to use them for any purpose we like? Unfortunately, it's not quite as straightforward as that. The point is best illustrated by considering what it might mean to explore validity for different uses of results.

5.2 As mentioned earlier, the central question at the heart of validation is this: are the inferences that we draw from our assessment results sufficiently accurate (for the uses to which they will be put)? This has become the standard technical definition, and the word "inference" is significant because different kinds of inference may be drawn—from the same assessment result—to support different kinds of use. This is not at all obvious, so it warrants a brief technical detour.

5.3 Assessment instruments are designed to support specific kinds of inference. So, an end-of-Key-Stage test will be designed primarily to support an inference concerning a student's "level of attainment at the time of testing". Let's call this the primary design-inference. And let's imagine, for the sake of illustration, that

our assessment instrument—our Key Stage 2 science test—supports perfectly accurate design-inferences. That is, a student who really is a Level X on the day of the test will definitely be awarded a Level X as an outcome of testing.

5.4 In fact, when the test result is actually used, the user is likely to draw a slightly (or even radically) different kind of inference, tailored to the specific context of use. Let's call this a use-inference. Consider, by way of example, some possible use-inferences associated with the following result-based decisions/actions.

1. A placement/segregation use. The inference made by a Key Stage 3 head of science—when allocating a student to a particular set on the basis of a Key Stage 2 result—may concern "level of attainment at the beginning of the autumn term".
2. A student monitoring use. The inference made by a Key Stage 3 science teacher—when setting a personal achievement target for a student on the basis of a Key Stage 2 result—may concern "level of attainment at the end of Key Stage 3".
3. A guidance use. The inference made by a personal tutor—when encouraging a student to take three single sciences at GCSE on the basis of a Key Stage 2 result—may concern "general aptitude for science".
4. A school choice use. The inference made by parents—when deciding which primary school to send their child to on the basis of its profile of aggregated results in English, maths and science—may concern "general quality of teaching".
5. A system monitoring use. The inference made by a politician—when judging the success of educational policy over a period of time on the basis of national trends in aggregated results in English, maths and science—may concern "overall quality of education".

5.5 Each of these result-based decisions/actions is premised on the use of Key Stage 2 test results.[iv] Yet, in each case, a slightly different kind of inference is drawn from them. None of these use-inferences are precisely the same as the primary design-inference (the inference that the Key Stage 2 test result was primarily designed to support). Indeed, some of the use-inferences are radically different in nature from the design-inference.

5.6 So, when it comes to validation (establishing the accuracy of inferences from results for different purposes) the implication should be clear: accuracy needs to be established independently for each different use/inference. Results will inevitably be less accurate when used as indicators of future attainment than when used as indicators of attainment at the time of testing. And results may be less accurate still when used as indicators of general aptitude rather than as indicators of attainment. When it comes to using results as indicators of quality of teaching, or quality of education, we should expect less accuracy still, since the qualitative difference between the design-inference and the use-inference is so great.

5.7 This begins to ground the most important observation of the present report: an assessment system which is fit for one purpose may be less fit for another and could, conceivably, be entirely unfit for yet another.[v]

5.8 Recall that, for the sake of illustration, this section has focused purely upon the exploration of validity for different uses of results. The full story of evaluation needs to be far more embracing.

6. *Can we construct a framework for system-level evaluation?*

6.1 The fifth and final confusion concerns what an overall evaluation ought to look like. This is where we begin to explore the concept of "fitness" in requisite detail. There are at least six more-or-less discrete perspectives from which assessment systems need to be evaluated:

1. technical accuracy
2. moral defensibility
3. social defensibility
4. legal acceptability
5. economic manageability
6. political viability.

Each of these will be considered briefly below.

6.2 Technical accuracy

6.2.1 The first evaluation perspective is technical accuracy; essentially, the concept of validation. It poses the question: overall, how accurate can we expect inferences from results to be? And, as explained previously, this question needs to be explored independently, for each discrete use of results, ie, for each discrete use-inference.

6.2.2 Unfortunately, it isn't always obvious which inference underlies (or ought to underlie) each use, which complicates the matter greatly. An example from system monitoring is helpful here. When considering trends in the percentage of students who attain at or above Level 4 in science at Key Stage 2, are we (or ought we to be) drawing inferences concerning:

- the level of attainment of specific cohorts of students from one year to the next (where attainment is defined in terms of an explicit programme of study in science)?
- the level of proficiency of the national cohort over time (where proficiency is defined in terms of an implicit "fuzzy set" of essential core competencies in science)?
- the level of performance of teachers of science over time?
- the overall effectiveness of policy and practice related to the teaching of science over time?

6.2.3 The first of the above use-inferences will be closest to the design-inference (being defined in terms of an explicit programme of study) and would, therefore, be likely to facilitate greatest accuracy. However, it's arguably of least interest as far as system monitoring goes, because it's furthest away from the ultimate system monitoring ideal of identifying whether "things are better now than before". For instance, in the first years of a new curriculum for science: we would expect average attainment to increase gradually as teachers became better at delivering the new curriculum (with more practice and training in teaching the new elements, with an improved selection of curriculum-specific text books and resources, and so on); and we would expect average test performance to increase gradually as teachers became better at preparing students for the specific form of assessment associated with the new curriculum. Such gradual increases would seem to be inevitable.[vi] However, they would not necessarily imply that teachers were becoming better at teaching, per se; nor even that they were necessarily becoming better at teaching science; nor would it necessarily mean that students of the new curriculum were more accomplished than students of the old curriculum. As far as system monitoring is concerned, we probably ought to be validating in terms of more distant use-inferences (eg, inferences concerning the performance of teachers, or the overall effectiveness of the system), since these have greater real-world significance. Unfortunately, these are correspondingly much harder to validate.

6.2.4 In theory, the analysis of accuracy is largely technical, using established methods for eliciting evidence of content validity, predictive validity, reliability, and so on. However, in practice, exactly how the various sources of evidence are synthesised into an overall judgement of accuracy is often not clear and, consequently, not that technical after all.

6.2.5 The logic of this perspective is essentially that: all other things being equal, more accuracy is better; and that accuracy must significantly exceed a threshold of chance.

6.3 Moral defensibility

6.3.1 The second evaluation perspective is moral defensibility. It poses the question: given the likelihood of inaccurate inferences from results, and the severity of consequences of error for those assessed inaccurately, is the specified use of results defensible?

6.3.2 This perspective starts by acknowledging that—within any assessment system—there will be a proportion of students who get assessed incorrectly and, consequently, for whom incorrect decisions will be made (be those selection decisions, provision eligibility decisions, placement decisions, and so on). It then proposes that—even if the system is just as far as most students are concerned—if it is sufficiently unjust for a sufficiently high number of students, then the system may have to be judged morally indefensible. This is analogous to why many countries refrain from executing serial murderers. It's not that execution, per se, is necessarily judged to be morally indefensible; it's the risk of executing even a small number of innocent people. So the assessment parallel might be:

- when the stakes are low for students—as is often true of everyday formative assessment—it would not matter too much if it were fairly error-prone (such errors can often be identified quickly through ongoing dialogue)
- but when the stakes are high for students—as when examination results are used for selection—it would matter (such errors can often negatively affect life chances time and time again).

6.3.3 This results in a utilitarian analysis (emphasising the minimisation of "horror" more than the maximisation of "utility") for which two kinds of evidence need to be taken into account:

- technical judgements—concerning the amount of inaccuracy that might be expected (stemming from the analysis of technical accuracy)
- value judgements—concerning the severity of negative consequences for those students who are assessed incorrectly.

6.3.4 This final point raises a fundamental question for the evaluator: whose value judgements ought to be taken into account in this analysis, and how? The answer is far from clear, especially since different stakeholders (eg, politicians, students, evaluators) are likely to have different values.

6.4 Social defensibility

6.4.1 The third evaluation perspective is social defensibility. It poses the question: is the trade-off between the positive and negative impacts from operating the assessment system sufficiently positive?

6.4.2 On the one hand, there will inevitably be a range of intended positive outcomes. In particular, the assessment results ought to empower users to make important educational and social decisions appropriately (such as selection decisions, placement decisions, school choices, and so on); to enable society to function more fairly and effectively than it otherwise would. Although, in theory, it may be judged entirely possible to draw sufficiently accurate inferences to support a range of important decisions; in practice, that doesn't guarantee that users will actually do so. So this needs to be investigated empirically. In addition, features of the assessment system itself may well be designed to facilitate important educational and social impacts (such as the improved attainment of students when assessed through modular rather than linear schemes) and these impacts need to be investigated as well.

6.4.3 On the other hand, there will inevitably also be a range of unintended, and possibly unanticipated, negative outcomes. In particular, features of the assessment system which appear to be innocuous may turn out not to be so. Consider, for example, standards-referenced systems, which employ a single scale to report absolute level of attainment at key stages of an educational experience that may span many years (eg, the national curriculum assessment system). The theory is that this should be motivating for even the lowest-attaining students; since it enables them to see that they are making progress as time goes by.[vii] However, such systems could conceivably turn out to be demotivating for precisely this group of students. Not only does the assessment reveal them to have attained lower than their peers at each key stage; they also see the gap between themselves and others widen on each assessment occasion.

6.4.4 The evaluator needs to be careful to distinguish those impacts which relate to the legitimacy of the assessment system, *per se*, and those which relate primarily to broader evaluation questions; for example, those concerning the legitimacy of educational or social policies or practices. School choice, for example, (even if based upon entirely accurate inferences concerning the general quality of teaching at a school) could conceivably result in a more socially divided society, which might be judged to be a bad thing. These are obviously important issues to be evaluated. However, they are issues for an evaluation of the policy of school choice, rather than for an evaluation of the assessment system which enables it. In practice, it is actually quite complicated to judge which impacts bear primarily upon the legitimacy of an assessment system and which relate primarily to broader evaluation questions; but it is useful to recognise the distinction and to try to work towards separation where possible. A rough rule-of-thumb might be: would we expect a different kind of impact if an alternative assessment system was in operation? If so, then the impact probably ought to be considered. If not, then the impact is probably attributable primarily to a broader policy or practice and, therefore, probably ought not to be considered. In the example above, relating to school choice, the "divided society" impact might be expected to occur regardless of the system used to generate results; so this impact might therefore not be relevant to scrutinise during an evaluation into the legitimacy of the underlying assessment system.

6.4.5 As with the moral defensibility perspective, the social defensibility perspective requires that two kinds of evidence be taken into account:

- empirical evidence—concerning the nature and prevalence of relevant intended and unintended impacts; and
- value judgements—concerning the costs of the negative impacts and the benefits of the positive impacts.

The synthesis of this evidence is based upon the utilitarian principle that: if, on balance, there appears to be too little benefit, for too few individuals, then the system may have to be judged socially indefensible.

6.5 Legal acceptability

6.5.1 The legal acceptability perspective asks: can the assessment system be operated without contravening the law?

6.5.2 This is becoming increasingly salient, both nationally and internationally. In England, the 1995 Disabilities Discrimination Act introduced legal rights for people with disabilities covering employment, access to services, education, transport and housing. The 2005 version of the Act included a new chapter which specifically covered qualification bodies; a provision which is intended to be extended to general qualifications from 1 September 2007.

6.5.3 The new legislation raises questions such as whether it is legally acceptable, within high-stakes general qualifications like GCSE English, to require specific forms of competence. For example, to be competent in English, is it absolutely necessary to be able to speak and listen fluently? Might there be a legal right for speaking- and hearing-impaired students to access this crucial "gatekeeper" qualification? Nowadays, we routinely need to consider whether our systems can be designed to be more inclusive without unduly compromising them.

6.5.4 The analytical bases for evaluation, from this perspective, are the principles and precedents of law; the basic premise being that contravention of the law is unacceptable. Significantly, judgements from the legal acceptability perspective can, and sometimes will, contradict judgements from the perspective of technical accuracy. Indeed, it may occasionally be necessary to make an assessment less valid in order for it to comply with the law. This is because technical analyses typically elevate the majority (sometimes at the expense of minorities) while legal analyses often elevate minorities (sometimes at the expense of the majority).[viii] Legal experts and assessment experts do not always share the same concept of fairness.

6.6 Economic manageability

6.6.1 From the perspective of economic manageability, the evaluator asks: is the burden of the assessment system upon society manageable?

6.6.2 The idea of burden does not reduce simply to financial cost, but also extends to issues such as: human resources (eg, the availability of skilled examiners); workload (eg, the time spent by students and teachers in preparing coursework); processing infrastructure (eg, the demands made of the postal system when delivering scripts); and even ecological impact (eg, the "rainforest cost" of the paper which flows through the system each year).

6.6.3 The analytic basis for answering this kind of evaluation question is economic, grounded in the principles that: all other things being equal, less expense and consumption is better; and that there will be a threshold of expense and consumption which cannot reasonably be exceeded.

6.7 Political viability

6.7.1 The final perspective is political viability, which poses the question: is society prepared to buy into the assessment system?

6.7.2 Clearly, if society is not prepared to buy into the system then—no matter how good it might seem to be from the other perspectives—it will remain unviable. Unfortunately, such failures are not uncommon in the world of educational assessment. In England, the following might be mooted as examples: S papers; Records of Achievement; the Certificate of Pre-Vocational Education; parity of esteem between academic and vocational qualifications.

6.7.3 Unlike the other perspectives, the underlying principle here is essentially arational. It is best illustrated by platitudes of folk psychology such as: the customer is always right; or, you can lead a horse to water but you can't make it drink.

7. *How should we conduct system-level evaluation?*

7.1 Turning the framework for system-level evaluation into a real-life evaluation is far from straightforward. As noted, each discrete use of results ought to be evaluated, independently, from each of the six perspectives. This clearly implies a very large amount of research; and the more uses to which results are put, the more research is required. Moreover, it ought not to be restricted to the intended or "official" uses either, since the unintended uses and even the proscribed ones are important too.

7.2 The example of national curriculum testing is useful here. Certainly, test results are not used for licensing nor for the certification of higher professional skills. Diagnosis and provision eligibility probably require results from more specialist tests; while selection and qualification would typically be based upon results from exams taken later in an educational career. Whether test results have (or ought to have) a role in guidance and national accounting is less clear. What does seem likely, though, is that results from national curriculum tests are used for the remaining 14 purposes, whether legitimately or not. Again, the question of legitimacy would need to be explored independently for each use.

7.3 At some point, evidence and argument from the independent analysis of specific uses, and specific impacts, needs to be brought together into an overall evaluation argument. This will require judgements concerning the acceptability of compromises and trade-offs, with reasoning along the lines of: "the system may not be particularly good for this use, but it's probably better than nothing; admittedly it does have a big negative impact for a small number of students, but perhaps not too many; and, ultimately, the system is quite good for that purpose, and that's the principal purpose, after all . . ." (obviously, this is simply a caricatured microcosm of an overall evaluation argument).

7.4 The previous paragraph hints at another important point: to reach overall evaluation conclusions, it is necessary somehow to weight the importance of alternative uses and impacts. There needs to be some indication of which are the most valued uses of results, and impacts of system operation, and which are more like fringe benefits. This might be a problem if there is neither any consensus among stakeholders nor formal specification from policy makers. Again, whose value judgements ought to be taken into account in this analysis, and how?

7.5 Ultimately, the legitimacy of the assessment system cannot be judged in isolation, but only in relation to:

1. a new-improved version of the same assessment system; or
2. an entirely different assessment system; or
3. a suite of more tailored systems, operating in parallel; or
4. no assessment system at all (which, admittedly, would be unlikely ever to triumph as an evaluation conclusion, but which is still useful as an anchor point).

7.6 This raises yet another complication: that each alternative will need to be put through the evaluation mill in its own right. That is, an overall evaluation argument will need to be constructed for each of the alternatives, to pit them against the present state of affairs. Unfortunately, since these are likely to be largely hypothetical at this stage, the construction of evidence and argument will inevitably be patchy and indirect.

7.7 Finally, it is worth emphasising that the aspiration of system evaluation is not perfection, but legitimacy; and this is also true for the flip side of evaluation, design. This legitimacy is a real-world, pragmatic aspiration, which might be characterised as: "overall, at least satisfactory, and preferably good, but inevitably not perfect". So, for example, whereas the principle of maximising validity (from a technical accuracy perspective) might go so far as to recommend a separate system for each discrete use of results, the principle of minimising burden (from an economic manageability perspective) might recommend just one. The overall evaluation conclusion (bearing in mind all perspectives) might recommend something in-between; say, two or three separate systems, operating in parallel, each supporting a distinct set of three or four different uses of results, and each with its own particular impacts. Compromise and trade-off are fundamental to the design and evaluation of assessment systems.

8. *Is system-level evaluation feasible?*

8.1 Given all of the above, is it humanly possible to undertake a rational and rigorous system-level evaluation? Is it possible to reach a straightforward conclusion to a straightforward question like "should the system of national tests be changed?" This is a challenging issue. It's probably true to say that no educational assessment system has ever been evaluated quite as rigorously as recommended above. Indeed, it's an inevitability of real life that decisions are generally made in the absence of complete evidence and argument; and the world of educational assessment is no different in that respect. Having said that, the inevitability of falling short of the ideal evaluation does not detract from the importance of constructing as rigorous an evaluation as is possible.

8.2 Frameworks like the one presented above can help inquirers to scaffold useful answers to thorny evaluation questions. They can be particularly helpful for identifying holes in the overall evaluation argument: where research still needs to be undertaken; and where argument still needs to be constructed. And they can also help stakeholders to reflect upon, to clarify and to articulate their different priorities for national assessment; to distinguish between the crucial uses and impacts and those which are more like fringe benefits.

9. *Using the framework to identify common limitations*

9.1 Finally, frameworks like the one presented above can also help inquirers to identify limitations in evaluation arguments presented to them by others. In this last section, a few common limitations will be illustrated.

9.2 The conflation of different evaluation questions

9.2.1 In constructing a robust evaluation argument, it is important to put to one side issues which appear to be relevant, but which actually fall under a broader evaluation remit. For example, when evaluating the use of test results for school choice purposes, it is clearly relevant whether the system supports sufficiently accurate inferences concerning differences in the general quality of teaching between institutions. However, as suggested earlier, the positive and negative impacts arising from school choice, per se, are probably not directly relevant and, as such, should not enter into the evaluation argument.[ix] Of course, they are crucial to evaluating the policy of school choice, and this evaluation needs to happen independently.

9.2.2 A particularly common limitation of many formal and informal evaluation arguments is the failure to distinguish between the impacts attributable to testing, per se, and the impacts attributable to the high-stakes uses of results which the testing is designed to support. So, for example, to the extent that high stakes can trigger behaviour which corrupts the validity of test results and the effectiveness of teaching, high stakes can similarly trigger behaviour which corrupts the validity of teacher assessment results and the effectiveness of teaching. In short, it may not be the operation of the assessment system, per se, which is problematic, but

the policies or culture underlying those high-stakes uses. Having said that, there are important differences in this situation from the one described above. First, although the impacts might be primarily attributable to the high-stakes uses, they directly affect the accuracy of results from the system; which thereby renders those impacts directly relevant to the evaluation. Second, the impacts upon teaching and learning, even if primarily attributable to the high-stakes uses, are likely to be different across different assessment systems; which again recommends that they enter into the evaluation.

9.2.3 When there is a range of equally valid, although logically separable, evaluation questions to ask, then these ought somehow to be arranged within a meta-framework. For example, it makes sense to interrogate the purposes of curriculum and qualification, before interrogating the purposes of assessment. Or, to put it less formally: the assessment-tail should not wag the curriculum-dog. At least, not too much; where the rider 'too much' is essential. In fact, the process of meta-evaluation needs to be iterative and will necessitate inevitable trade-offs and compromises. By way of extreme example, it would not be legitimate to promulgate a radically new curriculum for school-leaving examinations, if the learning outcomes which were elevated could not be assessed with sufficient accuracy: we need to remember that the examination results have important functions in their own right, as the basis for making the kind of qualification and selection decisions that are necessary to support a fair society.

9.3 The lack of a specified alternative

9.3.1 A common limitation of evaluation arguments is the lack of a specified alternative system. It is not foreseeable that society would tolerate the rejection of educational assessment entirely. So it would seem to be incumbent upon any critic of present arrangements to explain, in some detail, how an alternative system would, on balance, be more legitimate. The key issue, here, is one of detail. For instance, the "test versus teacher assessment" debate is literally meaningless unless the detail of the alternative system is spelled out.

9.4 Too incomplete an analysis of uses and impacts

9.4.1 Even when two or more systems are specified in sufficient detail, and pitted against each other, it is often the case that the evaluation argument remains incomplete, through omission of central components. This frequently occurs when an alternative system is proposed which is particularly effective in relation to certain uses and impacts—perhaps genuinely more so than the present system—but which leaves crucial other uses or impacts unaddressed. In England, numerous protagonists have argued recently for employing moderated teacher assessment (for certain uses and impacts) alongside a national monitoring unit (for others); instead of the present system of national curriculum testing. Few protagonists, though, have also grappled effectively with how best to support uses which require the comparison of teachers and schools in a high-stakes context. This particular debate is very important—because the arguments in favour of certain forms of teacher assessment alongside a national monitoring unit are persuasive. However, due attention also needs to be paid to satisfying the demand for trustworthy data on school effectiveness.

9.4.2 Another limitation of many evaluation arguments is the lack of available evidence, or a reliance upon evidence which is easy to challenge. A particular example of this at present is the impact of national curriculum testing upon teaching and learning, especially at Key Stage 2. Despite the system having been in operation for over a decade, and despite considerable anecdotal evidence of negative washback, there is remarkably little systematically documented evidence. This greatly hinders effective evaluation.

9.5 The gulf between real and hypothetical

9.5.1 Finally, while extant systems must inevitably be evaluated in the context of real-world operation—mired in the kind of intricate relationships which give rise to unforeseen problems—alternative systems will typically be evaluated as promising-hypothetical. In this context, it is easy to give the alternative system undue benefit of the doubt, without recognising that its implementation will inevitably necessitate certain compromises and will result in its own unforeseen problems. At the very least, the root cause of the problems which beset the extant system need to be extrapolated to the promising-hypothetical.

Endnotes

i The term "assessment" is used generically, to refer to any instrument or process through which student competence or attainment is evaluated (eg, test, teacher assessment, examination, etc). The term "system" is used to encapsulate, in a broader sense, the detail of the structure and mechanism through which students are assessed. In relation to national curriculum testing, for instance, this detail would include procedures

for test development, distribution, administration, marking, reporting, evaluating (and so on), as well as the technical, professional, managerial and administrative employees required to develop and operate those procedures.

ii It is based upon insights from the international literature on validation and evaluation, although references to specific sources have not been included (further information can be provided on request).

iii Although there is a huge debate in the technical literature on the precise extension of the term "validation", this does not significantly affect the tenor of the argument developed in this report.

iv In reality, it would be advisable to use more than one source of evidence to support important decisions (such as placement, monitoring, guidance and so on). Indeed, assessment professionals are increasingly preaching this dictum. However, that does not change the basic principle that, when results are used to support different purposes—whether alone or in combination with other sources of evidence—different kinds of inference are drawn from them.

v There are different ways of emphasising the point that results which are fit for one purpose may not be fit for another. The approach in the text is to focus upon the different inferences which need to be drawn from results. Another approach would be to stress that systems need to be designed differently for different purposes and different design compromises will be made. (Compromises are made so as not to over-engineer the system, because increased accuracy comes at a price; assessment design aspires to sufficient accuracy, for a specific purpose, rather than maximum accuracy.) Ultimately, design characteristics and compromises which are legitimate for one use may be illegitimate for another.

vi Note that this is not to implicate the phenomenon of "teaching-the-test" (whereby, over time, teachers reduce the scope of their teaching, excluding those aspects of the curriculum that the tests tend not to cover). This practice is neither appropriate nor inevitable. Were it to occur, it would occur in addition to the impact of practice, training and improved resources (described in the text).

vii This contrast with norm- or cohort-referenced systems, in which the lowest-attaining students may be awarded the same very low rank at every stage of their educational career, despite making real progress in learning and despite achieving respectably given their particular situations.

viii Any technical analysis which is based upon an average (which is frequently the case for large-scale educational assessments) thereby tends to elevate the majority.

ix Other than when considering the negative impacts which arise from inappropriate school choices, consequent upon inaccurate results data (the moral defensibility perspective).

Paul E Newton
Head of Assessment Research, Regulation and Standards Division, QCA

June 2007

Paper 2

NATIONAL CURRICULUM TESTS

Building Effective Assessment

Analysis and Monitoring of Current Practice and its Impact in England

1. *Current national curriculum assessment arrangements*

1.1 Very close contact with schools is essential for QCA to carry out its work effectively. QCA regularly monitors the nature and development of curriculum and assessment in schools in England. The outcomes of monitoring are reported annually, and used to shape directions for the future. The key element of this contact is regular and frequent reference on a broad range of issues to a network of more than 1,000 schools that work in partnership with us, and represent a cross-section of school types nationally. The following observations are drawn by QCA on the basis of systematic and formal consultation with this network.

- The national curriculum tests do a very good job in doing what they are designed to do—measure pupils' performance in reading, writing, mathematics and science. They stand comparison with any similar large-scale assessment tool across the world for reliability and validity.
- The tests provide an objective, nationally comparable snapshot of pupil performance in key areas of learning on an annual basis, and have been the foundation of statistical analyses of pupil and school performance over a number of years.
- The design of the tests encourages teachers to cover a broad curriculum within the areas being tested (for example, requiring pupils to respond to a range of different written texts to assess their reading; to produce two different kinds of writing without any prior notice of the form, purpose or audience; and to apply mathematical and scientific skills and knowledge).

— At Key Stages 2 and 3, large groups of markers have been trained, many of whom are practising teachers. Marking refines their understanding of national standards, and the insights gained from this experience are taken back into their schools and classrooms to further improve curriculum, teaching and assessment.

— At Key Stage 1, where since 2004 the tests have been used more flexibly to inform teachers' overall assessment of pupils at the age of seven, there has been a stronger and very beneficial focus on teachers' ongoing observations and on assessment directly influencing future planning, teaching and learning.

— The high profile of the tests has focused the attention of schools on maximising pupil attainment at key points in their educational progress. Evidence confirms that they have contributed to a significant rise in pupil attainment over the last 15 years.

— Like any tests, however well designed, they can measure only a relatively narrow range of achievement in certain subjects on a single occasion and they cannot adequately cover some key aspects of learning.

— The focus on the core subjects leads to comparative neglect of the full range of the national curriculum. Ninety per cent of primary and 79% of secondary schools report that the testing has led to pupils being offered a narrower curriculum.

— Although both teacher assessment and test outcomes are reported at Key Stages 2 and 3 it is the test results which are given greater public attention and which form the basis for judgements about school performance and effectiveness.

— Most schools prepare pupils extensively before they undertake the tests. To prepare for the Key Stage 2 tests, 68% of primary schools employ additional staff, 78% set additional homework, and more than 80% have revision classes and use commercial or QCA practice tests. In 80% of primary schools, the amount of time spent on test preparation has increased over the past 10 years, and in the second half of the spring term 70% of schools spend more than three hours per week on test preparation. There is a similar pattern of responses from secondary schools in terms of time spent in preparing for the tests.

— Ofsted reports that schools often deploy their most effective teachers in the particular year groups at the end of a key stage (years 2, 6 and 9), and that teachers in other year groups feel less responsibility for assessing pupils' progress.

— Investment needs to continue to be made into strengthening teachers' ongoing assessment skills. With an increasing focus on personalised learning and monitoring individual pupil progress, teachers' professional judgements about the achievements of their pupils are the most fruitful source of information when identifying targets for improvement and providing feedback for pupils and their parents/carers.

— Schools' perceptions of the accuracy of teacher assessment and national curriculum tests vary between primary and secondary schools. At Key Stage 2, 64% believe that teacher assessments are more accurate than tests and 9% say that teachers' judgments need to be supported by test results. At Key Stage 3, the figures are 37% favouring teacher assessments, with a further 41% believing that teacher assessments are as accurate as tests. Twenty per cent see value in tests to support teacher judgement.

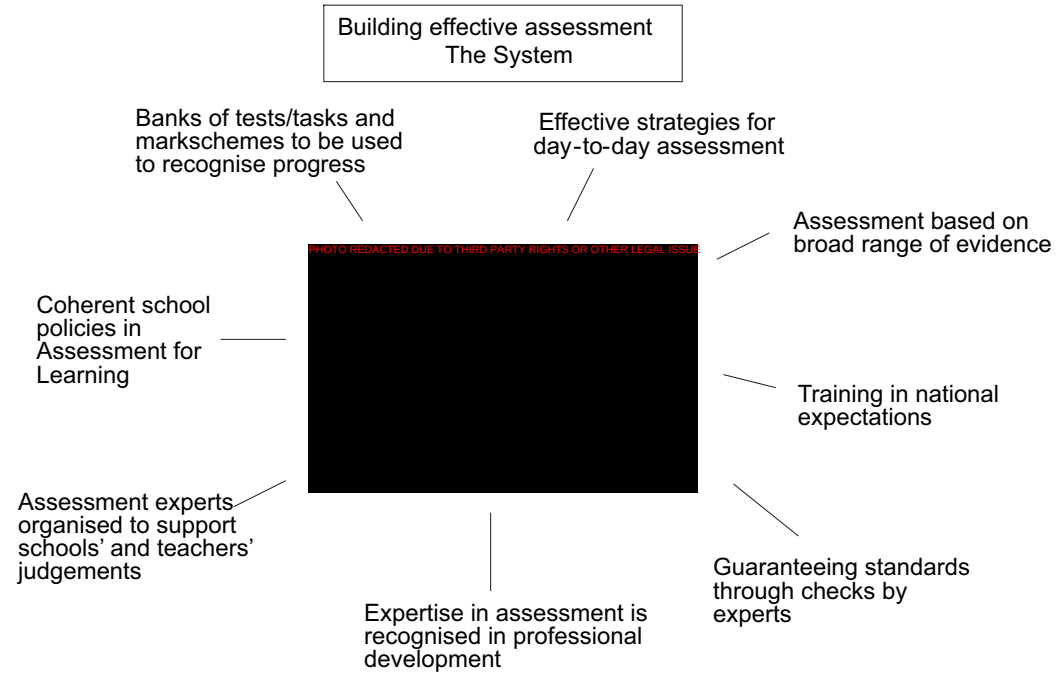
— Schools report that they often mistrust the results from the previous key stage and re-test using different measures.

2. *Building effective assessment*

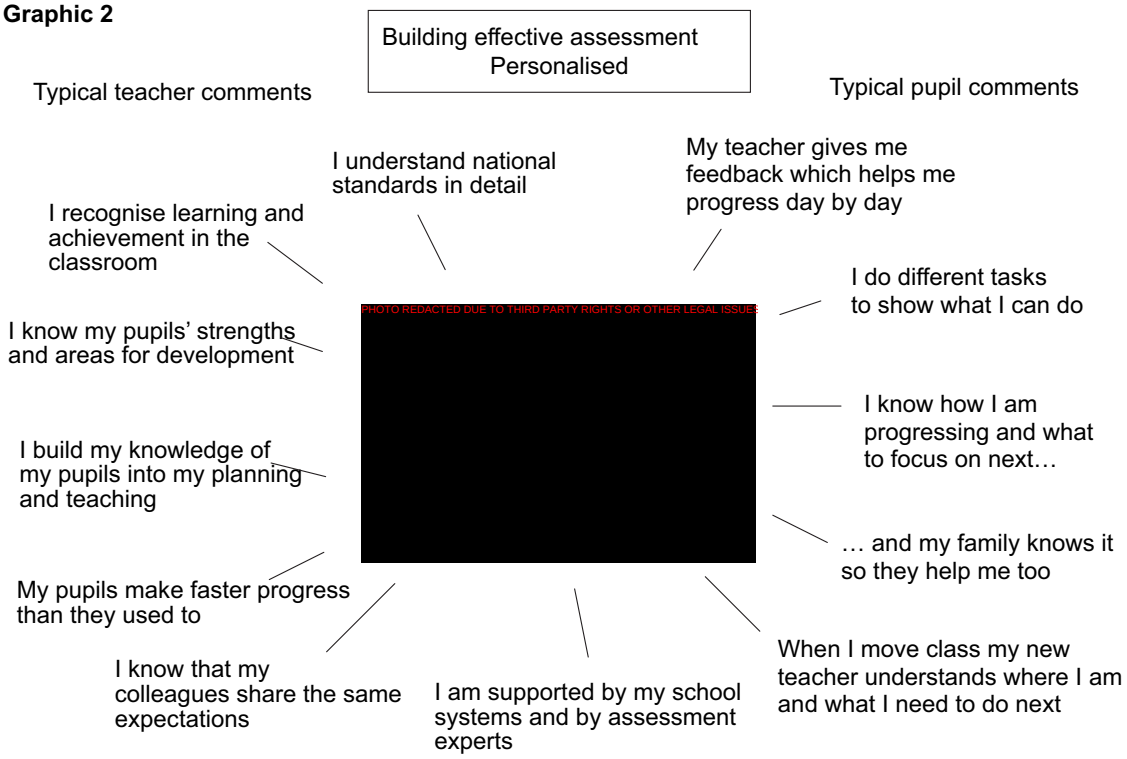
2.1 The strengths of the current arrangements provide a sound foundation on which to build. The following graphics show the directions of travel in supporting teachers and schools (Graphic 1), personalising assessment (Graphic 2) and making assessment more effective (Graphic 3).

June 2007

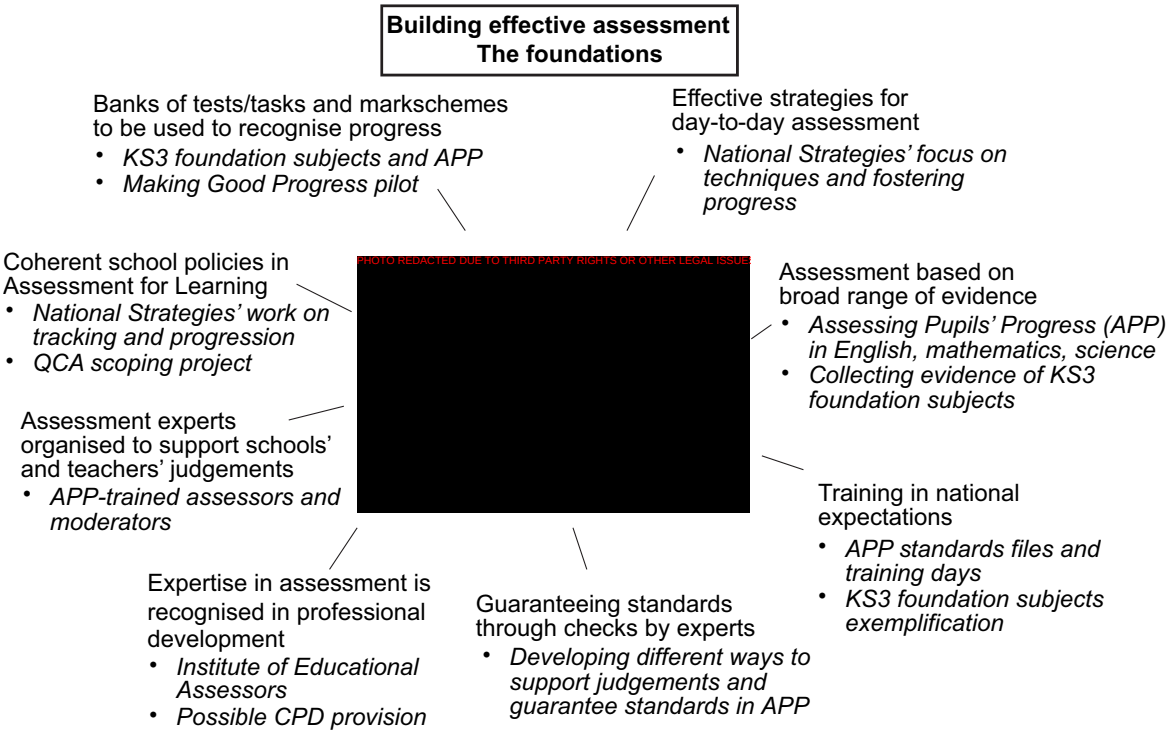
Graphic 1



Graphic 2



Graphic 3



Witness: **Dr Ken Boston,** Chief Executive, Qualifications and Curriculum Authority (QCA), gave evidence.

Q55 Chairman: I welcome you, Dr Ken Boston, to our deliberations. It is the first time that you have appeared before this Committee—we saw you in a previous Committee on a reasonably regular basis. It was good of you to come here at short notice, given that people—certainly those in Parliament—are close to the time when they disappear from London for their Christmas break. You were good enough to enable us to keep the momentum of our inquiry this side of Christmas, so that we can reach a conclusion early in the new year. We appreciate your taking the trouble to do that. This is an historic day for testing and assessment, although we did not plan it that way. We usually give witnesses a chance to say something at the start, after which we ask questions. Would you like to make a brief statement?

Dr Boston: I should like to take a couple of minutes to make a statement. Thank you for giving me the opportunity to give evidence to the Select Committee. I shall give a brief preface on standards and national performance. In its regulatory capacity, it is the job of the Qualifications and Curriculum Authority to ensure that assessment standards are maintained year on year for national curriculum tests, GCSEs, GCEs and other qualifications. The assessment standard is the height of the hurdle that is to be jumped in any examination or test—it is the degree of difficulty. Our regulatory task and the task of our division, the National Assessment Agency, which delivers the national curriculum tests, and the task of the awarding bodies, which deliver the general qualifications, is to keep the hurdle at the same height year on year. The performance standard is different. It is the number of students who clear the hurdle in a particular year. When we say that standards are rising—as they are—we mean that increasing numbers are clearing the hurdle. I make that point at the start because the two uses of the word "standards" are critically important and have been the source of much confusion. In areas other than regulation—the areas of curriculum, assessment and qualifications development—our role is to work with the Government to drive up performance standards and increase the number of those who clear the various hurdles. We are partners with the Government and other bodies in the national enterprise of raising performance standards overall. The QCA has been absolutely scrupulous in ensuring that our regulatory decisions are not influenced by political considerations. In my time in the job, at least, Ministers and civil servants have been similarly principled in ensuring that they remain totally disengaged from the QCA's regulatory functions. However, there has always been a logical inconsistency in the body accountable for maintaining assessment standards reporting to Ministers whose job is to drive up performance

standards. The Government's decision announced this morning to establish a new body from within the QCA to take over its regulatory responsibilities and report to Parliament, not Ministers, will resolve that difficulty and is therefore very welcome. At the same time, it will allow the QCA to become, in due course, a new organisation to focus on the role of curriculum and assessment, and qualifications, in raising national performance standards. I would like to say a couple of words about national performance standards and how to drive them up. Performance standards are rising, but in England, as in school systems across much of the western world, the rate of improvement in educational performance has slowed in recent years. If you look at the graph of our performance and those of many other western nations, you will see that the lines are not moving up as steeply as they were a few years ago. In some counties, the graph has virtually reached a plateau. There seems to be, internationally, a glass ceiling at about the 80% competence level: that is, at the level at which about eight in every 10 young people reach the agreed national bench marks, such as Level 4 at Key Stage 2. However, we are by no means unique. Fullan, Hill and others have shown that the conditions for breaking through that glass ceiling already exist and the difficulty here and elsewhere has not been in finding what to do, but in bringing together in the country's classrooms the things that need to be done. There are three approaches to teaching and learning that, if brought together effectively within classrooms, will cause individual, school and national performances to move upwards more sharply, with national performance standards potentially rising to the 90% competence level and perhaps above that. The first of those is personalised learning, which is a term that I quite dislike, because it is commonly characterised as putting the learner in charge of the learning, with all the implications of the secret garden of curriculum that we have heard in the past, without the edge of challenge and discipline in grappling with difficulty, which are fundamental to all real learning. Personalised learning is better described as highly focused teaching, where the teacher is firmly in charge of the process of instruction and designs it to stretch the individual beyond the level of what we might call the comfort zone. There is an educational theory of 30 years' standing underpinning that, which focuses on drawing the learner into new areas of learning that are beyond his reach at that point, but which, with effort and application, are achievable. As I have said, there is ample evidence over the past 30 years to show that that works. Personalised learning is deeply rooted in curriculum, but requires a three-dimensional curriculum that has depth, rather than a two-dimensional curriculum. It should be a deep, rich resource from which a teacher can draw bespoke material to take each young person to their next level of knowledge, skill and understanding. The second component is systematic and precise measurement in the classroom of the current stage of learning to enable the teacher to shape the next stage for each child. If personalised learning is to drive up performance at individual, school and national levels, it needs to stand on a foundation of frequent, low-stakes assessment of individual performance. That testing needs to happen routinely and incidentally within the classroom as a matter of course. Some of it can be supported by technology, such as a child taking a 10-minute task on a computer to prove for himself and the teacher whether he has yet mastered, for example, percentages and can be challenged with something more demanding, or whether more work on percentages is needed to make him secure. We need to enable teachers to use more of that sort of assessment in schools. There is an immense professional thirst for it and, because youngsters come to see frequent and incidental assessment as integral to their learning and as hurdles to train for and take pleasure in leaping, in that sense they do take charge of their own learning. The third and final component is professional learning for teachers to enable them to assess teacher performance better and to use the assessment information on each student to design and implement personalised instruction. Teachers need to be able to convert the formative assessment data into information that will enable them to make instructional decisions not at some time in the future—nor at the start of next year or at the end of the key stage—but tomorrow. That is when decisions on intervention need to be implemented. In England, significant progress has been made on each of those three essential prerequisites, achieving further improvement in school and system performance by bringing them together in classrooms. The new secondary curriculum has been designed to support highly focused teaching in the sense that I have described. That will also be an objective of the forthcoming review of the primary curriculum and of our work with Sir Jim Rose in the context of the broad view of the curriculum in the Children's Plan. The Children's Plan puts £1.2 billion into supporting the personalisation of learning over the next three years. The pilot single-level tests are also a significant step forward in providing information that has the additional potential to provide summative data on school and system performance. The tests represent a substantial investment in addition to the current, Key Stage tests, which they are expected to replace in due course. There are also, of course, growing data banks of test items produced by the QCA at the request of Government, such as the Key Stage 3 ICT test, and other assessment instruments developed by the private sector, which will support assessment of separate components for programmes of study. The assessment of pupil performance programme, which is now being rolled out nationally in both primary and secondary schools, goes to the heart of the teachers' professional learning in making instructional decisions based on assessment information. The Government is committing £150 million over the next three years for the development of staff in assessment for learning. To conclude those initial remarks, let me say that at the moment I am pretty optimistic about the future. There seems to be a willingness across Government, the teaching profession and the broader public to engage in

genuine discussion about the future of testing and assessment and to come out of the trenches to some extent. There seems also to be a real recognition of the importance of three things—personalised learning, formative assessment, and professional development for teachers—which are the essential keys to raising performance standards and the only way in which this country will drive itself through the glass ceiling at around 80%.

Q56 Chairman: Thank you for that introduction, which was a pretty thorough look at the whole field. If we are going to get through all our questions in the time available, the question and answers will have to be quick-fire. I want to start by asking why all that was necessary? You gave evidence to the Committee not very long ago, when you seemed to be an extremely happy chairman of the QCA. You did not say to us that there is a fundamental problem with the QCA structure and that if only the Government would listen there should be some fundamental changes. Nevertheless, fundamental changes are what we have here. Some of us who know the history and the origins of the changes, over the past 10 or 15 years, feel that we have kind of been here before. Why do you think that the changes have come about now?
Dr Boston: Our private, but consistent, advice to Government has been that there is a perception that the regulatory decisions could be manipulated by Government, given the way in which we report to Ministers rather than to Parliament. That argument is strong, and we have made it again and again. The Government have accepted the argument in so far as it relates to the regulatory side of our work. The other side of our work will continue much as it is. I believe that that is a step forward.

Q57 Chairman: Do you understand that the regulatory part will be in parallel to what has been established as the relationship of Ofsted to Parliament?
Dr Boston: I am not precisely sure what the governance arrangements will be, except that it will have its own board, its own chairman and its own chief executive—I do not think that anyone is sure yet and lawyers are looking at the matter. The issue of whether it is a non-ministerial department, or reports to Parliament in some other way, still needs to be worked through as part of the consultation process.

Q58 Chairman: When it was believed that Ofsted was responsible to and answerable to Parliament, there was a hard-fought battle to ensure that it did so through this Committee, or its predecessor Committee.
Dr Boston: Yes.

Q59 Chairman: So, I assume that constitutionally, the parliamentary relationship will be mediated through a Select Committee.
Dr Boston: That would be my assumption, but those matters are being considered within the Department, not the QCA.

Q60 Chairman: In broad terms, do you think that this morning's proposals are to be welcomed?
Dr Boston: Yes.

Q61 Chairman: In their entirety—there is no hesitation, qualification? I won't say the Australian equivalent of welcome, but you know what I mean.
Dr Boston: With a modest, restrained British approach to things, Mr Chairman, yes, these proposals are to be welcomed.

Q62 Chairman: Let us drill down a little. In this Committee, and the previous one, we did not see great public demand for these changes. Do you believe that the public were knocking on people's doors—they were certainly not knocking on my door—saying that they wanted a more independent relationship? Or is it that they were worried about standards? There was always a fuss in August when the results came out—the *Daily Mail* would always tell us that standards were going down and that there was grade inflation and much else. Is that what people are responding to? Is that what the Government have responded to—the furore that goes on in August?
Dr Boston: Certainly, the Government have listened to and heard our concerns about the ambiguity present where there is a body that, among other things, is responsible for regulation and reports on the maintenance of assessment standards to a Government who are committed to driving up standards to meet particular targets. As I said, in reality, we have not been troubled by this. I do not think that anyone could point to an occasion when pressure has been put on the organisation by the Government or civil servants with regard to standards— certainly, I am totally unaware of it, and I am certain that it has never happened. However, if we consider one of the causes of the August debate to be that the separation of the regulator from Government is not perfectly clear, then that August debate might be diminished if the separation were made more apparent. Of course, there may be other issues in the August debate that are not resolved by that situation.

Q63 Chairman: As you know, August is a slow news time. They always bring the education correspondents back for August, so if they have to write about something, I am sure that they will do so. What is your view of the balance between the agency and the other body? How will it be handled, and how will the two organisations develop?
Dr Boston: The Secretary of State has asked us to set up an interim regulatory authority. That should be done virtually immediately, and there should be as much distance between the regulatory body and the parent body—the QCA—as is possible by the summer examinations. Of course, the legislation will not be passed and take effect until 2009.The way we are looking at setting up the interim arrangements is for the QCA board, which cannot be discharged of its regulatory responsibilities without a change in the Act, nevertheless carrying out those responsibilities, not through me as Chief Executive, but through

Isabel Nisbet, the Head of Regulation and Standards, who, it has been announced today, will be the Acting Chief Executive of the new regulatory authority—Ofqual, or whatever shorthand we might finally use to describe it. That organisation will be operating in shadow form from April. I will not be dealing personally with the awarding body chiefs on matters of standards and I will not be setting levels in relation to national curriculum tests, as I do at the moment. That will be done by David Gee as head of the NAA. I will be responsible for managing the affairs of the board. I will remain the accounting officer for the entire organisation, but the shadow regulator's funds will be ring-fenced. An interim board with an interim chairman will be established for the shadow regulator, and the proposal is that, to all intents and purposes, it should function as a separate body from about April. Not only will it function separately, but it will do so from Coventry, because many of them would otherwise be moving to our temporary premises in the old Adult Learning Inspectorate.

Q64 Chairman: We must get on to the last thing. We have dipped our toe into the area of testing and assessment. We have already had a lot of written evidence and we have had a seminar. People mostly wanted to talk about, not the constitutional role of the two organisations, or the split between the roles of the organisations—that was hardly mentioned—but too much testing, grade inflation, and a range of things that concern parents, students and commentators. It seems that this is to take our eye off the ball, so that we can say, "Look, this is all alright. We are making some big, grand, but complex changes out there," whereas most parents and students are worried about other things entirely, such as too much testing. Everywhere in the world they say that there are too many tests. Academics come before us and tell us that we test the wrong things or too many things. Those are the real issues, are they not?
Dr Boston: Yes, they are. Certainly, during the interim period, we will not be taking our eyes off those balls.
Chairman: Let us get drilling now with David.

Q65 Mr Chaytor: To pursue today's announcement a little further. What will it cost?
Dr Boston: I do not have an answer to that, but we will be meeting to establish the shadow regulatory authority for which we will need completely new front-of-house facilities. From April, if you ring the regulatory authority, you will not want someone from the QCA answering the phone. The media will need to be different, as will the presentation and delivery. We are looking at that, with a view to presenting a budget bid to the DCSF for putting it in place.

Q66 Mr Chaytor: Do you know at what stage your budget bid will be presented?
Dr Boston: It will be presented within the next few weeks; by early January.

Q67 Mr Chaytor: In your opening presentation, you put a lot of emphasis on the distinction between assessment standards and performance standards. In 1996, the QCA's predecessor and Ofsted published a report on assessment standards, saying that there had been no weakening in the previous 20 years. In 2007, can the QCA say that there has been no weakening in assessment standards in the previous 11 years?
Dr Boston: Yes. I would also have to say that being able to say that is the product of vigilance and monitoring. Of course, when looking at standards, which are made by humans, and evidence produced by full-cohort papers—a new, different paper each year—judgments have to be made about the way in which one paper and performance equates with previous papers and performance, and so on. Much of our work on maintenance of standards is looking back over a period of time. The reviews that we undertake of groups of subjects over a period of time indicate, from time to time, that in one area there might have been a drift, and that needs to be corrected. In a report earlier this year we looked at music, including elements of the music curriculum and music performance, and there appeared to have been a drift there over five years. That then needs to be corrected by altering criteria with awarding bodies. It is a process of monitoring, review and adjustment, but taken in balance as a whole—as an overview of the situation—my answer clearly and unambiguously is yes.

Q68 Mr Chaytor: But will today's announcement about the split of the QCA's functions in any way reduce the likelihood of drift in assessment standards over the next 10 or 20 years? Your argument seems to be that there has been some drift here and there, which is largely the inevitable result of human error and weakness of human judgment that has been corrected. But is there anything in the new structure that will stop that happening?
Dr Boston: No. The new body—the regulatory authority—will use codes of practice similar to those we have used in the past. It will use monitoring processes with awarding bodies. It may choose to extend its work beyond the work we fundamentally do, which is at the front end of the qualification, developing the criteria and then accrediting the qualification submitted to meet those criteria, and at the end of the process, after the examination is running, looking at whether the code of practice has been applied in the awarding process. As we move forward with regulation—since Isabel Nisbet[1] has been with the organisation, she has driven this very hard—we need to be regulating more on the basis of the assessment of risk and going into particular points through the process, rather than focusing initially at the start and, finally, at the end.

[1] Isabel Nisbet, Acting Chief Executive of the new interim regulatory body which will begin operations next year. Isabel is currently the Director of Regulation and Standards at QCA.

Q69 Mr Chaytor: But none of those issues could not be grasped by the QCA in its present format. Is not that the case?
Dr Boston: That is true.

Q70 Mr Chaytor: There is nothing about the new form of regulator that will give an enhanced guarantee of no reduction in assessment standards.
Dr Boston: It is precisely the same style.

Q71 Mr Chaytor: What I am trying to get at is this: is the conclusion, therefore, that the only argument for change is to somehow deal with the annual two-weeks-in-August hysteria in the tabloid press?
Dr Boston: Well, I would not describe it as dealing with the two weeks of hysteria, because while the basis for that might be diminished I am not sure that it is going to go away. The basis of the separation that is occurring is, as I see it, the logical one: a regulatory authority should not be reporting to the political party that is currently trying to drive up standards.

Q72 Mr Chaytor: In terms of structural change within the QCA, will the existing structure of the organisation adapt itself neatly to a division into the two new functions or will this require a major overhaul?
Dr Boston: No. This will require some major separation of the organisation. The regulation and standards division is clearly at the core of regulation, although not all that it does will go to the new regulatory authority. There are other elements in our curriculum division and in the qualifications and skills division, where regulatory work is done. The re-accreditation of A-levels, for example, which is essentially regulatory, is done through the qualifications division as a 14–19 qualification. We have to unpick those functions and make provision for that work to transfer to the regulator.

Q73 Mr Chaytor: Within the QCA as it stands, there are three main divisions. The structure of the organisation is based on three main areas.
Dr Boston: There are four: regulation, qualifications and skills, curriculum and the NAA, which is the operational arm that delivers the national curriculum tests and the modernisation agenda.

Q74 Mr Chaytor: In terms of assessment standards and performance, this is a blurring of these two functions across the four divisions.
Dr Boston: Yes, organisationally there is a bit of a blurring. This is meant to clarify it. Regulations and standards or Ofqual—or whatever we end up calling it in shorthand—sitting at Coventry, will be purely to do with assessment standards and nothing else.

Q75 Chairman: We have a QCA. You are the experts on the curriculum. The Government have just announced yet another inquiry into curriculum, not by you, but by Jim Rose. What is he doing being pulled into that? You are the competent body. You know more about this than Jim Rose. Why are you not doing it? I would be sulking if I were you.
Dr Boston: The intention announced by Government is that the inquiry will be led by Jim Rose, but that we will work with him as the chief source of advice on evidence and as the body organising and managing a consultation, which presumably will be very widespread. We need to take this out and get genuine consultation with the professionals.

Q76 Chairman: Have they appointed Jim Rose because he is more of a political fixer than you?
Dr Boston: I have no comment on that, Mr Chairman.

Q77 Chairman: Some of us on the predecessor Committee were not too keen on the Rose report. He went totally overboard on synthetic phonics, but we hope that he will do a better job with you on the curriculum.
Dr Boston: He is certainly a very valued member of our board, and I believe that we will be able to work together very effectively to achieve this. Finally, of course, it will be his advice that goes to the Government. There is no question about that, but we will provide the horsepower in shaping that advice and carrying out the consultation.

Q78 Ms Butler: We are all aiming for the same goal: to ensure that our children are very well educated. We also want to ensure that schools are properly evaluated. In your opinion, are there any other ways in which the effects of national policy on the state schooling system could be effectively evaluated? Do you have any ideas or opinions on how it could be further improved?
Dr Boston: I am not quite sure that I get the question. Do you mean methods other than the current assessment system?

Q79 Ms Butler: Other than the current system and how it works.
Dr Boston: That question takes us fundamentally to the issue of the fitness for purpose of assessments. What are we assessing and why? That is the area in which the paper that Paul Newton from the QCA prepared for the Select Committee is very helpful. The current Key Stage tests are absolutely fit for the purpose for which they were designed. That is full cohort testing in reading, writing, maths and science for our children at two points in their careers and for reporting on the levels of achievement. They are assessments that are developed over two and a quarter years, and are pre-tested. They are run through teacher panels, pre-tested again, and run through teacher panels again. The marks scheme is developed over a period of time. In terms of the way in which they are put together, if your purpose is full cohort testing, in these dimensions, these are the Rolls-Royce. You are not going to get better; they are fit for purpose. The issue arises with any assessment when, having achieved an assessment that is fit for one purpose, you strap other purposes on to it. As Paul's paper shows, there are 22 purposes currently being served by current assessments, and 14 of those are in some way being served by Key

Stage test assessments. Some of those purposes are very close to what is the design purpose, the essential function—the design inference, as Paul calls it. Some of the user inferences—the purposes to which they are put—are much more distant. One of the things that attracts me to the single level tests is that the Government are now looking at a new suite of tests that will have, not only the summative role—potentially when you add up what children have achieved at the end of the Key Stage, to get similar data to the summative data that you get now—but potentially a formative and development role because they are taken during the Key Stage test, and will potentially have less impact on preparation for the test because you are not preparing everyone to take the test at a particular time. You are building children up to take the test when they are ready. My judgment is that, given that there are so many legitimate purposes of testing, and Paul Newton lists 22, it would be absurd to have 22 different sorts of tests in our schools. However, one serving 14 purposes is stretching it too far. Three or four serving three or four purposes each might get the tests closer to what they were designed to do. To take a very simple analogy, Barry, if you want to cut paper or cloth, you have scissors; if you want to slice an apple up, you have a knife; if you want to turn a screw, you have a screwdriver; if you want to open a bottle, you have a corkscrew. To some extent, we are not building tests, we are building Swiss army knives here, and when you put all of these functions on one test, there is the risk that you do not perform any of those functions as perfectly as you might. What we need to do is not to batten on a whole lot of functions to a test, but restrict it to three or four prime functions that we believe are capable of delivering well.

Q80 Ms Butler: Do I take it that you believe that the Government's direction of travel in stage not age is the right direction to be travelling in?
Dr Boston: Yes. That is a very important step forward, and I think that the single level tests that are still in pilot stage have the prospect of combining both a summative assessment and a formative assessment. They are across a whole programme of study; it is not simply like my example of testing a youngster on percentages and moving along. It will provide us with progress data—summative data as they go through it—as well as formative data as they go through the key stage itself.
Chairman: Annette will lead us through the next section on the purpose of testing figures for purpose, which we have started, but we are going to continue.

Q81 Annette Brooke: Can we backtrack slightly and look at degrees of error, certainly in validation? We had the statistic in an earlier sitting that up to 30% of candidates in public examinations are awarded the wrong grade. We can touch on the issues of consistency of marking and actual mistakes in adding scores together, but what sort of percentage error are we looking at that is due simply to the nature of the design of the test? It may be that a student hits a whole set of questions and does not know the answer to those particular questions. In other words, what aspects are there other than the obvious mis-marking and adding-up errors?
Dr Boston: I cannot say that in any test there will be this percentage of error, but there are sources of error. The figure of 30%. is a very high figure, which I have heard before and it certainly pulls you up. What are the controls we have over this? We have the nature of the mark scheme and how precise and definitive that is, in terms of allocating scores. We have performance around grade boundaries, where a score might be just above or below a grade boundary. More and more information is now being given by awarding bodies to candidates, including the return of scripts at GCSE and A-level, if you want them, and there is greater diagnosis of performance, particularly from Edexcel. If there is error, the objective is to detect it and then resolve it. The process of lodging an appeal after a result and having that heard and the paper re-examined is a legitimate and important part of the whole thing. We cannot say that the system works impeccably unless there are such robust appeal processes and they are seen to work.

Q82 Annette Brooke: Given that the 30% figure has been mentioned, surely that is something that you have investigated fully and looked at the evidence for? Can we really say that the Government are quite justified in being confident in the test results that are finally published?
Dr Boston: Yes, I can certainly say that we are confident in being published. However, it must be said that there are various views of comparability which compound all of this, and make people wonder whether the standards or grades are being met. One of the most recent arguments about grade inflation has been the work that Robert Coe has run from Durham, which has been interesting work to look at. He has taken what he called the test of developed ability, which was a notion of innate ability—developed ability—in an individual, and he took the example of the person getting A-levels and said that A-levels had become two grades easier over the last 20 years, and that that was a problem.

Q83 Annette Brooke: I am not really talking about grade inflation at the moment. I am actually talking about fundamental errors and confidence in the system. I agree that grade inflation is undermining confidence, but in this sitting we are not concentrating on that.
Dr Boston: Error exists. As I said before, this a process of judgment. Error exists, and error needs to be identified and rectified where it occurs. I am surprised at the figure of 30%. We have been looking at the range of tests and examinations for some time. We think that is a very high figure, but whatever it is it needs to be capable of being identified and corrected.

Q84 Annette Brooke: What is the primary purpose for which Key Stage tests are designed? We were talking about a variety of purposes. What is the No. 1 purpose?

Dr Boston: The No. 1 purpose is to decide the level that a child has reached at the end of a key stage.

Q85 Annette Brooke: Given that the tests are also used for the purpose of assessing a school's performance by parents, local authorities and the Government, and that we want more teacher assessment—you said that yourself—and full professional development, do you think it is reasonable to ask teachers to judge pupils' performance when they themselves and their schools are being judged by the results? Is there not a major conflict here?
Dr Boston: The use of diagnostic assessment and assessment of pupil performance, and training teachers to have an understanding of standards and to be able to decide where their children rest, where their achievement is, is very sound. I have talked about teacher assessment before and see immense value in it, and in the Institute of Educational Assessors in moderation, but I am not signed up to the abolition of external tests and to the elimination of external marking. I certainly think that it has a place and that in any assessment system a balance is needed between internal and external, but I certainly would not sign up for a summative assessment process that did not include a significant component of external marking.

Q86 Mr Chaytor: Of the 22 purposes to which the assessment results can be put, you stressed what you think are the most effective purposes, which are best served by the current system. Which of the 22 are least well served by the current arrangements?
Dr Boston: With regard to the personal value of students' achievements and the formative assessment to identify students' proximal learning needs and guide subsequent teaching, the national curriculum tests are less effective than the new tests—the single-level tests—will be. In respect of student monitoring to decide whether students are making sufficient progress in attainment in relation to targets, the single-level tests will do that better; and no, these are not the tests to deliver the diagnosis of learning difficulties. We could develop better and simpler tests to identify the general educational needs of students to transfer to new schools. We can use Key Stage tests to segregate students into homogeneous groups or screening to identify youngsters who differ significantly from their peers, but we could simply design better ones as well. I will not go through the 14; it is a matter of stripping down. The tests are good at assessing institution performance; a standard test is applied to all schools in the country to children of the same age and it will give you at one level a measure of the performance of that institution. You might want to moderate that when you to come to setting targets for that institution in terms of its intake, but they are pretty good at that.

Q87 Mr Chaytor: In terms of institutional performance, does it follow that the function of school choice is effectively served by the current tests?
Dr Boston: It could be served by a better test.

Q88 Mr Chaytor: You have been very strong on the effectiveness of the tests and the importance of full cohort testing. But full cohort testing is not the only way of getting the information that the Government and the public require. Why has the QCA been so resistant to techniques of light sampling?
Dr Boston: I do not think that we have been resistant to it. In fact I think we were the first people to start talking about it publicly. We offered advice to the Government and the Government were not at that stage heading in that direction. They were heading in the direction of the progress tests, as they were then called, or the single level test, which I think is fine. But one of the issues with the Key Stage tests is that they are a full cohort test. There is a new test each year. They take a long time to develop and then all the test items can no longer be used again. The Government set great store by sample tests such as PIRLS, PISA and TIMSS. In other countries such as America, for example, the national assessment of educational progress is a test of a statistically valid sample, which takes the same test items each year. It is slightly changed, but it is basically the same thing. It will give you an absolute measure of whether standards on that test are rising or falling. It is horses for courses. There are ways in which this can be organised. The way that the Government are moving is to go for the single level tests, which I strongly support. But we need to be wary, if we are to have single level tests but phase out Key Stage tests, that we do not saddle the single level tests with these 14 functions. We should use the single level tests for some of the functions and have other sorts of tests for other functions.

Q89 Chairman: If we have been using all these tests for 14 different things all this time, is it legitimate for people like us to say to you, well where was the QCA? Have you been telling Ministers over all these years that this is a ridiculous system of testing and that it is so wide that we are picking out 14 different outcomes and that you need to divide into four very specific groups—your corkscrew, your screwdriver and so on? Where have you been? Have you been telling the Government this for a long time and they just would not listen?
Dr Boston: No. I do not think that that would be fair for you to say. The discourse on what assessment is about and how we do it is a public debate.

Q90 Chairman: I am sorry, but most of my constituents do not believe that. Parents of children taking tests believe that you are the person who looks after this sort of stuff, and that if you do not like what is going on, you should tell the Government that they should do something about it, and, if it really came to it, that you would come out from your corner and say that tests are not fair.
Dr Boston: I am certainly not saying that the key stage tests are not fit for purpose. I am saying that there are some purposes for which they are far fitter

than others. They can be used for these purposes. There is no question about that. But for many of them there is a better way to do it.

Q91 Chairman: That is what our expert witnesses have been saying: there are too many tests. You have not really answered that. We went to New Zealand and they said that they would like to know more about our students, but that to test at 7, 11, 14, 16, 17 and 18 we must be crazy. Why does the QCA never seem to say anything about the number of tests and the fact that the other expert witnesses say that those tests are not fit for purpose?
Dr Boston: I do not believe that there are too many tests, particularly in primary education. There is undue pressure on preparation for the tests, but if we consider the amount of time that is actually taken up by the testing process, it is not high, and it is certainly higher in some other countries. In secondary education it is far more intense. There is no question about that. Our concern—or my concern—has not been with the burden of assessment, as people sometimes refer to it, but with the high stakes put on the assessments because, in the case of Key Stage tests, they carry 14 different functions.

Q92 Chairman: That is what we keep coming back to. Why have you not blown the whistle on those 14 different functions and said that they should not be used in that way?
Dr Boston: I provide advice to Government. I am not out there as an independent commentator.

Q93 Chairman: Are you saying that you have told the Government that they are not fit for purpose for a long time and they have not reacted to that?
Dr Boston: No. I have never told the Government that these tests are not fit for purpose because I do not think that that is the case. I think that they are fit for purpose. I have certainly said that there are many purposes that would be served better by different sorts of tests. Indeed, as you know, some time ago I raised the issue of sample testing, on which the Government were not keen for other reasons.

Q94 Chairman: What about the other point that we picked up on in the evidence—that people said that because you have not blown the whistle on the tests, they drive out the ability to teach a decent curriculum; that the teachers are just teaching to the test and cannot explore the curriculum?
Dr Boston: Fundamentally, our task has been to develop, deliver and build these tests and to make sure that the results from them are valid. Although I admit that there are some errors in them, we make sure that there are processes for that error to be identified and for the problem to be resolved. We have been extraordinarily forward in pushing for the introduction of more technology and scanning, precisely for reasons of improving the quality of marking.

Q95 Chairman: We visit examining boards and often the progress and innovation comes from them, not from you. I get the impression that you are running behind Cambridge Assessment and Edexcel. They are teaching you how to do that stuff.
Dr Boston: Edexcel, which was the first to get into online scanning and marking in this country, would not have got there without the very strong support that it had from QCA, both publicly and through the Government. The fundamental argument related to improvements in the quality of marking. You will remember the fuss that occurred at the time when the contract went to Edexcel—or Pearson—about bringing in a private, overseas company to run marking when previously it had been done by charities. The argument that we publicly and strongly ran then was that that was the way forward. It was the way to guarantee quality in marking and to eliminate problems because second marking would take place alongside first marking with the material coming up on the computer screen.
Chairman: We will drill down on that in a minute if you do not mind. I want to call Stephen now to talk about test targets and tables.

Q96 Stephen Williams: How do you go about deciding what a child should know at each particular stage in their life? Ten days ago we had the Key Stage 2 league tables reporting that children by age 11 are meant to reach Level 4 across the subjects. How was it decided what the content of Level 4 is and what the target is for an 11-year-old to get to that level? What process is gone through to reach those two things?
Dr Boston: That is a very technical question that I am sure someone behind me could answer if you were prepared to let them, or they could slip me notes and I would attempt to make a fist of it.
Chairman: It must be the latter and not the former, otherwise *Hansard* will be driven up the wall, so if you do not mind, we will be happy to give you some time for someone to supply a note. Stephen, do you want to change the drift of your questions until you get an answer on that?

Q97 Stephen Williams: I will ask the question in broader terms because, in the introductory session, Ken basically said that there was a constitutional nicety, a separation of powers between a regulator of standards which ensured that everyone had confidence that nothing was being politically manipulated. Is it the QCA, all the advisers sitting behind you and those behind them back at headquarters who decide what Level 4 is and how many children should reach it by a given age, or is that box-ticking mentality started in the Department and you are told to design a curriculum to deliver that?
Dr Boston: I understand the question, but I know that the experts sitting behind me will give a better answer than I could, so can we move on and I shall take the question in a moment?
Chairman: Lynda, did you want to come in on that point?

Lynda Waltho: No, I wanted to follow on later.
Chairman: Carry on, Stephen.

Q98 Stephen Williams: Ken, you said that the primary purpose of all Key Stage tests was to assess the individual performance of a child, yet what gets all the attention—I refer to the August media frenzy—is the performance of a school, which is an aggregation of the performance of all the individual children. Do you think a fair outcome of the tests is that schools should be held to account or do you think that is a minor delivery of the system?
Dr Boston: No, I am firmly of the view that schools should be held to account. I believe in full cohort testing. I believe that full cohort testing and summative assessment have a place and that holding schools to account for what they achieve is important.

Q99 Stephen Williams: When you say that you believe in full cohort testing, I can see that means that you believe in testing every child, but do you also believe that the publication of the aggregate results for every child is fair on the school, or would a national sample be a better way of measuring standards across the country?
Dr Boston: Yes, I certainly believe in reporting the achievements of that school. These are the other schools.

Q100 Stephen Williams: That is based on the aggregate of the results for each child in the school, so would you reject alternatives completely?
Dr Boston: It depends on the purpose. If your purpose is to find out whether children are writing and reading as well as they did 10 years ago nationally, the best test is to give a sample of them virtually the same test as was given to a sample of them 10 years ago. That will tell you whether we have gone up or down. If you want to report on the performance of a school this year in relation to the school next door, the sample will clearly not do that, but the full cohort test will. It is again about purpose. Both purposes are legitimate and some of the difficulties with the testing programme or the examination programme when looking at whether standards have changed significantly year on year or over a 20-year period, are that the curriculum, teaching methods and other things such as class size have changed. If you really want to know whether people are better at reading than they were 20 years ago, give them the same test.

Q101 Stephen Williams: I see that you have had time to look at the note that has been passed to you.
Dr Boston: This is Dr Horner's contribution in handwriting. The report of the task group on assessment and testing in 1989—while I was still elsewhere, Barry—decided on a 10-point scale and proposed a graph of progress that identified Level 4 at age 11. That report was then published, presumably. We are now on an eight-point scale—aren't we?—so that 10-point scale has been reduced. That does not fully answer your question, Stephen.

Q102 Stephen Williams: No, four out of 10 is 40% and, on the same scale, four out of eight is 50%, so it seems to be a completely different target. Perhaps we are getting a bit too technical. I was trying to get at whether the Government were feeling the QCA's collar in respect of how we set the standards for children. Therefore, is it right that we have a separate regulator of standards for the future?
Dr Boston: Barry, I should very much like to give you a written statement tomorrow in answer to this question.[2]
Chairman: Okay.

Q103 Stephen Williams: Before the note, I was going to mention the difference between how a child's performance is assessed and how a school's performance is assessed. You were going into how children's performance was assessed over time. Is there another way in which you can assess a school's effectiveness apart from the league table mentality that we have at the moment? Is there an alternative? After all, they do not have league tables in Wales or Scotland.
Dr Boston: You can certainly assess the performance on the basis of teacher reporting, as against the school reporting its performance against a template of benchmarks, perhaps, as occurs in some other countries, and reporting its testing of its students against national averages in literacy, numeracy and so on. I know of cases where that occurs.

Q104 Stephen Williams: You are a man of international experience. Do you think that anywhere else does it better than England—whether a state in Australia or anywhere else—without this sort of national frenzy every August, with people wondering whether things are going down the pan or the Government saying, "No, things have only ever got better"?
Dr Boston: I think England is pretty rare in the way it does this in August, although I would not say unique.

Q105 Chairman: Is that good or bad?
Dr Boston: The annual debate about whether too many have passed and whether standards must have fallen is a very sterile debate and I would be glad to see the back of it. If it is right that this new regulator will lead to the end of that, it is a good thing. We are not so sure that it will. There are other, better, ways of celebrating success and achievement—not questioning it.

Q106 Stephen Williams: Do you think that any particular country does it a lot better than England?
Dr Boston: No. I think that in other countries where the results come out there is less public criticism of youngsters on the basis that, because they have three

[2] *Note by witness:* In 1988 the Task Group on Assessment and Testing (TGAT) designed the assessment system for the national curriculum. This included the development of a then 10 level scale to cover the years of compulsory schooling. Level 4 was pitched as the reasonable expectation for the end of the primary phase, to ensure pupils could move on with confidence in their skills to tackle the secondary curriculum.

A grades, the result must be worthless. Such criticism is a very bad thing. From my previous experience, I have a great interest in Aboriginal education. There was 200 years of Aboriginal education in Australia with absolutely no impact on the performance of Aboriginal kids until we introduced full cohort testing and reporting at school level. Then, suddenly, people took Aboriginal education seriously and it began to improve.

Q107 Stephen Williams: If nobody else wants to come in on this section, I should like to ask one last question, going back to the start, about introducing the new regulator. We have only had the report, *Confidence in Standards*, from the Secretary of State today, and we have not been able to digest it fully yet. When were you consulted about the split in the QCA's responsibilities? Was it before, during or after the August round of exam results that we had a few months ago?
Dr Boston: It was after.

Q108 Fiona Mactaggart: You have been talking clearly about the difficulty of tests fulfilling 14 different purposes. The fact is that they fulfil some of those inadequately. You suggested that the best way to see if children over time are able to achieve the same standard is through sampled testing. Do we do very much of that, and if not, why not?
Dr Boston: Those tests will tell you whether performance on a particular task has improved over time. We do not do that as a country. We pay a lot of attention to PIRLS and PISA and the national maths and science study. In developing its Key Stage tests from year to year, the QCA does pre-test. Part of those pre-tests in schools, which youngsters think are just more practice tests, are pre-tests for what we will use in 18 months' time. In them we often use anchor questions, which are the same questions that have been asked a few years before or in consecutive years. They might be only slightly disguised or might not be changed at all. That is to help develop tests that maintain standards so that Level 4 is Level 4 year on year. The boundary between the levels is set by the examiners. It might be 59 in one year and 61 in another, but they know that in their judgment that is a Level 4. They draw on those tests. We have not used the tests systematically enough to say, "We used these six questions for the past eight years and we know that students are getting better at reading or worse at writing," but that is the basis on which we develop and pre-test those emerging assessments.

Q109 Fiona Mactaggart: I am struck by this. You are saying that we do it a bit to ensure the comparability of tests over time. We all accept that some of that kind of work is a necessary function of getting accurate summative tests, but there is a constant threat in debate in assessment about whether standards have changed over time. I do not think that I properly understand why we have not bothered to invest what does not strike me as a very large amount of resource in producing that kind of sampling over time to see whether standards are improving or weakening, and where. We would then have a national formative assessment about where the strengths and weaknesses of our education system are over time. Do we have a mechanism that is designed to do that? If not, why not?
Dr Boston: No, we do not. We use PIRLS and PISA and in the recent results, they confirmed what we already know; for example, at PIRLS level, the line I talked about is not as steep as it was before. It has flattened off, but has not come to a plateau. The notion of a sampling programme is something that we have raised with Government. Some years ago, before I came into this job, there was the Assessment of Performance Unit, which did some of that work. That is no more. I do not know the background and the reasons why the work was not pursued, but it was work of this sort. It would seem to me that we need to be thinking not of either/or. That is the message that I really want to get across. We are not thinking of Key Stage tests or single level tests or sample tests. If we want to serve those 22 legitimate purposes of testing—I am sure there are more—we need a number of tests that will deliver between them all those things, but which are designed so that they are very close to what Paul Newton calls the design inference, where the user inference and the design inference are very close indeed.

Q110 Fiona Mactaggart: What I do not understand about the proposed new system is that if we developed a wider range of tests to separate some of these functions more precisely so that we get more accurate information rather than trying to infer information from tests that are designed to do something else, which is what we do at present, who would take the lead in developing the sample tests and introducing them? Would it be the QCA or the new regulatory authority? I have not had time to read through the document, but I do not understand whose job is what.
Dr Boston: It would be the QCA, and it would do its work partly through stimulating the private sector market and the awarding bodies to work with it. Presumably the QCA would take the initiative on remit from the Government. That would be critical: the Government would decide that they wanted a set of new tests. We did not go out and invent single level tests. We were remitted to do them. We produced them at Government request, and with our very strong support. So the initiative would rest fundamentally with the Government, but the body that would lead on it would be the QCA, or whatever the QCA might end up being called some time in the future. The regulatory authority is to ensure that, once the product—the assessment—is there, it delivers on standards and maintains standards. The regulator is not a development authority; it is an authority to regulate products and ensure their quality once they are there.

Q111 Fiona Mactaggart: When you were remitted to develop the concept of single level tests, were you remitted to develop a test that was a one-way street, rather than a test that could be re-administered? I gather that the National Foundation for Educational Research is concerned about the fact that this is just a

single pass test and that someone who chooses when they do it might pass then but might not necessarily pass it a month later.

Dr Boston: We were remitted to produce a test which would be taken as a one-off. Further down the track if we get to a point, as I think we might, where single level tests are available virtually on line, on demand, we would need to go to a data bank of test items. What we have at the moment is a Level 3 test or a Level 4 test. A judgment is then made on the score you get about whether you are secure in Level 4. That test is then finished with. The time may come in the future, as with Key Stage 3 ICT tests, where there is a computer in the corner on which you can take at any stage your Level 4 or Level 5 reading test. That would depend on a data bank. In that sense it is constantly renewable, if I understand the question correctly.

Q112 Fiona Mactaggart: It was not so much about whether it was renewable. If the teacher of the child can choose the moment at which the child takes a single level test and it is a propitious day for that particular child, the child may do well in the test and succeed, but it might still be rather a frail attainment. There is anxiety about whether that is a fully accurate picture of the child's capacity and the general learning level even though they can do it on a fair day with wind behind them.

Dr Boston: I am remiss in that I have not fully explained the relationship between the Assessment of Pupil Performance and the tests. The APP programme is designed essentially to produce greater understanding among teachers about what is represented by a level—the profile of a Level 4 in reading, the profile of a Level 5 in reading and the difference between them. It represents the different indicators that show a child is either at Level 4 or Level 5, and the child is then entered for the test. The test is meant to be confirmation that the teacher has made the judgment correctly.

Sitting suspended for fire evacuation.

On resuming—

Chairman: Dr Boston, we are back in business, although only briefly. I suspect that we will have to call you or your team back at some stage, because this has been unfortunate. I will give a question to each member of the team, and you will answer speedily. I will start with David, followed by Stephen, then Fiona, and I will finish.

Q113 Mr Chaytor: On maintenance of standards, will the new A* grade at A-level have the same pass rate in all subjects across all examining boards?

Dr Boston: No.

Q114 Mr Chaytor: Does the existing A-level threshold have the same pass rate in all subjects?

Dr Boston: No.

Q115 Mr Chaytor: Does that cause a problem?

Dr Boston: No.

Q116 Mr Chaytor: Will there not be a huge discrepancy between different subjects in different boards?

Dr Boston: The A/B boundary is set by professional judgment. The reality is that subjects are different; there is no attempt to say that, for example, 10% must pass or have an A grade in every subject. No country in the world achieves precise comparability between subjects in terms of standards. Australia tries to do so: it takes all the youngsters who get a certain grade in, for example, English, geography, and art, and, if they find that a lot of the youngsters who are taking those three are getting higher grades in geography than in the other two subjects, then they deflate the mean of geography. Some pretty hairy assumptions underlie that. Here, an A/B boundary is set by professional examiners broadly at the level that a hard-working, well-taught, student who has applied himself or herself fully would achieve on a syllabus or specification.

Q117 Mr Chaytor: Are the thresholds for subjects on examining boards matters of public record? That is, is the percentage score that triggers a B, an A or an A* on the record and available to pupils and parents?

Dr Boston: The answer is no, I believe.

Q118 Mr Chaytor: My next question is, should it be?

Dr Boston: I would think not.

Q119 Mr Chaytor: Why not?

Dr Boston: The essential point is that you might have a harder paper one year than another, in which case the boundaries might change significantly. The point is not the numerical score where the boundary is drawn. The fundamental point is the professional judgment of the examiners, who decide where the A/B boundary is and where the E/U boundary is. They do that on the basis of their experience and past statistical evidence using papers of similar demand.

Q120 Fiona Mactaggart: Does the fact that schools are held accountable through tests that are really designed to be summative tests of children's achievement mean that teachers teach a less-rounded curriculum?

Dr Boston: My only reaction to that is absolutely anecdotal. We have a network of 1,000 schools to which we relate intensively, and I have been told by people at the QCA who work closely with schools, and from what I hear from professional bodies, head teachers and so on, that their answer to that question is frequently yes. I do not run a school, and I do not have first-hand evidence of that, but all the evidence that I hear in my position is about the narrowing of the curriculum that results from these tests. Presumably, there may be some better approach to that with the single-level tests. I have also spoken to many head teachers who are probably the exception to the rule and say, basically, the objective is good educational nutrition for these youngsters, and if they have got that they will pass the tests. That is a better way than simply narrowly training them to take the assessment.

Q121 Fiona Mactaggart: I am sure that they are right. However, because of lack of self-confidence and other things among many teachers, such teachers

are not in the majority, I suspect. Would it be possible for you to devise a test? I have listened to you speak about testing. Your commitment is to using testing to improve the quality of education for children, yet here seems to be some evidence that in one respect testing in Britain is narrowing the quality of education for our children. Could you devise a separate way of holding schools accountable, which could avoid that difficulty so that that function is dealt with differently from the way in which we were assessing children's attainment?
Dr Boston: Holding them accountable for what?

Q122 Fiona Mactaggart: For the quality of teaching. At the moment, they are held accountable by the attainments of the children through examinations.
Dr Boston: I see the desirability of the aim, but at the moment I cannot neatly and glibly say, "Yes, we could do this, this and this." I see the point of the question.

Q123 Fiona Mactaggart: In the meantime, is there anything you can do to reduce the burden of testing in terms of the rest of the curriculum?
Dr Boston: Apart from providing advice to Government on assessment reform, I cannot see a way in which, within the ambit of the QCA itself, we could be catalytic in producing that change.

Q124 Stephen Williams: Perhaps I can go back to the subject of what was called the historic day—I assume that that reference was to the announcement on *Confidence in Standards* that was made earlier today. In my earlier question to Ken, I asked him when he was consulted about the split, and about the setting up of the new organisation. Have you been consulted on the structure? I have been reading chapter 2 during our sitting, which does not make it clear whether there will be a sort of Ofsted, with a chief inspector and a board. I think that I heard you refer to a board—is that right?
Dr Boston: We have certainly been consulted, and our advice has been sought on where we might go from here now that the Government have made the decision to go ahead and now that consultation has happened. The intention, as I understand it—I thought that it was set out in the document—was that there should be a non-departmental body with its own board and its own chief executive. I have no detail beyond that at this stage. We have been consulted and have been asked for quite detailed advice on how we might set up shadow arrangements—I described our proposals on that earlier. They have still to be accepted by Government, but they seem to be an intelligent way forward.

Q125 Stephen Williams: If we assume that there will be a board—I cannot see that in the document, but I have only skim read it so far—what sort of people should be on it? In relation to A-levels, do you agree that it would be sensible for universities to be represented on the board, given that roughly 90% of children who achieve A-level standards now continue to higher education?
Dr Boston: The regulator will of course be responsible for all qualifications—not just the general ones but vocational and adult ones, too. The regulator will clearly have a role in devising new approaches to the recognition of awarding bodies, including the post-Leitch recognition of employers as both awarding bodies and training providers. The board of the new body would, I think, need to consist of higher education representatives, business representatives and teaching profession representatives. It would probably be pretty similar in composition to the current QCA board.

Q126 Chairman: We shall have to finish now, but is it right that you have a choice as to which way you jump? Can you choose which organisation you opt for?
Dr Boston: No, I will continue as Chief Executive of the QCA.

Q127 Chairman: I have one last question. When we pushed you today, you tended to say, "But I'm a regulator." In a sense, therefore, some of your answers have persuaded me that the reforms are right. When I asked you why you did not push for the reforms or take a certain course in advising the Government, you showed a certain unhappiness. The indication was that there was a functional stress between the two roles. Is that right?
Dr Boston: There is a stress, yes. I am not an independent commentator on education. I certainly have a responsibility under the current legislation to be absolutely separate from the Government and from everyone on maintenance and regulation of standards. My position has always been that the minute any Government attempted to interfere with that, I would be the first to declare it publicly. On issues such as the curriculum and provision of qualifications, the current role is to advise the Government. We do not have the capacity to go out and say that we are simply going to introduce a new form of testing in two years' time. Those decisions are for the Government—they always have been, and they always will be. There has been tension, and you have exposed it cleverly in our discussion.
Chairman: Ken Boston, it has been a pleasure to have you here. I am sorry that we were disrupted and that there is unfinished business that perhaps, when you return from Australia, we can revisit with you. Thank you to all those who have attended. I wish a happy Christmas to everyone.

Monday 14 January 2008

Members present:

Adam Afriyie	Mrs. Sharon Hodgson
Annette Brooke	Fiona Mactaggart
Mr. David Chaytor	Mr. Andy Slaughter
Mr. John Heppell	Lynda Waltho

In the absence of the Chairman, Fiona Mactaggart was called to the Chair

---

Memorandum submitted by Association of School and College Leaders (ASCL)

A. The Present Situation

Assessment in Britain requires a radical review

1. In England, young people take externally set and marked examinations at the ages of 7, 11, 14, 16, 17 and 18. The system is at breaking point as more and more examinations have been added to an already over-examined system. The total number of examination papers sat by young people in schools and colleges each year in national curriculum tests at 7, 11 and 14, GCSE examinations, GNVQs, AS and A2 examinations and key skills tests is over 30 million. No other country has so many examinations, taking place so frequently in the life of a young person. Whilst Wales and Scotland are in a slightly better position than England, their examination and assessment systems are also heavily over loaded.

2. The ASCL paper *Examinations and Assessment* (SHA, 2002), stated:

> *We do not argue against assessment. Far from it. High quality assessment is an important part of good teaching. [But] the purposes of assessment have become confused. This has happened largely because external examinations have assumed too much importance in the system. Examinations have become the master of education, not the servant.*

3. The Tomlinson report, published in 2005, recognised the problem of too many examinations and advocated greater reliance on in-course assessment by teachers, recommending the use of chartered assessors, as proposed by ASCL [SHA] since 2002. The Daugherty report on assessment in Wales also advocated a reduction in assessment and the Wales Assembly Government has put this into place, although the replacement system is proving unnecessarily bureaucratic.

4. The current problems on assessment may be summarised as follows:

- Young people are subjected to far too many external examinations. These take place more frequently than in other countries. The relentless pressure of external examinations can interfere with the enjoyment young people take in learning, can lead to excessive levels of stress, and in extreme cases to mental health problems.
- Schools and colleges spend too much valuable curriculum time in directly preparing for, and conducting, external examinations.
- The purpose of external examinations is confused between diagnostic, summative and qualification (for the examinee), component of performance management (for the teacher), accountability (for the school) and indicator of national achievement (for the nation).
- The examination system is very costly (see paragraphs 5–12 below).
- The complexity of the examination system has led to concerns about the accuracy and consistency of marking and results, with increasing numbers of re-marks being sought at GCSE, AS and A levels.
- It is becoming very difficult to find sufficiently qualified and experienced staff to be the markers, moderators and examiners of the external examination system. As a result, some papers are being marked abroad.
- There is a lack of trust in the professional ability of teachers to carry out rigorous internal assessment.

5. The cost of external examinations is excessive and uses too high a proportion of school and college budgets. The cost comprises three elements:

- Examination fees.
- Administration time (carried out by support staff since September 2003).
- Invigilation (carried out by support staff since September 2005).

6. The PricewaterhouseCoopers (PwC) report on examination costs, commissioned by QCA in 2003, published in 2005 a figure of £610 million as the cost of the examination system. ASCL has carried out its own surveys from time to time and our figures suggest that the cost is at least that figure. The costs are broadly consistent between institutions of comparable size.

Table

THE COST OF THE ENGLISH EXAMINATION SYSTEM

	Direct Costs	*Time Costs*	*Total*
QCA Core costs	8	—	8
QCA NCT costs	37	—	37
Awarding body costs	264	—	264
Exam Centres—Invigilation	—	97	97
Exam Centres—Support & Sundries	61	9	70
Exam Centres—Exams Officers	—	134	134
Total costs (£m)	**370**	**240**	**610**

Source: PricewaterhouseCoopers

7. Since the PwC survey costs have risen further. ASCL does not have aggregated figures (though these may be available from the DfES) but it is clear from a small sample of schools and colleges that the direct cost to institutions has increased. Some examples are:

8. An average sized sixth form college in the West Midlands with roughly 1300 full time students spends £300,000 on examination fees, invigilation, and administrative staff employed solely for examinations work. In larger sixth form colleges, the cost of external examinations is now well in excess of £400,000—often the second highest item on the college budget after staffing.

9. A large tertiary FE college in the North West has an annual expenditure for examination fees alone of approximately £650,000, and employs three dedicated staff at a cost of £75,000. The principal estimates that about 4% of the college annual budget of £20million goes on external assessment.

10. In a 1,500-pupil comprehensive school with a sixth form in Wales, the cost of examination fees is approximately £100,000. The cost of administration of the external examinations is over £17,000, and the cost of support staff for invigilation is approximately £13,000. A total of £130,000.

11. The cost of examination fees in a typical 11–16 school of 960 students in the Home Counties is £60,000.

12. None of these figures includes the opportunity cost of the time of staff whose main responsibilities lie elsewhere, though teachers, heads of department, and senior leaders all devote a proportion of their time to setting up, supervising and analysing external examinations, and supporting students through them.

B. Tests, Examinations and Their Purpose

13. The purpose of tests and examinations has become confused with school accountability and the performance management of teachers. The same assessments are used for the following purposes:

- Diagnostic assessment.
- Formative assessment.
- Summative assessment.
- Evaluative assessment.
- Ipsative assessment.

They are also used for:

- a component of the qualifications structure;
- monitoring progress;
- teachers' performance-related pay;
- performance management of teachers;
- school and college performance tables;
- accountability of schools, colleges, local authorities, the Learning and Skills Council and the DfES; and
- meeting national targets.

14. Of the last group of seven purposes, five are evaluative, demonstrating how the Government has skewed the assessment system from its prime purposes of diagnostic and formative towards the evaluative. The assessment of the work of young people has become primarily for the accountability of schools and colleges, rather than to be of value to the students themselves.

15. The use of assessment for learning has improved the quality and extent of formative assessment, encouraging students to think more about their own learning and helping teachers to mould their teaching style more effectively to the needs of the students. Assessment for learning has become an important element in student voice, in that it provides students with a structure in which to feed back to their teachers information on the effectiveness of their learning. It is therefore a major contributor to personalising learning.

16. Teachers have been criticised for teaching to the test but, if the system is geared to constantly monitoring progress and judging teachers and institutions by outcomes, it is hardly surprising that the focus is on ensuring that students produce the best results. Particularly at Key Stage 2, this results in over-preparation for the tests in May of year 6, followed by a period with much less emphasis on the tested subjects. By September, when the children enter year 7, they have had four months of this post-test phase—hardly the best preparation for the start of secondary education. Many secondary school leaders believe that this is a major contributory factor in the so-called Key Stage 3 dip in performance.

17. Intelligent accountability for schools and colleges is not helped by the use of test scores to produce league tables, nor by the way in which the Government is trying to produce a single measure of accountability—the contextualized value added measure—as a precise indicator of the effectiveness of a complex institution such as a school or college. Schools and colleges expect to be held to account for their performance, but measures should not claim greater rigour than they can stand and confidence intervals should always be included.

18. By producing league tables of performance at age 14 and by using Key Stage 3 test results as an indicator for Ofsted inspections, the importance of Key Stage 3 tests is magnified unnecessarily. The critical test results in secondary education are at age 16 and 18—no employer or university has ever asked an applicant what they scored in Key Stage 3 tests. A check on the progress of 14 year olds in the major subjects is necessary for schools' planning and self-evaluation, but this could be achieved without the use of an elaborate series of external tests.

19. In a 14 to 19 qualifications system, the importance of GCSE at age 16 will also be played down from the huge external examination industry that it has become. In its early papers on 14 to 19, the Government itself described the future role of the GCSE as a progress check and we agree with this as the 14 to 19 system matures.

20. Nobody criticises A level teachers for teaching to the test, because the test is widely respected and the syllabus provides an excellent education for the students following it. Schools want to focus on developing deep and sustained learning with assessment systems supporting that process and this is possible at A level.

21. ASCL does not support the introduction of the A* grade at A level, believing that there is adequate information available to highly selective universities to distinguish between the best candidates on the basis of their module grades, their raw marks and their wider achievements, information on all of which is available to admissions tutors.

22. The progress of the education system as a whole could be monitored more efficiently and effectively. The aggregation of individual test scores creates a high-stakes testing system in which the pressure is bound to create a false picture of progress. National curriculum testing should not therefore be used to monitor progress towards the achievement of national targets. Instead, random sampling tests should be carried out by a new body, similar to the former Assessment of Performance Unit (APU). Monitoring of progress should be by national sampling, not by national saturation, as we have at present.

C. Chartered Assessors: Using the Professional Judgement of Teachers

23. At all levels of external assessment, greater trust should be placed in the professionalism of teachers who have, in recent years, become more rigorous and skilful at assessment. Internal summative assessment should play a greater part in the examination system.

24. National curriculum tests at 11 and 14, GCSE, AS and A level examinations should rely more on in-course assessment through the professional judgement of teachers.

25. A problem with relying more on internal assessment by teachers is that there is a lack of public trust in the professional ability of teachers to carry out such assessment rigorously. A change in the balance between external and internal assessment must take place in a way that maintains public confidence in the qualifications system.

26. ASCL has proposed the establishment of a cohort of chartered assessors, a system of in-course assessment that will produce no loss of rigour in examining and will thus secure public confidence. Chartered assessors will be experienced teachers, externally accredited to carry out in-course assessment to external standards. The chartered assessors will be responsible for carrying out or overseeing rigorous in-course assessment that will form a substantial proportion of externally awarded qualifications. It will be the responsibility of the chartered examiner to mark and grade work at the standard of the external qualification to which it contributes.

27. Chartered assessors would develop expertise in formative assessment and assessment for learning, as well as understanding and enforcing rigorous standards in tests leading to the award of qualifications. Assessors from one school might also support another school where colleagues were inexperienced in assessment or where there were problems in teacher recruitment and retention.

28. ASCL proposals for chartered assessors are being taken forward by the Institute of Educational Assessors (IEA) and the use of chartered assessors is envisaged in the current development of 14–19 diplomas.

29. Precedents exist for the role of chartered assessors, both in the qualifications for teachers who assess vocational courses, and in the accreditation awarded to modern languages teachers to carry out A level and GCSE speaking tests. Teachers apply for accreditation and undergo training before they carry out oral examinations or in-course assessment to external standards.

30. In-course assessment, if carried out rigorously and to external standards, gives a truer picture of a student's standard of attainment than an external examination taken on a particular day. A combination of externally set tests and internally set work would form the basis for the assessment.

31. One way in which chartered assessors could be deployed has been described by the chief executive of the Qualifications and Curriculum Authority (QCA). In a speech in May 2006, Dr Ken Boston stated,[1] "*no other country devotes as much time and expertise to developing measures of student progress*". He went on to outline ways in which the system could be re-balanced to rely less on external testing without sacrificing rigour in the assessment process:

> *"If teacher assessment were taken to mean that teachers should set their own tests, and decide on that basis whether a child is, say, a Level 4 in KS2 English or a C in GCSE Maths, then I personally would reject such a proposition—not because of any lack of faith in the professionalism of teachers, but because of the impossibility of being able to strike a common standard nationally across all the classrooms in this country.*
>
> *"If teacher assessment meant, however, that teachers in primary schools and in the early years of secondary education had access to a national bank of standard-referenced tests and examinations which had been trialled and piloted by test developers and awarding bodies under QCA regulation; that the tests and examinations were administered within a specific window of time; that the papers were marked using a mark scheme on which teachers had been trained; that their marks were externally and independently audited by chartered assessors belonging to the Institute of Educational Assessors; and that the system for doing so was demonstrably as rigorous and robust as the current system in maintaining standards nationally and producing valid and reliable data on national performance—then it might well be a better process than the current one, and something which the QCA could recommend to Government".*

32. ASCL strongly supports the approach being recommended by Dr Boston. Furthermore ASCL believes that unless there is recognition of the role that chartered assessors can play, the delivery of the proposed 14 to 19 qualifications framework will not be viable.

33. The proposal to create chartered assessors will raise the status of teachers and of in-course assessment in schools and colleges. It will improve the quality of school- and college-based assessment and thus contribute to the raising of standards in schools and colleges. It will provide a new step on the continuum of professional development for teachers. It will provide important professional development opportunities for aspiring classroom teachers. It will make just-in-time testing more viable and reduce the length of the examination period each summer. Above all, it will make the examinations system more manageable whilst retaining the credibility and standards of the external examination system.

D. Progress Measures

34. The use of pupil progress measures, as proposed by the Secretary of State in 2007, is in principle a move in the right direction of intelligent accountability for schools. Good teachers measure the performance of individual pupils on progress made and it is right that the same principle should be used to measure the performance of schools. However, the proposals as set out in the consultation paper will not have the desired effect. There are several specific aspects about which ASCL has major concerns. The response of ASCL to the consultation is appended at Annex A, which includes an alternative proposal from ASCL for the operation of the progress measure so that it acts as an incentive to schools to raise the achievement of all pupils and not just the group of pupils defined by the threshold measure in the consultation paper.

[1] Speech by Ken Boston at the launch of the Institute of Educational Assessors, 9 May 2006.

E. Key Stage 3 Review

35. ASCL strongly supports the Key Stage 3 review proposals from the QCA, but believes that the purposes of the review in re-thinking and broadening the curriculum may be threatened by the continuing narrowness of the Key Stage 3 tests.

F. Diplomas

36. The assessment systems of the proposed diplomas are as yet not fully defined. Experience of previous attempts to introduce quasi-vocational qualifications, for example GNVQ, lead ASCL members to be concerned that the assessment of the diplomas may be too much like those of GCSE and A level. Effective vocationally-oriented courses cannot be assessed in the same way as academic courses. Much of their purpose and value is lost if they are forced to be so assessed. The diplomas should be different from GCSE and A levels and their assessment should fit the purposes of the qualification, not a pre-determined single view of external testing. Parts of a diploma course, such as functional skills, may be most appropriately tested by external tests (quite likely using ICT). But most other aspects should rely on teacher assessment, using chartered assessors as outlined above.

G. Systemic Reform

37. ASCL welcomes the effect of the workforce reform agreement in transferring examination invigilation from teachers to support staff. This is having a beneficial effect in reducing the burdens on teachers.

38. ASCL also welcomes the modernization agenda being carried out by the National Assessment Authority (NAA), which is seeking to streamline the work of the examinations office and reduce the bureaucratic burden in that area.

H. University Entrance Tests

39. ASCL is concerned at the proliferation of university entrance tests. It is extremely difficult, especially for maintained schools and colleges, to prepare students for the many tests that now exist and thus we believe that these tests discriminate against some students and act against the policy of widening participation in higher education.

June 2007

Annex A

Making Good Progress

Response of the Association of School and College Leaders

A. Introduction

1. The Association of School and College Leaders represents 13,000 members of the leadership teams of colleges, maintained and independent schools throughout the UK. This places the association in a unique position to see the progress measure from the viewpoint of the leaders of both secondary schools and colleges.

2. ASCL welcomes the opening of a debate by the Secretary of State on progress measures. Helping every student to make good progress is the function of the education system and so consideration of policies aimed specifically at supporting that is long overdue.

3. The use of pupil progress measures is in principle a move in the right direction of intelligent accountability for schools. Good teachers measure the performance of individual pupils on progress made and it is right that the same principle should be used to measure the performance of schools.

4. However, the proposals as set out will not have the desired effect. There are several specific aspects about which ASCL has major concerns. The association looks forward to a period of consultation and piloting in which the best features of the proposals can be developed and the worst amended or dropped.

5. To support such a process an alternative measure is proposed in section E of this document which the association believes would command much greater support, avoid the faults of the measure proposed in *Making Good Progress*, and would better lend itself to target-setting at all levels.

B. Assessment and Testing

6. The document's clear statement of assessment for learning and endorsement of it is welcome. ASCL shares the belief that this approach, always used by good teachers to some extent and in some form, can be usefully extended; and has for some time championed it. However, it should be remembered that high-stakes, externally marked tests are antipathetic to assessment for learning. For testing to be supportive of learning it must be kept closer to home, with frequent assessments (of all kinds) devised or chosen by the teacher and marked by the teacher.

7. The association welcomes a move away from age-linked testing. The further idea of "testing when ready" is also welcome. Sadly, what the paper sets out is not that but rather a proliferation of the existing testing regime. Modern technology can surely lead us to aspire to forms of testing that enable students to be tested whenever they are ready, not on a given day in an examination hall in six months time. To propose tests on the current model, but more often, is to miss an opportunity to devise something better for our children, and potentially to exaggerate the faults and costs of the present system.

8. In both of these areas the paper is hidebound by the prevailing orthodoxy of testing, which has prevented any genuinely creative ideas.

C. Personalised Learning

9. ASCL also welcomes the renewed emphasis on personalised learning. It is closely bound to assessment for learning, and is again not new; good teachers, and good schools and colleges, have always tried to personalise their offering to students.

10. This is recognised in *Making Good Progress* and in the *2020 Vision* report, one of the good features of which was its recognition of the good practice already in the current education system. The present document somewhat loses sight of that by extracting (on page 14) a list of approaches that schools "will need to adopt" as if they were not all in the usual repertoire of school behaviour. Some may need more emphasis in some schools.

11. The "personalised classroom" as set out in the first paragraph of page 16 is an attractive prospect, but for it to be realised it is imperative that the teacher not only has ready access to the necessary data but also can rely on it. The present high-stakes testing regime and the weakness of the national curriculum tests prevent any such reliance. Many schools make use of CAT, Midyis or other diagnostic tests for example because they do not feel able to rely on the National Curriculum tests, which were devised as summative tests, as a good baseline for predicting the future performance of each pupil.

12. A more rapid response to pupils who are falling behind is clearly welcome, provided that that does not translate into ever more frequent, stress-inducing, external tests. Our young people have become the most tested in the world, and their stress levels have risen markedly as that has happened. There is now a need to take greater care with their mental health and normal development.

13. The document does recognise at this point that teachers are already skilled at discovering the progress of their individual pupils and tailoring their courses to their needs.

14. ASCL welcomes the clear statement that personalisation does not mean devising a separate plan for each student, and the renewed promise of greater flexibility in the secondary curriculum to allow schools room to be more creative in devising programmes suited to their particular students.

15. There is a contradiction between the idea of personalised learning, which recognises the different needs and abilities of each person, and the setting of systemic targets, which presupposes that all young people should ideally travel the same path at the same rate.

16. The suggestion in the *2020 Vision* report that students from disadvantaged backgrounds should receive additional support is welcome, and the document does no more than reiterate this. However, at one of the DfES presentations to stakeholders this was extended to an intention to provide 10 hours of individual tuition to students not "on trajectory", possibly at home, at weekends or in the school holidays—provided by local authorities. ASCL cannot welcome this interpretation of the *2020* suggestion. It would be very expensive, costing far too much to administer as well as overlooking the possibility of joint work with small groups of students in similar states of learning and with similar needs. It would be very unlikely to be good value for money.

17. If additional funds are available for this type of support they should be delegated to schools, which are closer to the individual students and will be better able to apply them than local authorities.

D. Measures and Targets

18.ASCL strongly opposes the proposals in section five of the document. In this section the proposals go badly wrong in ways that would ensure that the good intentions of the earlier sections could not be realised.

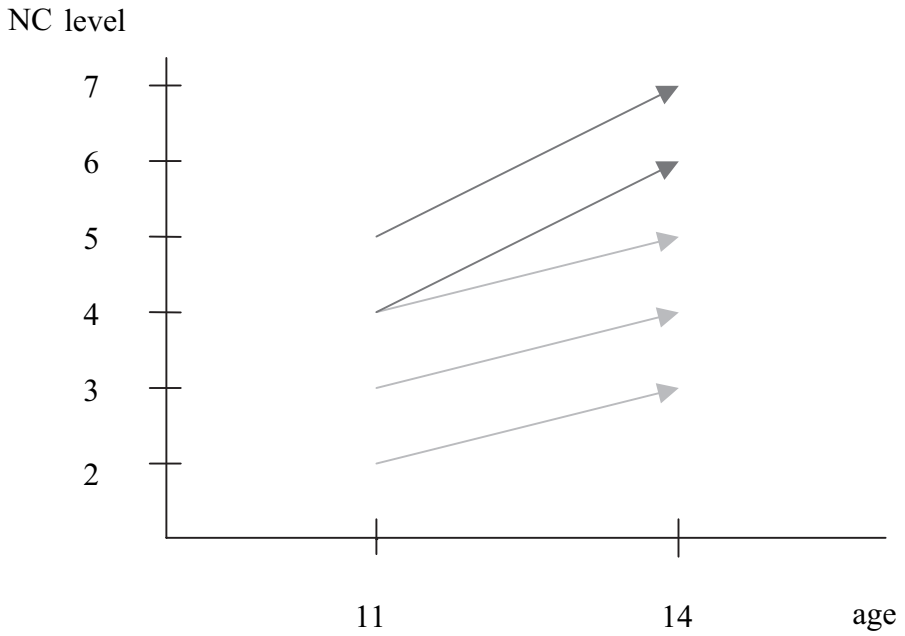
19. First, measures framed as "the percentage of children who . . ." are bad measures of progress. They concentrate the attention of teachers, schools, partnerships, local authorities, inspectors, government and the media on those children on the borderline of making the grade when we should all be interested in *the progress of all children*.

20. Sensible measures in this area should look at the distance travelled by each child related to how far we might reasonably expect a child with that starting point and that set of attributes (disadvantages for example) to travel in the time.

21. Secondly, the proposal that every child should be measured against an improvement of two national curriculum levels is absurdly crude. It may be easy to understand, but will mislead most of those who see it, and will create new perverse incentives as damaging as those caused by the some of the present measures. *For every complex and difficult problem there is a simple and straightforward solution . . . that is wrong*.

22. That a child who is badly behind at the start of a key stage, a child who is a high flyer, a child with a strong leaning towards or away from a particular subject, a child with every sort of support, a child with a profound disability, a child with severe social disadvantage, a child simultaneously learning English, a child who learnt English at a previous stage should all somehow move on two levels bears no examination. In fact, any research that has been done into these and other interrelated factors is ignored here.

23. The information that is set out in *Making Good Progress* points to a further weakness in the proposals; that they would systematically favour selective schools and other schools that have a more able than average intake. This is illustrated clearly by looking at Key Stage 2 to Key Stage 3 mathematics:



This diagram is highly simplified, some children make greater or less progress than indicated, but the arrows show the most common movements during Key Stage 3. Note that on the whole those children who start at lower levels mostly do not make two NC levels of progress, whilst those who start at higher levels mostly do. The effect at school level would be to create a measure that would imply that a school with a "good" intake is doing better than a less favoured school, almost independently of the excellence of the schools themselves.

24. Thirdly, the key stages are of different lengths, and the national curriculum levels not designed to be of equal size. So it is inevitable that, quite apart from the many individual differences between children, a target of two levels per key stage will be much harder to reach at some stages than at others. Figures in the document itself make it quite clear that this is the case. Adopting such a set of measures would therefore invite media attention of the most unwelcome and ill-informed kind: "only 30% of children make acceptable progress in English between age 11 and 14", for example.

25. Encouraging such misunderstanding cannot be to the advantage of children, schools or the Government itself. Many children will be given the false impression that they have "failed", when in fact they have made perfectly normal progress. Secondary schools will be painted as failing their pupils even when they have made above average progress. And the Government will be accused of allowing a systematic failure of education at Key Stage 3.

26. Targets framed in this way will set up new perverse incentives. Students who part-way through a key stage have already clearly made their two level improvement, or who clearly cannot do so, will not be targeted as intensively as those who may or may not make that improvement.

27. A two key stage improvement may be an appropriate aspirational target for many individual students, but it is not appropriate for all, and it is certainly not appropriate as an accountability measure for teachers, schools or Government at Key Stage 3.

28. Fourthly, the relationship between the national curriculum tests and GCSEs is not close and not well understood. The document seeks views as to how a measure based on a percentage of those moving up two levels on these two incommensurable scales should be formulated. It should not be formulated at all. Any such formulation will fail to measure anything meaningful and will create perverse incentives of the worst kind.

29. Fifthly, this whole section is predicated on the national curriculum tests as robust and reliable measures of attainment. These tests are better now than in their early days but are still not capable of bearing the weight of all the many uses to which they are already put. It is not sensible to erect a further edifice of measures and targets on them.

30. Finally, this section asserts that these new measures should be added to all the existing measures and not replace anything. This is simply wrong. The English education system already has more tests and measures than any comparable system, a larger proportion of scarce resources is diverted from actual learning into setting tests, preparing for them, administering them, analysing the results, reporting the results, and dealing with the inevitable misunderstanding of them by children, parents, governors, the media and others.

31. The assertion that nothing can ever be removed from a bureaucratic system does not sit well with recent attempts by Government to reduce bureaucracy and improve the intelligence of accountability systems. In this case it rests upon a separate assertion, made at page two, that it is the elaborate system of tests, targets and performance tables that has driven up standards in recent years. No evidence is brought to support that idea, which has taken on the aspect of a dogma. Indeed, on the very same page of the document it is undermined by the assertion that it is the Government's increased investment in education that has had the beneficial effect. This seems a more likely explanation: our schools are better led, are better staffed, have better facilities and are better resourced than before, and this has been reflected in better progress. Pupils and teachers are also more experienced at the tests, which is bound to have had a beneficial effect on scores nationally.

32. ASCL urges as a matter of general principle that initiatives should not be taken, in a system already at full stretch, without indicating what it is that they should replace. In this case progress measures are welcome but must replace some of the alternative measures that have now had their day, and done whatever good they may have been able to do. There are plenty of candidates . . .

E. An Alternative Measure

33. As already stated ASCL welcomes the idea of a measure of progress. As a constructive response to *Making Good Progress* what is set out in this section is an outline of such a measure that would command the support of school leaders.

34. Any such measure should avoid the perverse incentives inherent in "the percentage of students who . . ." but should depend on the progress made by all students. It should also intelligently reflect the starting point of the student in question.

35. What is needed as a first step is a more complete and careful analysis of a cohort of students, say those that took the various tests in 2005 that will give an expected outcome for each student based upon the starting point. This will also answer the question, left open in *Making Good Progress*, of what progress can be expected between Key Stage 3 and GCSE. What ASCL proposes is the use of this statistical relationship as a baseline for expected performance against which performance in future years can be judged.

36. Thus each student's result can be compared to the expected outcome and a positive or negative "residual" determined. These residuals can readily be aggregated to give an average for a class, school, local authority and the country as a whole.

37. This should be familiar as the approach taken by value-added measures. Like them it would avoid perverse incentives and be based upon careful research into the actual performance of real students.

38. Traditional value-added measures have the drawback that they are *cohort referenced**, meaning that they relate an individual or a group with the averages for the year group to which they belong. This has some disadvantages. The individual's score is partly determined by the performance of the peer group. Such measures do not really reflect change from year to year; a teacher, department, school, partnership or local authority can improve in performance but find that that is not reflected because others have improved too. And in particular they hide improvement in the system as a whole—the average residual for the whole group must by definition be zero.

39. So what ASCL proposes is different in a crucial respect: the performance of future students should be compared not to their own peers in their own year-group, but to the fixed 2005 reference group. In the sense in which we are using the term here this would make the measure *norm referenced** rather than cohort referenced and therefore avoid the drawbacks outlined in the previous paragraph. It would allow for year to year comparison of the performance at all scales from the departmental to the national.

40. In particular any improvement would be reflected in the measure which could thus be used for any target-setting. The average residual for the nation as a whole would no longer necessarily be zero; if the education system improves new cohorts of students will do better than the 2005 group and this will be reflected in a positive residual. These residuals are expressed as fractions of a National Curriculum level (or GCSE grade) and are therefore relatively easy to understand if not to calculate.

41. There is at least one precedent for such an approach in the SAT (formerly Scholastic Aptitude Test), widely used and respected in the USA. Before 1941 this test was cohort referenced but after that date comparison was made each year with the cohort of 1941 thus making the test norm referenced.*

42. It is not likely that in the more rapidly changing modern world a reference group could be retained for more than sixty years, and the measure could be rebased from time to time, but it is envisaged that the same base should be used for, say, a decade at a time in order for systemic progress to be observed.

43. As far as the secondary phase of education is concerned ASCL suggests that the main measure of progress should be that between Key Stage 2 and GCSE. This reflects the most common patterns of secondary organisation with 11–16 and 11–18 schools. In such schools measures involving Key Stage 3 are clearly subsidiary and should be primarily for internal use rather than used to rate the school as a whole.

44. ASCL will be pleased to help develop these ideas further as part of its commitment to more intelligent accountability.

F. Progression Premium

45. This idea is particularly unwelcome. Teachers and their leaders are motivated by a desire to do right by those in their charge, not by a desire for a bonus. Such a premium, especially one built upon a measure that lacks full professional confidence, would either reward in a capricious fashion or would systematically reward those that need no such reward (ie those schools teaching the best supported pupils with the fewest disadvantages).

46. ASCL would remind ministers that the School Achievement Award was scrapped for very good reasons. They should not seek to reintroduce a similar, but equally flawed, reward.

47. ASCL strongly suggests that this idea be dropped forthwith.

G. Conclusion

48. ASCL welcomes the basic idea of *Making Good Progress* and its aspirations. School and college leaders have always striven to help all their students achieve as much as they can.

49. However, the actual proposals contained in the document, especially those in section five, would not help in any way to do this, and would in fact do far more harm than good.

50. ASCL would strongly suggest that the whole of sections five and six, and some of section three as outlined above, should be set aside. There is a need for some genuinely new thinking about these important matters so that the whole system of assessment, testing, reporting and accountability can be amended to better serve the worthy aims of *Making Good Progress*.

51. ASCL stands ready to contribute to such thinking and trusts that the major amendments proposed above should be made before the pilot begins.

January 2007

* *Thanks to Professor Dylan Wiliam, Deputy Director of the London Institute of Education for the explication of the difference between cohort referenced and norm referenced, and for drawing attention to the norm referenced nature of the SAT. Norm referenced is often used as opposed to criterion referenced to describe measures that would more properly be called cohort referenced. To be clear: in this paper norm referenced means comparing the performance of a child or group of children with a fixed reference group (say those that took tests in 2005) whilst cohort referenced means comparing a child or group of children with those taking the tests in that year.*

Reference

Wiliam, D. (2007). Balancing dilemmas: traditional theories and new applications. In A Havnes & L McDowell (Eds), Balancing dilemmas in assessment and learning in contemporary education (pp 269–283). London, UK: Routledge. (To be published later this year).

Supplementary memorandum submitted by the Association of School and College Leaders (ASCL)

Further to its previous submission, ASCL would like to make the following points:

1. The over-assessment of young people in England is causing them considerable stress and this is incompatible with the first outcome of *Every Child Matters* (being healthy). The stress is evidenced by international comparisons (for example, by UNESCO) of the happiness and wellbeing of young people, in which England has not featured at all well.

2. Many of the costs of external examinations were highlighted in our original submission. We would like to add further that the supervision of the large number of additional support staff required to make the current examination systems work has fallen upon senior staff. This is an additional load in itself and an indirect cost not included in our previous paper.

3. The cost of the national testing regime is wholly disproportionate to the gains made at intermediate key stages, most especially (in secondary schools) at Key Stage 3. In looking for efficiency savings, the Government should be aware that there is an obvious return to be had in streamlining over-costly and only marginally useful tests at this stage.

4. Age-related tests are also antagonistic to the personalisation "test when ready" philosophy underpinning *Making Good Progress*.

5. We would like to see the development of student portfolios of work in which objective e-testing components are a part. The e-testing would happen when an individual is ready, so would be spread across a school year and not prove unmanageable (as the earlier ICT e-Tests at KS3 proved to be for many schools).

6. The portfolios should be moderated by accredited Chartered Assessors, as suggested in our original paper.

7. Chartered Assessors should have an obligation to moderate other schools and, in turn, be moderated by others of their rank (thereby avoiding any conflict of interest).

8. There are current and successful role models for this approach—most obviously BTEC at all levels post-14 and is, we understand, to be QCA's recommended assessment regime for the new Diplomas.

9. The implication of this is that a Chartered Assessor would be required in each school for each of the core subjects currently tested at Key Stage 3 (English, maths and science). Such a less costly system could further be extended to ICT, thereby helping to ensure that 14-year olds reached functionality in the four major areas of the curriculum.

10. Functional skills should be a subsumed part of GCSEs. Functionality should be assumed to be achieved by an individual securing a GCSE grade C or better on papers that have been designed to include that as an objective. They should also be available as stand-alone qualifications for those not expected, or subsequently proven unable, to reach that level in the standard school examination. We understand this to be the intention for English, maths and science but, as ICT GCSE is generally an optional subject in KS4, functional skills must be tested in some other way. The portfolio approach would lend itself to this assessment, particularly in the many schools in which ICT is consciously taught as an embedded part of the whole curriculum.

December 2007

Memorandum submitted by Association of Teachers and Lecturers (ATL)

EXECUTIVE SUMMARY

ATL outlines the current excessive burden imposed by the current assessment and examination system particularly under the yoke of performance league tables, and gives a brief history of the system's development.

Using research evidence, ATL finds:

- That the data provided by the testing and examination system is compromised by the number of purposes for which it is used.
- That these purposes can be met through a system of cohort sampling, with evidence that this works in other countries.
- That the current high-stakes national assessment and testing system:
  - narrows the curriculum and reduces flexibility in curriculum coverage;
  - undermines the *Every Child Matters* agenda;

- has a negative impact on pupil attitude; and
- depresses staff morale and leads to "teaching to the test".

— That the current system of Key Stage tests:

- leads to duplication of testing between stages, particularly between Key Stages 2 and 3;
- provides data which is not used by teachers upon which to build further learning;
- does not accurately reflect changes in performance over time;
- does not provide valid information about students' attainment;
- undermines Assessment for Learning approaches;
- produces performance levels that are not sustained;
- assesses a limited range of skills;
- measures schools on indicators that are not only too narrow but are damaging to learning;
- leads to a narrow teaching focus; "teaching to the test";
- excludes many higher-level cognitive skills; and
- produces simplistic grades which often of little value in diagnosing learner needs.

ATL proposes a fundamental change to the assessment system, where we propose assessment for learning as the primary method of assessment throughout pupils' learning careers in a league-table free environment that uses cohort sampling to provide data for national monitoring purposes.

ATL believes that there should be no national assessment system prior to a terminal stage and international evidence links high pupil achievement to such systems which postpone national assessment and selection.

ATL outlines the need for schools to provide their students with the skills, understanding and desire for lifelong learning, something which the narrowness and high-pressure of the current assessment system may prevent.

ATL believes that assessment for learning principles and practice should underpin teacher assessment which should be, in the main, formative. This submission provides a wealth of research evidence about assessment for learning (AfL) and teacher assessment in the following areas:

- the positive impact of AfL on standards;
- the tension between AfL and summative assessment;
- personalised learning and AfL;
- AfL and the measuring of achievement;
- how AfL's vision of learning and ability is undermined by age-dependent levels;
- teacher assessment and the needs of a diverse school population;
- perceptions of bias in teacher assessment;
- resource needs of AfL; and
- workload implications of teacher assessment and AfL.

ATL strongly believes that this proposed system cannot exist alongside performance tables which already have a pernicious effect on the current national testing system.

ATL's recommendations for action are for the Government to do the following:

Initially:

- Review the current assessment system with urgency in light of its impact on curriculum coverage and on teaching and learning.
- Investigate the purposes applied to the present national assessment system.
- Develop AfL pilots in schools exempt from national testing during the pilot period.
- Prioritise CPD for teachers in assessment, particularly AfL techniques and strategies.
- End the use of national testing as market information and accountability mechanisms.
- Explore options of cohort sampling to meet national monitoring needs.
- Work with awarding bodies to produce a national bank of test materials as resources for teachers.

— Abolish school performance league tables.

— Explore alternative options to age-dependent levels.

And ultimately:

— Postpone national testing until a terminal stage.

ATL—Leading Education Union

1. ATL, as a leading education union, recognises the link between education policy and our members' conditions of employment. Our evidence-based policy making enables us to campaign and negotiate from a position of strength. We champion good practice and achieve better working lives for our members. We help our members, as their careers develop, through first-rate research, advice, information and legal support. Our 160,000 members—teachers, lecturers, headteachers and support staff—are empowered to get active locally and nationally. We are affiliated to the TUC, and work with Government and employers by lobbying and through social partnership.

2. ATL has recently produced *Subject to Change: New Thinking on the Curriculum* which questions whether our current curriculum and assessment systems are fit for purpose for the needs of society and our young people in the 21st century. This submission is based on these very arguments and we strongly welcome this Inquiry into testing and assessment, particularly around areas which challenge the efficacy of current national arrangements such as Key Stage testing.

Current Excessive Assessment—The Historical Picture

3. Our current pupil cohorts experience years of national assessment and testing; if you count foundation stage assessment, a pupil who goes on to take A-levels will have undergone national assessments and tests in seven of their 13 years of schooling. Yet prior to 1988, pupils faced only two external national tests—GCSEs and A-Levels—and a system of sample testing existed, which was overseen by the Assessment Performance Unit (APU). During that time, teachers had the power to design and carry out assessment for pupils not yet undertaking GCSE or A-level exams.

4. New arrangements for testing and league tables, including the assessment of all pupils by statutory assessment tasks and tests in core subjects at the ages of seven, 11 and 14 (at the end of Key Stages 1, 2 and 3 respectively) set up by the 1988 Education Reform Act have had, and continue to have, a huge impact on the primary and early secondary curricula as taught in schools.

5. 14–19 debates around curriculum and assessment have often concentrated on the issues of GCSE and AS/A2 provision with a resulting focus on the tensions between academic and vocational qualifications and the demands of external examination processes. The focus on difficulties of delivery has narrowed the debate and future thinking. For example, the 14–19 Diplomas, currently in development, from starting with a vision of integrating academic and vocational strands is becoming increasingly mooted as a vocational-only learning route due to the requirements of most stakeholders bar one, the learner.

6. The introduction of league tables of school exam and national test results through legislation in the 1990s has had an enormous and detrimental impact on the effects of the national testing regime in schools and has encouraged a risk-averse culture there. By placing such emphasis on "standards" as evinced through test results, league tables have encouraged "teaching to the test" and the regurgitation by learners of key "facts" leading to "surface" or "shallow" learning.

7. These measures represent a significant increase in the accountability to government of schools, teachers and learners concerning their performance, creating an imbalance between professional autonomy, professional judgement and accountability where the latter has assumed a disproportionate part of the experience of being a teacher.

The Data Machine—A Centrally Run System of Testing and Assessment

8. What the current centrally run assessment and testing system does give us is a large amount of data on pupil attainment and school performance; indeed at times, this seems to be its primary raison d'être. However, ATL questions whether that data in itself is helpful or useful enough to offset the detrimental effect it is widely acknowledged to have on the teaching of the current curriculum. The Daugherty Assessment Review Group in Wales, reviewing assessment arrangements at Key Stages 2 and 3, considered whether the "hard data . . . on pupil attainments and the targets it gives some pupils to aspire to, is of sufficient value to compensate for the evident impoverishment of pupils' learning that is occurring at a critical stage in their educational development".[1] Their conclusion can be inferred by their recommendation to the Welsh Assembly that statutory National Curriculum testing of 11 year olds at Key Stage 2 and 14 year olds at Key Stage 3 should be discontinued.

9. "*While the concept of summative assessment may be simple, the uses of data from summative assessment are varied and the requirements of different uses make varying demands in relation to reliability and validity of the assessment".*[2]

As outlined above by the Assessment Reform Group, the different uses of summative assessment data has a significant impact on its rigour and its fitness for purpose. Newton (2006) lists 18 uses for this data, currently:

1. Social evaluation
2. Formative
3. Student monitoring
4. Transfer
5. Placement
6. Diagnosis
7. Life choice
8. Qualification
9. Selection
10. Licensing
11. School choice
12. Institution monitoring
13. Resource allocation
14. Organisational intervention
15. Programme evaluation
16. System monitoring
17. Comparability
18. National accounting[3]

10. ATL questions whether one system can be fit for all these purposes. In terms of assessment, we understand validity to be the extent to which any assessment succeeds in measuring what it originally set out to measure. However, a plethora of purposes means that in fact we are measuring many other things in addition to the original focus of that assessment; for example, the aggregation of pupil's grades into broad level for the purposes of monitoring pupils, schools and systems will impact on the formative purpose of the assessment, making the outcome far less meaningful. Swaffield (2003) relates this to the notion of consequential validity: "This means that even a well-constructed test is not valid if the results are used inappropriately—which moves the idea of validity on from something which is the concern of test writers to something which is the responsibility of everyone who interprets and uses assessment results".[4]

11. ATL believes that clearer distinctions need to be made between the respective uses and purposes of assessment. Other countries' systems make this distinction clearer; strategies used include those which combine teacher led formative assessment with the utilisation of a national bank of tests applied for summative purposes when learners are ready. National monitoring needs are met through a system of sampling pupils' performance (eg cohort sampling), thus reducing the overall test burden whilst increasing the relevance and breadth of the learner evidence. While there is an economic advantage of collecting readily-available achievement data, eg the results of end-of-Key-Stage tests, we will demonstrate, throughout this submission, the lack of useful and relevant information it provides. If monitoring was separated from the performance of individual pupils, there would be no need for the central collection of individual pupil assessment data. As the Assessment Reform Group conclude, "this would remove the 'need' for high stakes testing and would ensure that assessment—and, more importantly, what is taught—was no longer restricted to what can be tested. The continuation in several countries of regular surveys of small random samples of pupils indicates the value of this approach".[5] In addition to the US National Assessment of Educational Progress (NAEP), there is New Zealand's National Education Monitoring Project (NEMP) and nearer to home, the Scottish Survey of Achievement (SSA).

Lessons from Across UK and the International Scene

12. The Scottish Survey of Achievement could provide a useful model for further investigation into restoring the place of teachers to the heart of curriculum and assessment. From the end of 2002–03, a new system of assessment in Scotland has been introduced. Teachers there have been provided with an online bank of assessment materials, based on the Scottish Survey of Achievement. The aim of these tests is to confirm the teachers' assessments of their pupils' attainment. These are to be administered to pupils when teachers deem they are ready to take them, rather than at a pre-determined time, making testing far more manageable within the school system and less likely to distort teaching and learning. Teachers have been supported in this process by the Assessment is for Learning (AiFL) programme. This has not led to any lack of accountability in the system; HMIE produce full reports on schools, based around a set of 33 quality indicators in seven key areas and the system strongly encourages schools to continually self-evaluate and assess achievements using these quality indicators. The Scottish Survey of Achievement also provides national figures, thus offering a way of measuring national progress over time without testing every child. The AiFL programme is being fully integrated into the national assessment system. In England, Assessment for Learning (AfL) still appears to be a separate strand from the national testing system, rather than an integrated part of a coherent whole.

13. International comparisons prove particularly interesting when we constantly hear of rising standards. Indeed, test results are improving, yet our international standing is falling in terms of our place on international league tables as evidenced by trends demonstrated in the PISA/OECD surveys. The UK's standing on international league tables for 15 year olds has slipped; although the UK's response rate to the 2003 PISA/OECD survey was too low to ensure comparability, the mean score that was produced was far lower than that achieved in the 2000 survey, leading to a fall in ranking within the OECD countries alone, a drop in place further increased by the inclusion of non-OECD countries within the survey.[6]

Impact of High-stakes National Testing and Assessment

14. A central proposition to the introduction of the national curriculum in 1988 was the entitlement of pupils to access a broad and balanced curriculum. However, the amount of high-stakes testing has had a well-documented narrowing effect on the curriculum, undermining this entitlement for many pupils, particularly in schools fearful of low scores on the league tables.

15. *Narrowing curriculum and reducing flexibility*

Webb and Vulliamy, carrying out research commissioned by ATL, document this effect in the primary sector; the standards agenda, through national curriculum testing in English, Maths and Science at various key stages and related performance league tables, "focused teachers' attention on curriculum coverage in literacy, numeracy and science to the detriment of the rest of the primary curriculum".[7] However, it is not just teachers and their representatives who are expressing this concern; Ofsted state in their 2005 evaluation of the impact of the Primary National Strategy in schools that the raising standards agenda has been the primary concern of most headteachers and subject leaders coupled with a far more cautionary approach in promoting greater flexibility within the curriculum. Ofsted also recognises the narrowing effect which Key Stage 2 tests have on teaching of the curriculum, in terms of time and also in terms of support for earlier year groups.[8]

16. *Undermining the Every Child Matters agenda*

The negative impact of current assessment mechanisms is not only diluting the principles of the curriculum vision of 1988, it is undermining the current *Every Child Matters* agenda. The longitudinal PACE project in primary schools in England observed that curriculum and testing pressures appeared to be "diminishing the opportunities for teachers to work in a way that enables them to 'develop the whole child' and address the social concerns of the wider society".[9] The Assessment Reform Group notes the lack of correlation between "the narrow range of learning outcomes assessed by tests . . . with the broad view of learning goals reflected in the DfES *Every Child Matters* policy document".[10] This tension at school level between narrow standards and school goals of engendering pupil enjoyment and creativity was strongly expressed by the headteachers who took part in ATL's research by Webb and Vulliamy.

17. *Impact on pupil attitude*

And what effect does this "tension" have on our pupils? A view across schools and colleges, observed by researchers, is that pupils have become very utilitarian in their views of what is "worthwhile to pursue"; Ecclestone and Hall (1999) call this a " . . . strategic and cynical compliance with assessment requirements" where passing tests is the primary focus and learning is "marginalised".[11] This is hardly surprising when we consider the high-stakes purposes of individual assessment data in our current system and the sheer volume of assessment which each pupil will face. But there are other pupils for whom such a utilitarian approach is not a possibility; for lower-achieving pupils, research has shown that the experience of frequently failing tests is demoralising, reducing self-esteem, including their belief in their ability to succeed with other tasks.[12] Thus, the gap between higher and lower achieving pupils widens, exacerbated by the fact that focus on test outcomes reduces the levels of early identification of under-achievement and appropriate interventions as noted by Ofsted in relation to the impact of Key Stage 2 testing.[13]

18. *Impact on education staff*

ATL's members, teachers and support staff, with pupils, are bearing the brunt of the testing overload and the high-stakes pressure. They are frustrated by the narrowing of the curriculum and the need to ready pupils for ever-increasing numbers of tests. This pressure encourages/drives many teachers to be complicit with the "strategic and cynical compliance" of students mentioned earlier and to be "presenters of content" to ensure that their pupils succeed in the narrow focus of the tests and that the school receives a good ranking on the performance tables. This process is ultimately de-skilling; an enforced focus on performance outcomes lessens and undermines richer assessment skills and feedback and will ultimately weaken these skills within the profession.

How Effective are the Current Key Stage Tests?

19. Key Stage tests are effective in producing a vast quantity of data on pupil performance as defined by the tests. However, we have earlier addressed the issues of validity around this data, particularly in regards to the myriad of uses to which it is put. Research has shown that Key Stage tests lead to a narrowing of curriculum, and within high-stakes frameworks which include school performance league tables, to "teaching to the test" and a destructive emphasis on testing rather than learning. To further explore the question, it is necessary to address the following issues which investigate this notion of their effectiveness.

20. *Limited value of test result data for further stages of learning*

An issue with the testing system currently in use is the limited value of its data for further stages of learning. The evidence for this is particularly strong in the transition between Key Stages 2 and 3. Many secondary schools carry out their own testing of Year 7 pupils in the autumn term, "a considerable duplication when pupils have already been assessed in most aspects of the core subjects at the end of Key Stage 2". (Ofsted)[14] It was also one of the main findings of the PPI survey, commissioned by ACCAC in 2002, that secondary schools did not make extensive use of the statutory assessment data available to them.[15]

21. *Do they adequately reflect levels of performance in children and schools, and changes in performance over time?*

Many of the purposes of assessment data can be linked to the standards agenda. Government is particularly concerned with proving through that agenda that their emphasis on, and investment in, education is resulting in rising standards over time. Pupils' grades in national curriculum tests and exams are, indeed, improving over time. However, Wiliam (2001) argues that any attempt to measure standards of achievement over time is "doomed" as we are not comparing like with like; what is taught in schools changes even if the official curriculum does not. We have already observed the evidence of growing focus on test-preparation and on teaching those subjects, or indeed aspects of subjects, which are tested to the detriment of untested aspects or subjects. Wiliam argues that the idea of measuring standards over time "in any real sense is nonsense" and that "while reported standards may rise, actual level of achievement could be falling—tests are no longer an adequate proxy for achievement across the whole domain".[16]

22. It is particularly those purposes which add high-stakes contexts to assessment that limit the value of achievement data. Tests do not usually test the full range of what is taught and in low-stakes contexts that limited range of achievement can indicate achievement across the whole subject.[17] Yet we know that once assessment occurs within high-stakes contexts, there is pressure on the school and the teacher to focus on the student's performance on the aspects of the subject likely to be tested—within an overburdened curriculum, those aspects will, inevitably, be given more time. Any such concentration of resources will inevitably mean that breadth, and indeed depth, of subject coverage will be sacrificed to the relentless pressure of targets, standards, tests and league tables. The purpose of assessment as an aid to the development of learning is shunted into second place.

23. Harlen and Deakin-Crick (2003) concluded from their research that current high-stakes testing does not provide valid information about students' attainment due to the narrow focus of tests and the consequences of being taught to the test leading to many students not actually possessing the skills or understanding which the test is designed to assess; the focus of teaching in this environment is to teach students to pass tests even where they do not have the skills or understanding.[18]

24. *Do they provide assessment for learning (enabling teachers to concentrate on areas of a pupil's performance that needs improvement)?*

A definition of assessment for learning which centres around its purpose and focus describes it thus; "assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils" learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence".[19] This definition demonstrates that Key Stage tests with their current emphasis on ranking, certification and accountability do not provide assessment for learning. Many good teachers use an assessment for learning approach working with learners to gather and interpret evidence to use to discover "where the learners are in their learning, where they need to go and how best to get there".[20] However, the pressures of test preparation and the importance of grade achievement have made it a secondary or "add-on" practice in many schools and classrooms.

25. *Does testing help to improve levels of attainment? Fall-off in pupil performance from Y6 to Y8 due to "hot housing"*

We hear all the time that standards are improving; ATL questions whether this means that our pupils are learning more and better. Research would suggest otherwise. Durham University carried out research, commissioned by the DfES, which noted the lack of evidence to show that pupils reaching Level 4 at Key Stage 2 will retain their learning, let alone progress to higher learning. They cite a study by Watson (2002) which showed how a "level focus" and booster classes temporarily raised pupils to mathematics Level 4 but that was not sustained over a period of six months to a year. Not only were learning outcomes not sustained but the Durham university report also details how high stakes assessment encourages a more rigid teaching style which disadvantages and lowers the self-esteem "of those who prefer more active and creative ways of learning".[21]

26. The current system is perceived as a selection system by pupils. The totemic importance of Level 4 at Key Stage 2 is now so huge that pupils who fail to achieve it cannot be blamed for feeling just that—failures. And we know that this is just how many of them do feel. We also know the effect this has on their future attitudes to learning. It is therefore no surprise that there is a dip in performance between Year 6 and 7. The policy of a differentiated offer post age-14 makes the Key Stage 3 tests an even clearer selection mechanism, determining how pupils' "choice" is to be "guided".

27. *Are they effective in holding schools accountable for their performance?*

Whilst ATL must again question the notion of effectiveness in this context, we acknowledge that Key Stage tests are "effective" in holding schools accountable for aspects of their performance, ie the performance of pupils in Key Stage tests. However, the cost of this excessive accountability is high. An IPSOS Mori poll in October 2006 found that the current target-driven culture was one of the top factors to demotivate teachers. Also, Key Stage tests are holding schools accountable for their performance across only a part of the curriculum; we have already documented research evidence around curriculum narrowing, the lack of sustainability of learning into subsequent key stages, the negative impact on attitudes towards learning amongst students and the lack of evidence of real attainment across the whole subject or curriculum.

28. *"Teaching to the test"—the high-stakes nature of test results leads to narrow teaching focus*

Despite the earlier mentioned demotivating effects of working within the current national assessment system, teachers are working so that their pupils have the opportunity to succeed within those same systems. There is strong evidence that rising test scores are not caused by rising standards of achievement but are rather the effect of growing familiarity amongst teachers and students with test requirements; research shows that changes in the tests are accompanied by a sudden fall in achievement, followed by a rise as teachers begin "teaching to the new test".[22]

29. *National curriculum tests and exams as assessment measures*

National curriculum tests and exams have long struggled to produce assessment instruments of high validity with optimum reliability and coursework and teacher assessment are examples of their attempts to ensure greater validity. However, these were add-ons, expected to fit in around the testing/examination system and thus were compromised in value and in practice. We have already noted that the limited coverage possible in tests combined with a high-stakes environment has a corresponding curtailing effect on the taught curriculum in schools. However, the format of the national tests which are written tests of limited duration also "excludes many of the higher-level cognitive and communication skills and the ability to learn both independently and collaboratively".[23]

30. Proponents for exams cite their objectivity, an assertion which needs to be briefly examined before we move onto a viable alternative. Public examination grades are not exact measures; they are approximate with known margins for error. These grades depend upon the judgements of examiners, who though very often highly professional, skilled and experienced people are also fallible human beings. Grades depend on snapshots of student performance under very particular conditions, at a certain point of time and in response to a certain set of assessment tasks. And e-assessment will not remove these features—it may bring many advantages of efficiency but "it won't by itself eliminate grade uncertainties".[24]

31. In addition, the needs of many of the assessment purposes outlined in paragraph 9 for simplistic grades mean that that much useful information about actual performance is lost. Sue Swaffield warns of the limitations of this data: "Summary statistics are often used to compare individual pupils or schools. In doing so, it is important to remember that any single score or level could have been arrived at from a wide variety of individual judgements, and so a level or grade gives no specific information about a pupil's performance. Much more information is needed if teachers in the next year group or school are to build upon pupils' prior attainment".[25] Furthermore, there is a danger that we "fail to appreciate the impact of test unreliability" (it is likely that the proportion of students awarded a level higher or lower than they should be because of test unreliability is at least 30% at KS2, for example) on the "reliability of change scores for individuals"[26] hindering diagnosis of a learning problem, should one exist.

32. Standardised tests can also obfuscate the meaning of pupil performance. For example, many tests offer multiple choice options to the pupil but these can confuse a reader who understood the text perfectly but was confused by the similarity of the choices offered—not by the text.[27] Without the teacher there to mediate, clarify and feedback the learning to the pupil, we, and they, lose the meaning and ultimately, it is the learner who loses out.

ATL Vision for the Future

33. *Change at a fundamental level*

ATL is arguing for a fundamental change in the assessment system; it is not enough to hand over the administration of summative assessment to teachers within a high stakes context and expect real advances in pupil achievement and engagement. Otherwise we are in danger of merely adding workload to teachers with no real addition in terms of professional autonomy nor a move to assessment which puts learning in first place. This fundamental change means that we are proposing assessment for learning as the primary method of assessment throughout the career of pupils in a league-table free environment that uses cohort sampling to provide data for national monitoring purposes.

34. *No national assessment system prior to terminal stage*

Due to the here- and elsewhere-documented detrimental effect of national curriculum testing on teaching and learning, ATL believes that there should be no national assessment system prior to a terminal stage. We believe that the present and future needs of our society requires an assessment system which focuses learners on learning rather than tests, maintains the breadth which was part of the vision of the National Curriculum in 1988 and which encapsulates part of the current vision for *Every Child Matters*, and which engages learners as participants in their learning and progress.

35. It can be argued that a system which postpones summative assessment at a national level fits within the earlier recommendations of the Task Group on Assessment and Testing (TGAT). The original vision of TGAT was for "an assessment system designed for formative purposes" which "can meet all the needs of national assessment at ages before 16 . . . only at 16 does it seem appropriate for assessment components to be designed specifically for summative purposes (paragraph 26)".[28]

36. International evidence now clearly links high pupil achievement with systems which postpone national assessment and selection. Finland's education system is a strong example of this as it is one which has gained it a high (often first) place on the OECD Programme for International Student Achievement (PISA) surveys of 2000 and 2003 with top ranking scores in mathematics, problem solving, science and reading and it defers national testing until a terminal stage. In fact, not only did Finland's students score highly in terms of performance and proficiency, but they demonstrated positive attitudes towards learning as this excerpt from the Executive Summary of the 2003 survey indicates: "For example, more than half of the students in France and Japan report that they get very tense when they have to do mathematics homework, but only 7% of students in Finland and the Netherlands report this. It is noteworthy that Finland and the Netherlands are also two of the top performing countries".[29]

37. *Focus on learning*

Across subjects, there are two key sets of goals: that pupils learn with understanding (develop understanding of concepts which can be applied in different contexts, identifying the links between different situations, applying the learning); and, understanding learning (that learners develop awareness of the process of learning). ATL has argued, and indeed it is widely recognised, that "students cannot learn in school everything they will need to know in adult life" [OECD, 1999][30] and therefore, schools must provide "the skills, understanding and desire needed for lifelong learning". This means that we need to look critically at our assessment systems, which have a huge influence on what is taught in the classroom and as we have demonstrated earlier in this submission, our current assessment system produces "strategic and cynical" test-takers rather than engaged and questioning lifelong learners with the flexibility needed for a rapidly changing society.

38. *Formative assessment, assessment for learning (AfL) and personalised learning*

ATL believes that assessment for learning principles and practices should underpin teacher assessment in schools and colleges. When assessment for learning (AfL) is talked of as a strong assessment model to support pupil learning and engagement, the formative aspects of assessment are highlighted, when evidence of pupil learning is used to identify learning needs and to adapt teaching work accordingly to meet them. The education community are fortunate to have an abundance of evidence to demonstrate the positive effects of formative assessment, even within the current system. Black *et al* (2002) answer the question, "*Is there evidence that improving formative assessment raises standards?*" with "an unequivocal yes, a conclusion based on a review, by Black and Wiliam (1998a), of evidence published in over 250 articles by researchers from several countries. There have been few initiatives in education with such a strong body of evidence to support a claim to raise standards".[31] They found that an increased focus on using formative assessment as principle and practice within the classroom produced gains in pupil achievement, even when measured in narrow terms such as national curriculum tests and examinations.

39. Research by the Assessment Reform Group endorses this finding regarding the weight of evidence that assessment for learning, with its formative assessment focus, has a positive impact on summative results, citing a quarter to a half GCSE grade improvement per student. However, their research does point

to the tension between assessment for learning and summative assessment which clouds the "improvement" focus of AfL, subsumed by information about successes and failures.[32] This argues for ATL's proposition that assessment for learning becomes the norm for teachers and pupils throughout the school careers of learners; it cannot fully realise its potential and vision within a system which has summative national tests and examinations at its core.

40. The advent of "personalised learning" on the horizon has brought AfL to the fore. This is unsurprising as assessment for learning is an approach which has the learning needs of individual students at its heart and is one which involves students far more directly in the assessment process. The DfES rightly sees the assessment for learning model as being school-based, collaborative, whole-school enquiry and yet this model cannot fit within a high-stakes assessment system which adds huge time and focus pressures to schools, creating a risk-averse school culture and through league tables, pits school against school. This is a fundamental flaw with the assessment for learning focus within the *Making Good Progress* project which will be hampered by its having to develop alongside more frequent national testing and targets.

41. *AfL requires a fundamental re-think in how we measure achievement*

This will require a culture change in schools and indeed, the wider community, about how we see achievement in schools. Many pupils and their parents will see learning tasks as competitions, achievement marked by a grade or a ranking within the class. One of the key problems with this "win/lose" view is that those who often lose no longer even try; better to switch off rather than risk "failure". Teachers working with researchers on formative assessment methods have found that "whilst pupils' learning can be advanced by feedback through comments, the giving of marks—or grades—has a negative effect in that pupils ignore comments when marks are also given".[33] Once grades were removed, pupils concentrated on the feedback given by the teacher and on how it could help them improve.

42. Research shows that grading and feedback have a big impact on pupil motivation and resulting willingness to engage in tasks and learning. Black *et al* detail key research findings on these effects:

- "Pupils told that feedback ' . . . will help you to learn' learn more than those told that 'how you do tells us how smart you are and what grades you'll get'; the difference is greatest for low attainers (Newman & Schwager, 1995);
- Those given feedback as marks are likely to see it as a way of comparing themselves with others (ego-involvement), those given only comments see it as helping them to improve (task-involvement): the latter group out-performs the former (Butler, 1987); and
- In a competitive system, low attainers attribute their performance to lack of 'ability', high attainers to their effort; in a task-oriented system, all attribute to effort, and learning is improved, particularly amongst low attainers (Craven et al, 1991)".[34]

This evidence shows that the returns for making this kind of change to how we assess learning will be significant, particularly amongst those who are currently losing out under the current system.

43. *Move away from age-dependent levels*

Target-setting, within the standards agenda, has led to a system of age-dependent levels. Again, researchers have argued that these mitigate against learning through an erroneous and demotivating belief about the nature of ability. Wiliam highlights the work of Dweck and her colleagues on students' views of the nature of ability and how that has a profound impact on how they react to challenging tasks. Those who see ability as a fixed entity, "how clever you are is how clever you stay" will tackle a challenging task if they believe their chance of success is high but will not engage if they believe that their chance of success is low. Those who see ability as incremental will see a challenging task as offering a chance to "get cleverer", ie to improve ability. As Wiliam observes, "in order to optimise the conditions for learning, it is therefore necessary for students to believe that ability is incremental, rather than fixed. A system of age-dependent levels would lead to a situation in which many students would get the same grade or level at ages 7, 11 and 14, thus potentially reinforcing a belief in ability as being fixed".[35]

44. *Teacher-led assessment and the needs of a diverse school population*

Our current curriculum and assessment models are based on the idea of "homogeneous knowledge to be owned by all". Shohamy (2000) observes this emphasis on homogeneous knowledge: "This is even more apparent in educational assessment. In a number of situations there is a gap between curricula and assessment as curricula may, at times, contain statements and intentions for the recognition of diverse knowledge, yet the tests are based on homogeneous knowledge".[36] It is not possible to de-contextualise assessment but ATL believes that local teacher-led assessment makes it possible to minimise the use of contexts which will have a detrimental effect on pupils' opportunities for achievement.

45. ATL believes that a fair assessment system is one which "elicit[s] an individual's best performance" and Gipps details the factors that need to be in place for assessment tasks or tests for this to occur: "This involves tasks that are concrete and within the experience of the pupil (an equal access issue) presented

clearly (the pupil must understand what is required of her if she is to perform well) relevant to the current concerns of the pupil (to engender motivation and engagement) and in conditions that are not threatening (to reduce stress and enhance performance) (Gipps, 1994). This is where teacher assessment can be more equitable since it is under the teacher's control (Gipps, 1994)".[37] Teachers are one of the parties who are in the best place to ensure that these conditions are in place and therefore teacher assessment is the method through which pupils have the opportunity to achieve to the best of their ability.

46. *Popular concerns regarding teacher bias: the evidence*

ATL acknowledges, in proposing a teacher-assessment focus using AfL, that there is a hurdle to be tackled in perceptions about teachers assessments. Harlen (2004) documents the "widespread assumptions that teachers' assessments are unreliable and subject to bias—despite their use in some countries as a main feature of national and state systems".[38] But Harlen goes on to pose ways in which that unreliability can be addressed; through provision of training around identification and understanding of assessment criteria by teachers and training which highlights sources of potential bias, as revealed through research.[39] Studies in Australia have shown that finer specification of criteria, describing progressive levels of competency, can lead to increased reliability of teacher assessment using assessment evidence from the full range of classroom activity.

47. The extent of evidence base for this perception regarding unreliability and bias is open to challenge. Harlen (2004) highlights a key concern with the process through which such a judgement has been reached in the past: "It should be noted that much of the evidence of bias in teachers' assessment comes mainly from studies where TA is compared with another measure and based on the questionable assumption that the benchmark measure is unbiased and is measuring the same thing as the teachers' assessment. So, whilst it has been reported that teachers under-rate boys more than girls in mathematics and science as compared with their performance in tests (Reeves *et al*, 2001), the conclusion might equally be that boys perform above expectation on mathematics and science tests".[40] Researchers have concluded that TA is prone to bias due to systematic variations between TA and standards task/test performance judgements, based on the assumption that the latter measures are unbiased. Yet bias in terms of gender, first language and SEN has also been found in the results of these standard tasks and tests so their original conclusion must be called into question. However, as we propose that teacher assessment, through assessment for learning, should be the only form of assessment throughout pupils' school careers, we acknowledge that bias and its effects must be a key part of training for teachers so that non-relevant assessment factors such as pupil behaviour and gender are recognised as potential sources of bias and influence and guarded against by teacher and moderators. The bias of unfamiliar situations is one which is a risk in national standard tasks and tests, a risk which lessens with teacher assessment.

48. *Resource needs of AfL*

Literature and research around assessment for learning yield a rich source of support, information and advice to teachers, through research observations, case studies and exemplifications of good practice. And much of that relates to involving the pupils to a far greater degree with their own learning in a conscious fashion combining subject/focussed skill learning with cognitive skills' learning. Teachers have access to examples of AfL techniques such as comment-only marking, peer and self-assessment, open questions that engage pupils and the promotion by the teacher of the liberating notion that wrong answers can as useful as right answers for learning, particularly with the exploration of ideas and concepts.

49. It is crucial that teachers are supported by training and resources. These resources can include exemplifications, concrete examples of good practice, diagnostic instruments, even task banks. Possibly most importantly, is the need for teachers to have space and time to collaborate to share examples of positive classroom experience (or perhaps examples of where/when things did not go so well), growing experience leading to fluency and efficiency with methods and to exploration of new ways of working with students. Students who are skilled and equipped to be self- and peer-assessors can check straightforward tasks. Sensitive and robust moderation procedures are a key part of this vision and here we can envisage a role for LEAs, consortia, clusters or networks of schools. Indeed each school needs to be an assessment community where assessment is something at the heart of each pupil's, each class's and each department's curriculum.

50. *Workload implications of AfL*

ATL is aware of the implications of this proposed assessment system in terms of new demands and workload. However, ATL believes that workload is not merely an issue of work level, it is also an issue of responsibility, autonomy, and professional satisfaction. It is important to remember that teachers already spend a large proportion of their time on assessment. Saving half of that time by removing or reducing the burden of national testing would more than compensate for the extra time needed for the embedding of assessment for learning practices and the process of moderation which is a vital component of it.

51. *Performance tables*

Assessment for learning does not lend itself to the narrow forms of data which currently feed performance league tables and ATL wishes to make it clear that the system which we have outlined would be negatively impacted by the continuation of these instruments of high-stakes pressure, particularly on schools and LEAs. Any such focus on narrow, hard data will undermine the learning focus of schools, and inevitably some schools will succumb to the pressure to conform to the rigid measures of the standards agenda. League tables also undercut any notion of collaboration between schools and yet any system which hopes to offer full and broad curricula and personalised learning, needs to promote cost-effective ways for schools to meet those needs through the sharing of resources, expertise and knowledge. This is not a form of accountability which promotes equitable access of opportunity to all and ATL has no hesitation in calling for its abolition—there are other, far more meaningful, forms of accountability and of school information.

Conclusion: Recommendations for Action

52. ATL's vision is for a system where assessment is a key part of learning, a central activity for teacher and pupil within a low-stakes context which does not create a culture of competition in which "losers" become demotivated or disengaged and in which teachers become empowered, further skilled and re-motivated.

53. ATL calls on the Government to:

Initially:

- Review the current assessment system with urgency in light of its impact on curriculum coverage and on teaching and learning.
- Investigate the purposes applied to the present national assessment system.
- Develop AfL pilots in schools exempt from national testing during the pilot period.
- Prioritise CPD for teachers in assessment, particularly AfL techniques and strategies.
- End the use of national testing as market information and accountability mechanisms.
- Explore options of cohort sampling to meet national monitoring needs.
- Work with awarding bodies to produce a national bank of test materials as resources for teachers.
- Abolish school performance league tables.
- Explore alternative options to age-dependent levels.

And ultimately:

- Postpone national testing until a terminal stage.

June 2007

Submission References

1. Daugherty Assessment Review Group (May 2004) *Learning pathways through statutory assessment: Key Stages 2 and 3. Final Report.*
2. Assessment Reform Group *Assessment Systems for the Future, Working Paper 1: Aims and outcomes of the first year's work of the project,* Draft 10 (2006).
3. Newton, PE *Clarifying the purposes of educational assessment* draft June 2006.
4. Swaffield, S (2003) *Assessment Literacy for Wise Decisions* Association of Teachers and Lecturers, London.
5. Assessment Reform Group (2006) *The role of teachers in the assessment of learning* London.
6. Organisation for Economic Co-operation & Development (OECD) *Learning for tomorrow's world: First results from PISA 2003* France 2004.
7. Webb, R and Vulliamy, G (2006) *Coming full circle: The impact of New Labour's education policies on primary school teachers' work,* Association of Teachers and Lecturers.
8. Ofsted (2005) *Primary National Strategy: An evaluation of its impact in primary schools 2004–05.*
9. Pollard, A, Triggs, P, Broadfoot, P, McNess, E, and Osborne, M. (2000) *What pupils say: changing policy and practice in primary education,* London, Continuum.
10. Assessment Reform Group *The role of teachers . . .*
11. Wilmut, J (2004) *Experiences of summative teacher assessment in the UK.*
12. Assessment Reform Group *Ibid.*
13. Ofsted (2005) *Ibid.*
14. Ofsted (2002) *Changing schools—effectiveness of transfer arrangements at age 11: an evaluation by Ofsted.*
15. Daugherty Assessment Review Group *Ibid.*
16. Wiliam, D (2001) *Level best? Levels of attainment in national curriculum assessment* Association of Teachers and Lecturers, London.

17. Wiliam, Ibid.
18. Harlen, W & Deakin-Crick, R (2003) *Testing and motivation for Learning* Assessment in Education, 10.2 169–208.
19. Black, P, Harrison, C, Lee, C, Marshall, B, Wiliam, D (2002) *Working Inside the Black Box* Dept of Education and Professional Studies, King's College, London.
20. Assessment Reform Group *Role of teachers . . .*
21. Beverton, S, Harris, T, Gallannaugh, F and Galloway, D, (2005) *Teaching approaches to promote consistent Level 4 performance in Key Stage 2 English and mathematics* University of Durham, School of Education.
22. Harlen, W (2004) *Can assessment by teachers be a dependable option for summative purposes?* Published in *Perspectives on pupil assessment* GTC, London.
23. Assessment Reform Group (2006) ibid *Role of teachers . . .*
24. Murphy, R (2004) *Grades of uncertainty: Reviewing the uses and misuses of examination results* Association of Teachers and Lecturers, London.
25. Swaffield, ibid.
26. Wiliam, Ibid.
27. Swaffield, ibid.
28. National Curriculum Task Group on Assessment and Testing. *A report* 1988, London: Department of Education and Science.
29. Organisation for Economic Co-operation & Development *Learning for tomorrow's world: First results from PISA 2003: Executive Summary* France 2004.
30. Assessment Reform Group *Role of teachers . . .*
31. P Black, C Harrison, C Lee, B Marshall and D Wiliam, (2002) *Working inside the Black Box* Dept of Education & Professional Studies, King's College, London.
32. GTC (2004) *The role of teacher in pupil assessment Perspectives on Pupil Assessment* GTC, London, November 2004.
33. P Black, C Harrison, C Lee, B Marshall and D Wiliam, (2002) *Working inside the Black Box* Dept of Education & Professional Studies, King's College, London.
34. Ibid.
35. Wiliam, ibid.
36. C Gipps & G Stobart (2004) *Fairness in Assessment Perspectives on Pupil Assessment* GTC, London, November 2004.
37. Ibid.
38. Harlen, W (2004) *Can assessment by teachers be a dependable option for summative purposes?* Published in *Perspectives on pupil assessment* GTC, London.
39. GTC (2004) *The role of teacher in pupil assessment Perspectives on Pupil Assessment* GTC, London, November 2004.
40. Harlen, W (2004) Ibid.

Memorandum submitted by the National Association of Head Teachers (NAHT)

General Issues

Why do we have a centrally run system of testing and assessment?

It is essential, first of all, to make the key distinction between *assessment* and *testing*:

> *Assessment* lies at the heart of all reaching and learning and is the fundamental professional activity of any teacher. It enables them to establish the performance and understanding of their students, to assist with ongoing learning and development.
>
> *Testing* covers the final, standardised awarding of an agreed qualification or level at a particular point. This applies to the SATs as well as to such qualifications as GCSEs, A levels etc.

It is where these two activities are not distinguished from each other that confusion and difficulties arise.

It must be recognised that the British centrally-run system of testing and qualifications at the end of compulsory education and beyond is respected internationally. Although there are ongoing difficulties in the way in which these qualifications evolve over time, there is no-one calling for the wholesale abolition of this highly valued system. However, the rationale for the current centrally-run test system stems for the Government's standards agenda, with its associated regime of targets, tests and league tables.

The current arrangements by which children are tested according to national tests are viewed as burdensome and damaging. A review of this system and of the narrow rationale securing it is of paramount importance.

What other systems are in place both internationally and across the UK?

Every school has its own arrangements for internal assessment, many highly praised during Ofsted inspections and many reflecting the skills of the teaching workforce. As part of the National Strategies, a focus on "Assessment for Learning" has proved to be of great value in enabling teachers to track and support students through their learning journey.

It is where these activities become directed solely to successful "passing the SATs" that they become weakened and potentially damaging.

In many of the countries who have been rated highly through such international projects as PISA, formal education begins later than in the UK, and there is no such systemised arrangement for formal tests. More recent information from countries such as Holland, Finland and Denmark suggests that there is a greater emphasis upon play and creativity at younger ages, formal schooling begins later, teachers have greater autonomy and the system of national testing and assessment is far less draconian, if it exists at all. Certainly there is no high stakes testing or publication of league tables and there is an acceptance that children develop in different ways and at different rates.

It is also worth noting that, in Wales, a decision was taken in 2005 to make Key Stage 2 tests optional and abolish league tables. Instead, the system is predicated on assessment of an individual's attainment and progress, rather than on accountability within the system, as in England.

Does a focus on national testing and assessment reduce the scope for creativity in the curriculum?

At its best, creativity releases the child from the rigid, formal framework of the national curriculum, to be able to explore and investigate in a holistic and practical mode, the wonders of the world around him or her. This approach, however, has to be extremely well structured and organised by the teacher and the school, as a framework of essential skills and knowledge, needs to underpin the curriculum so that the child is able to develop his or her creativity. The professional activity of ongoing assessment and understanding of a child's development will never reduce the scope for creativity. Rather, the encouragement by a skilled adult will nurture creative development of children through the early years.

If the time and energies of teachers, parents, and children are dominated by a narrow syllabus and a narrow range of activities which will be the subject of high stakes testing, we run the risk of this dominating the curriculum and this may well lead to a narrowing of opportunity. If children are straitjacketed by "teaching to the tests", whether this be at KS1, KS2 or KS3, there will not be time for the normal, essential creative development which needs to be a part of the whole educational experience.

Who is the QCA accountable to and is this accountability effective?

The brief of QCA is "to regulate, develop and modernise the curriculum, assessments, examinations and qualifications". It is described as "a non-departmental public body, sponsored by the Department for Education and Skills (DfES). It is governed by a board, whose members are appointed by the Secretary of State for Education, and managed on a day to day basis by an Executive Team".

In its regulatory capacity, its role is to ensure that the Awarding Bodies adhere to the clear rules relating to their examinations and, from time to time, conduct appropriate reviews of this work. It is for QCA to take on this role, to ensure that the trust which has built over time can continue. In this capacity, QCA is highly effective.

In terms of its role as developer and moderniser of the curriculum, QCA is extremely careful to involve all key stakeholders in its reviews and to use the expertise of the teaching profession, through a wide range of organisations. The integrity and skill of QCA officials is generally appreciated and respected by the education professionals.

The QCA is given clear remits relating to aspects of its work by the DfES and, where there can be frustrations expressed, it is largely because the remit does not necessarily give QCA sufficient freedom in aspects of its work. QCA offers sound professional advice to the DfES but the Secretary of State for Education is not bound to listen and follow this advice. However, there have been circumstances where QCA has offered strong recommendations for caution (eg over the abolition of coursework in GCSE) and the DfES has asked QCA to undertake further work.

QCA is generally effective but there are potential dangers in that it is so strictly controlled by the DfES that all it is empowered to do is offer advice.

What roles should exam boards have in assessment and testing?

The Awarding Bodies are highly respected for their work in ensuring that standards are maintained in external qualifications over time. In spite of recurrent negative publicity each August, there is evidence that employers, teachers, parents and pupils have great confidence in the qualifications that are offered and awarded. Even the GCE "A" levels have returned to their former status following the debacle of Curriculum 2000.

The ongoing development of e-testing, the development of the new diplomas, the support for teachers and students in working through the new GCSEs and other qualifications are aspects for which the Awarding Bodies are given due credit.

Questions about the role of coursework, the viability of the new Diplomas, the risks inherent in the greater use of the Internet for research and risks of plagiarism and the issues relating to the increased costs of examination entries for schools and colleges, all need to be viewed in the context of general recognition that the Awarding Bodies are successful as providers of a tried and tested system.

National Key Stage Tests

Current Situation

How effective are the current key stage tests?

The current Key Stage tests dominate the work of Primary schools and, for Secondary schools, during Key Stage 3. This is not healthy. As with any summative assessment system, the Key Stage tests only give a snapshot of a pupil's ability at a specific time and in a relatively narrow field.

The programmes of study and the range of the curriculum are not, in themselves, damaging, but the emphasis on the outcome of the tests means that the focus of much of the teaching, in particular in year 6 and in year 9 is on test performance and likely content. This is clearly insufficient and narrows the range of what is offered.

The current Key Stage tests are effective in testing the prescribed content and the schools' effectiveness in preparing children to undertake these tests. They are not effective in testing either pupils' broader range of educational achievement nor in testing the success of a school (except in its success in preparing pupils for the tests!) There is also a growing body of evidence that the plethora of testing "windows" is having a detrimental effect on individual children's health and well-being.

Do they adequately reflect levels of performance of children and schools, and changes in performance over time?

The Key Stage tests provide one source of helpful performance data for both students and teachers. Because the NAA draw on long-term, tried and tested skills which ensure that standards are maintained over time, the tests could be used as one broad indicator but it is hazardous to draw too many conclusions from the minutiae of the detail. A teacher's professional knowledge of the pupil is vital—statistics are no substitute for professional judgement.

As an overall national standard, statistically the tests are valid. Because of the small size of many of the individual school cohorts, where a single pupil may count for more than 15% of the overall score, the statistical validity of this data is severely limited. The tests only test one aspect of educational performance and need to be recognised as a single item of data, to be taken professionally alongside many other elements. Care needs to be taken over the interpretation of data—over-simplified interpretation can lead to flawed conclusions. Any use of data should be as an indicator, rather than a determinator.

Do they provide Assessment for Learning (enabling teachers to concentrate on areas of a pupil's performance that needs improvement)?

The Key Stage tests do have a value in giving teachers an indication of pupil performance and will provide some of the data which is helpful in enabling a teacher to understand the performance of the students. However, they only provide one measure and need to be treated in this respect.

Assessment for Learning is far broader than the Key Stage tests and information must be gleaned on an ongoing basis, from day to day course and schoolwork, and not from one measure, operated at identifiable points in a child's career, for which they may well have been overprepared. Assessment in the normal process presupposes the collection of information over a period of time rather than relying upon a snapshot of attainment, in order to ascertain where pupils are and plan where they need to go. Assessment for Learning is a broad principle, far wider than feedback from snapshot national tests and countless schools have developed sophisticated pupil tracking systems through it.

Are they effective in holding schools accountable for their performance?

The Key Stage tests represent only one measure of performance. Schools have a wide range of accountability measures, ranging from financial benchmarking through to full Ofsted inspections.

The development of the self-evaluation systems which take account of Key Stage test results, alongside other professional educational data, is far more reliable than the one-dimensional picture which is offered by the SATs. Schools now have the tools and are continuing to develop expertise and experience in self-evaluation and they need to be trusted to get on with the job.

How effective are performance measures such as value added scores for schools?

Value added measures are part of the rich array of professional data available to schools, local authorities, SIPs and Ofsted. To some extent they help to provide a context within which the narrow SAT information can be viewed. All elements of professional educational data has its place, but it is to be used in conjunction with other information, to pose hypotheses and lead to professional discussion about school improvement, rather than to make rigid judgements or be used to draw simplistic and potentially inaccurate conclusions. Whilst the principle behind value-added scores is reasonable, there is still disquiet about the validity of data in different contexts. Although the value-added data is in the public domain, its complexity is such that, at best, it remains meaningless to the majority of its readers. At worst, it is open to misuse and abuse.

Are league tables, based on test results, an accurate reflection of how well schools are performing?

League tables are hugely damaging to the educational system. They only use one of the many types of measures which should inform understanding of the context and the success of a school and its pupils. They should never be used to make simplistic comparisons between different schools, in different areas, teaching a different cohort of pupils. They should never be viewed as a total measure of any school.

League tables based on test results will only ever indicate how a school has enabled its pupils to perform in those particular tests and this can never give a full indication of how effective the organisation is in offering a wide, broad and appropriate education to those young people in its charge. Even modified by social deprivation or value added factors, they can only give a distorted snapshot of the work of a vibrant and organic community.

To what extent is there "teaching to the test"?

Because of the external focus on the results of SATs, there is far too much "teaching to the tests". Recent survey evidence indicates that, at year 6, for four months of the school year, schools are spending nearly half their teaching time preparing pupils for Key Stage 2 tests.

This has been actively encouraged by the DfES through the provision of "booster classes" and through the requirement to produce "intervention plans". These boosters and interventions have not necessarily been used as professional development plans for the wider education of children. Instead, they have had the prime focus of ensuring that a small identifiable cohort of children will be "boosted" to achieve a higher grade on the narrow range of work relating to particular tests.

This emphasis has narrowed the focus of the curriculum and introduced professional fear into the work of both headteachers and individual class teachers. A headteacher's or a Year 6 teacher's career can be blighted by a single poor performance (for whatever reason including the unfortunate absence of a couple of bright pupils). As referred to before, because of the relatively small cohort tied into any one school's results, the statistical validity of any test is flawed.

Very few teachers have the confidence to take risks and introduce dynamic and entirely appropriate rich activities with students approaching the SATs, if the content appears not to relate directly to that which will be examined.

How much of a factor is "hot housing" in the fall off in pupil performance from year 6 to year 7?

A pupil who has been coached emphatically and successfully to achieve a grade higher than they would naturally have obtained, may well, when coping with the pressures of transfer to a new and more adult environment, appear to have "slipped a level".

There is also a danger, reported by many professionals, that students may learn how to succeed in a particular type of test, which can give a distorted picture of their broader ability. There are many examples of year 6 students who have obtained high levels, particularly in Science SATs, who are not able to replicate this performance within the secondary curriculum. The results are not wrong. They merely indicate that the students have learned how to pass Science SATs and not developed scientific skills and absorbed scientific content. This can be extremely unhelpful for the receiving secondary school.

Another huge danger is that the "hot housing" may not be a stimulating activity and that this may have a damaging effect on the morale of the student. If booster classes and repetitive test practice activities are boring and continue to offer more of the same to the student, they are unlikely to foster a love of learning such as could be engendered by a rich and creative curriculum.

If pupils are not force-fed a diet of SATs, they may well also be able to prepare more broadly for the transition to the very different environment of secondary school.

Does the importance given to test results mean that teaching generally is narrowly focused?

Yes, see above. Recent studies have concluded that the standards agenda has focused teachers' attention to the detriment of the rest of the curriculum.

What role does assessment by teachers have in teaching and learning?

Assessment by teachers lies at the heart of all teaching and learning. Assessment may be formal and thorough, or brief and effective and undertaken through oral or other processes. Not all recent developments have been unhelpful in this respect: for instance, the teacher assessment component of national assessment prepared teachers for Assessment for Learning and the emphasis on personalised learning.

Every teacher is assessing the performance of his or her pupils at every point in the teaching and learning activity. It may not be formal; it may not be extensive but at every point, information about what a child knows or does not know is used by the skilled teacher in making decisions about the advice to be given, the encouragement to be given and the ongoing educational needs of the pupils. True personalised learning depends on skilled ongoing assessment by the teacher and on skilled self-assessment by the pupil.

It is vital that we do not confuse assessment with formal test structures.

THE FUTURE

Should the system of national tests be changed?

The tests themselves are not inherently the root of the problem. It is the emphasis and use of the results that has done and continues to do the damage. The high-stakes nature of the process is that which is leading to the skewing of the curriculum and the stress which is unhelpful and unhealthy for students and their teachers. The majority of our members do not have an issue with the principle of testing. The crucial issue remains the high stakes nature of the process and the emphasis on published league tables, coupled with the linking to inspection outcomes.

League tables need to be abolished and it needs to be recognised that SATs only offer one of many elements by which a school and its success should be evaluated. The current system needs to be changed. Whether or not the tests themselves need to be fundamentally revised is a totally different question.

If so, should the tests be modified or abolished?

League tables should be abolished as should the unhealthy emphasis on a single outcome measure.

If the current arrangements are significantly modified along the lines indicated above, a review of the content and style of the tests can be undertaken in a professional and non-emotional professional activity, through thorough and appropriate consultation with all interested parties. This consultation needs to be open and transparent, involving all interested parties and must look at the nature of and the rationale behind the continuation of testing.

The Secretary of State for Education has suggested that there should be a move towards more personalised assessment to measure how a pupil's level of attainment has moved over time. Pilot areas to test proposals have just been announced. Would the introduction of this kind of assessment make it possible to make an overall judgement on a school's performance?

The proposals included in the *Making Good Progress* consultation would lead to a different data set which can be used by schools. This would be different information which would have its own value used in a professional context. There is no reason to assume that this different data set would be any more accurate or any less damaging than the current data set if taken in isolation. Any *overall* judgement of a school's performance would be no more infallible and no less misleading than current information.

The new proposals are based on an assumption that a young child should make a particular path of progress at a particular rate. Children learn in different ways and at different rates. The underlying assumption, that there is an optimum and fixed rate of progress over time for all pupils, is flawed. The danger is that one inadequate measure may be exchanged for another. As stated previously, data provides an indication of knowledge and progress, it is not a definitive determinator.

However, as professional data, the information drawn about pupil performance from tests taken "when ready" will have significant value to the school and will fit with other elements of data to assist with school improvement, pupil support and true assessment.

Would it be possible to make meaningful comparisons between different schools?

No. If the pupil is put at the centre of learning, rather than maintaining the current system of school accountability, then the data gives assistance to the planning and developing of the learning for the pupil. It does not support the comparison between different schools.

What effect would testing at different times have on pupils and schools? Would it create pressure on schools to push pupils to take tests earlier?

It is not possible to guess with accuracy what the impact of the new style of tests might be. There will be schools where students are encouraged to take tests early. There may be other schools where students are encouraged to take tests at a later point when they are more likely to have perfected their performances in the named activities. Teachers, parents and students will learn the rules of the new game over time.

There may also be logistical difficulties in some primary schools if the testing has to take place over a longer period of time and there could potentially be greater costs and more disruption to the curriculum. Consideration must be given to the issues for pupils with special educational needs. The P levels used are not suitable for any summative approach.

If Key Stage tests remain, what should they be seeking to measure?

Key Stage tests should be used to test the skills itemised within the related programmes of study. They should be used within schools as internal professional data to assist in the process of individual pupil progress and overall school improvement. They should not be used to provide league table data.

It should be possible to develop a bank of external tests which can be used when a school feels that the pupil is ready. These tests should be linked to relevant programmes of study, should be skills-based and should be used solely for schools' internal data collection and analysis. This would enable cohort sampling to be built into this to help inform national trends from time to time.

If, for example, a Level 4 is the average for an 11 year old, what proportion of children is it reasonable to expect to achieve at that or above that level?

Children learn at different rates and in different ways. Some 11 year olds will have far exceeded a Level 4, whereas others may need longer to arrive at their destination. What is important is that schools encourage and support pupils to make the progress which they, as individuals, need to make. Local approaches to formative assessment and pupil progress measurements are, in most settings, highly effective. Schools are only too aware that children do not always progress in a regular, linear manner.

We must not label as failures 11-year-olds who learn more slowly or who have skills in different aspects which cannot be described in such concepts as "Level 4". What is a Level 4 Happiness or a Level 5 Social Responsibility? How can we expect a certain, arbitrary percentage to succeed or fail? More importantly, why should we?

How are the different levels of performance expected at each age decided on? Is there broad agreement that the levels are meaningful and appropriate?

The current descriptions and levels relate to one narrow aspect of the educational and curricular experience. If they are agreed to be criterion-referenced measures relating to specific programmes of study, then it is possible to decide which children have achieved the desired level. The mistake that is too often made is to assume that the output data relates to a far broader range of skills. It does not.

Testing and Assessment at 16 and After

Is the testing and assessment in "summative" tests (eg GCSE AS A2) fit for purpose?

The current "summative" tests and qualifications at age 16 and after are generally respected and regarded as fit for purpose. There are a number of modifications due to come into force from September 2008 and these have been the subject of professional consultation.

While there are some aspects which will continue to need to be modified to keep up with wider developments, generally GCSE, AS and A2 are not in need of major imposed revisions. Answers to other questions will give further information relating to those aspects which need to be kept under review.

Are the changes to GCSE coursework due to come into effect in 2009 reasonable? What alternative forms of assessment might be used?

The concerns and the media furore about coursework and the inevitable increase in plagiarism as a result of the accessibility of materials on the Internet have been largely out of proportion to the potential difficulties to the system and takes no account of the changed learning patterns and environment that students have today.

Where a student has copied large quantities of material from the Internet, the teacher is usually able to detect the fraud. Discrepancies in the pupils' style, poor blending of the plagiarised material with the student's own work, and teacher common sense will largely reduce the impact of this growing trend. It is not new. Pupils have always tried to use extraneous material (and where does research end and plagiarism begin?). English teachers have long been accustomed to challenging the inappropriate use of other materials in student essays.

The initial reaction to get rid of coursework was inappropriate and draconian. Thankfully, a more balanced approach has been adopted since and, by treating each of the different subject disciplines at GCSE in different ways, an appropriate solution appears to be on the horizon.

Coursework will always be an entirely appropriate and important part of any student's work throughout study. Whether or not the coursework becomes part of the summative test which gives the final grade for the qualification is another matter. It may be that coursework could be a part of the teacher assessed element. Alternative approaches are being considered as part of the consultation on coursework in conjunction with the Awarding Bodies and QCA.

What are the benefits of exams and coursework? How should they work together? What should the balance be?

Students need to be capable of undertaking independent research and study. Coursework, with varying levels of teacher intervention and assistance, is one of the best ways of ensuring that this can be undertaken. This is recognised and, as part of the Diplomas, an Extended Project is viewed as an essential element. This is entirely right.

To be so fearful of the dangers of plagiarism and the Internet would be to deny both teachers and students a vital part of the educational experience. The balance between coursework, assessed coursework and terminal examination will, quite rightly, vary from subject discipline to subject discipline.

Will the ways in which the new 14–19 Diplomas are to be assessed impact on other qualifications such as GCSE?

The new 14–19 Diplomas offer the opportunity for a radical and imaginative approach to assessment. Whether or not this opportunity will be taken remains to be seen.

The Extended Project, modular study, "when ready" testing and e-assessment are all aspects which will have implications for other qualifications.

However, it would be a mistake to regard the Diplomas as a completely new departure from conventional assessment. There have been for many years, innovative and varied forms of assessment in existing GCSE and A levels and it is hoped that the knowledge and experience of these can be a solid foundation for summative assessment in the future.

It is ironic that, as we remove the GCSE coursework from many of the subjects, we are seeking ways of assessing and evaluating Extended Projects at Level 2. One might ask, just what are the key differences between these two types of assessment?

Is holding formal summative tests at 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?

Until the formal leaving age is accepted as 18, it will be necessary to have some form of summative testing and qualification at age 16. GCSEs, Level 1 and 2 Diplomas and other suitable qualifications (which may include i-GCSEs) will need to remain until it becomes the norm for all students to proceed to education and training post 16. The Tomlinson report offered a widely respected and viable alternative but when this was rejected, the educational world had to return to ensuring that the current system was as effective as possible.

It will remain necessary to have a summative examination so that a reliable, standardised award may be given at the end of a Level 1, 2 or 3 course.

There are some subjects where there have been too many, too complex modules but these are the subject of further consultation. The question of re-takes is also under review. It is this which places too great a burden on students and takes them away from study and the course to focus on excessive examination. Generally, the existing system is fit for purpose.

To what extent is frequent, modular assessment altering both the scope of teaching and the style of teaching?

Frequent modular assessment is not new. In the early days of GCSE Mode 3, this became an excellent method of ensuring ongoing motivation for students for whom a terminal examination and traditional methods was not attractive.

The new Diplomas will contain considerable elements of modularisation and it is anticipated that these individual elements will have the possibility of being counted for different awards at different levels and in different combinations. The Minerva software, currently being developed, is intended to be the basis for the management of this new system.

Teachers have welcomed the moves towards modularisation because of the positive benefits in terms of motivation, and because students can achieve credit for key aspects of the course in spite of finding some parts of the final qualification too challenging or inappropriate.

If anything will assist the reintegration of some of the NEETs (young people not in education, employment or training) it will be the further, suitable development of modular, component assessment within the new vocational diplomas.

How does the national assessment system interact with university entrance? What are the implications for a national system of testing and assessment from universities setting individual entrance tests?

Universities have been worried about the rise in the number of students who achieve grade A in the A levels. They argue that this had made it more difficult to select the truly high achievers. Making the actual points level detail available to universities should have gone some way towards indicating which of the students are the highest achievers.

Whether or not it will be possible to introduce PQA (post qualification application) will depend on negotiations between Awarding Bodies, schools and universities on the question of timescales. If the universities can move their start dates back, it may be possible to complete the A level assessment before they make the firm offers. Moves to bring forward the A level results dates and curtailing the marking period for the Awarding Bodies will also assist with this.

It is to be hoped that universities will accept and welcome the new Diplomas. The Secretary of State for Education has urged them to join with the rest of the educational world in giving the new qualifications a fair and successful start. Some universities, however, will inevitably seek to develop their own admissions criteria and we must not arrest the new developments to pander to their views.

Far more worrying must be the trend of the independent schools to turn to alternatives such as the i-GCSEs and the International Baccalaureate. It will be essential that QCA and educational organisations work together to ensure that we have a consistent, coherent system of examinations and qualifications at the end of Key Stage 4 and at the end of compulsory schooling.

May 2007

Memorandum submitted by the General Teaching Council for England (GTC)

Executive Summary

The GTC hopes that the Education and Skills Select Committee (ESSC) will, as a result of this inquiry, urge the Government to undertake a fundamental and urgent review of the testing and assessment regime in maintained schools.

England's pupils are among the most frequently tested in the world, but tests in themselves do not raise standards. Tests are used for too many purposes and this compromises their reliability and validity. The tests can depress pupils' motivation and increase anxiety. They do not adequately serve the interests of parents or pupils and they lead to a narrowed curriculum and encourage "teaching to the test". The system diminishes teachers' professional judgements because summative outcomes reached by the teacher carry less public weight than the outcomes from end of Key Stage (KS) tests, although the received wisdom that KS tests and public examinations are error-free methods of assessing pupil attainment is misleading.

GTC's Proposals

Ongoing classroom assessment combined with a timely use of a nationally devised bank of tests

- Continued Government support for teachers' use of assessment for pupil learning to ensure it has maximum impact across schools.
- The development of a nationally-devised bank of tests/tasks to be used during the key stage when the teacher judges that the pupil/pupils are ready.
- Teachers overseeing all forms of assessment including the bank of tests, and their professional judgments on pupils' performance being given increasing weight over time.
- Increased Government investment in teachers' assessment skills.

School Improvement and Accountability—focused away from the centre and towards the community, parents and pupils

- The development of a richer dialogue between schools and parents based on enhanced information resulting from teachers' assessment of their pupils.
- An entitlement for parents to be fully and regularly informed about progress and attainment.
- Using the School Profile to communicate a broader range of school accountability information to parents.

Monitoring National Standards—a more cost effective and efficient system of collecting data

- Introducing a system of cohort sampling involving a limited number of pupils in a limited number of schools to collect data for monitoring national standards.

Assessment in the Future: Building the Case for Change

Introduction

1. The General Teaching Council for England (GTC) warmly welcomes the Education and Skills Select Committee's (ESSC) inquiry into testing and assessment. We hope that the report the Committee will publish as a result of this inquiry will persuade the Government to undertake an urgent and fundamental review of the testing and assessment regime in maintained schools. We also hope that this leads to the implementation of changes to rebalance the role of assessment in education and refocus the importance of teachers' professional judgement in how pupils are assessed in the future.

2. The proposals in this memorandum are based on the view that a single approach to pupil assessment is currently being used for too many purposes and that assessment in all its forms should be fit for purpose, place the least possible burden on pupils, teachers and schools and have the least possible adverse effect on the curriculum. The GTC's proposals have been widely discussed with teachers, head teachers, parents, governors, national agencies and representatives of local authorities through a series of over 20 events across England and a major national conference. Further GTC events on assessment will take place over the summer.

3. The GTC continues to be convinced that the existing assessment regime needs to be changed. Evidence from teachers at GTC consultative events in 2006 and 2007 shows that the current system finds schools giving too much emphasis to end of Key Stage (KS) test results and performance tables at the expense of the longer-term needs of the children and young people they endeavour to serve. A summary of the views expressed by teachers at four principal GTC events is attached at Appendix 1. It is a concern shared by many that the accountability regime inhibits the capacity of schools to deliver sustainable personalised learning and limits local influence on schooling.

4. The GTC is the independent professional body for teaching. Its main duties are to regulate the teaching profession and to advise the Secretary of State on a range of issues that concern teaching and learning. The Council acts in the public interest to help to raise standards in education.

The Extent and Impact of Testing and Assessment

5. England's pupils are some of the most frequently tested in the world. The average pupil in England will take at least 70 tests and examinations before leaving school. The system employs 54,000 examiners and moderators dealing with 25 million test scripts a year (Skidmore, P 2003). However, despite the very significant resources required to conduct them, tests do not in themselves raise standards, as the DfES Primary Strategy *Excellence and Enjoyment* 2003 acknowledged.

6. Pupils understand that the KS tests represent high stakes for themselves, their teachers and their school. Evidence from teachers indicates that high stakes testing has a narrowing effect upon the curriculum, by moving the focus of curriculum delivery away from being broad and balanced to a narrower one based on test content. Research studies indicate that high stakes testing can depress pupils' motivation and increase their anxiety (Harlen, Wynne and Crick and Ruth Deakin 10:2, p169–207).

7. The tests are not integrated into pupils' normal classroom work; they are set at arm's length from teachers' professional judgements. Summative outcomes reached by the teacher carry less public weight than outcomes from the end-of KS tests with the exception of the arrangements at the end of KS1. Throughout the system, assessment for accountability is given precedence over on-going formative assessment that supports learning.

8. Assessment for Learning (AfL) is the type of formative assessment that supports learning. One of the substantial benefits of AfL is that it encourages learners to take a role in their progress and development and to develop the capacity for self-assessment and peer assessment. AfL also supports teacher planning and teaching and the away in which curriculum and resources are organised to optimise learning. The system of national tests ignores these processes.

9. Parents' legitimate wishes to know about their children's learning and progress are not best served by the single measure that is the outcome of end-of KS testing. Nor do the outcomes best demonstrate schools' accountability to their parents and local community. Evidence indicates that when parents make judgements about the quality of a school they do not use the school's position on published league tables as their main criterion (GfK NOP Social Research 2005). Parents require broadened and enriched sources of information about their local schools.

The case for change

10. The assessment system relies upon the use of any given test for too many purposes and, as the GTC argues, this compromises the reliability and validity of the information obtained. The system creates tensions that have had a negative impact upon the nature and quality of the education that some of our children and young people receive. These tensions may impede the full realisation of new approaches to education, including more personalised learning.

11. The received wisdom that KS tests and external public examinations are error free methods of assessing pupil attainment is misleading. All methods of assessment are prone to error. As Professor Paul Black argues, it is unproven that assessing pupil attainment by the use of tests is less error prone than relying on teachers' assessment. Other evidence suggests that on a particular day, at KS3, 30% of pupils and at KS4, 40% of pupils have been given the wrong level. (Wiliam D. 2007)

12. The outcomes from the end-of KS testing are also used as a measure of standards over time. The technical limitations of the end of KS tests and the tensions within the national assessment system may mean that the use of these data in this way is flawed. Questions have been raised about a significant margin of error that could be involved in the testing process and therefore its reliability as the basis of long term policy formation.

13. Furthermore, there have also been issues raised by teachers and others about the extent to which the tests assess the actual attainment of pupils as opposed to their performance on a particular day. Other well documented concerns include the narrowness of the tests, the "drilling" of pupils in preparation and the backwash effect on curriculum breadth and flexibility.

14. *2020 Vision* recognised that national assessment tests are not primarily diagnostic tools to ascertain pupils' learning needs. Nor do they recognise or adequately record the extent to which pupils have developed the desired skills and aptitudes. The review recommended that the Government commission a group to report on the national curriculum and its assessment "as a matter of priority". The Council strongly supports the group's recommendation for this review.

15. The Government should shift the balance of schools' accountability away from the centre and towards the community, parents and pupils, enabling improved dialogue with parents and less undue focus on national performance measures.

16. The DfES *Making Good Progress* document proposes twice yearly externally-marked "progress" tests and targets in addition to the current end-of KS attainment tests and threshold targets. This assessment approach would increase the number of tests a pupil must take and the pressures created by performance tables would remain.

17. A preferable system would be the GTC's proposals for a national bank of tests/tasks which teachers could use when pupils are ready, rather than tests that meet the needs of the system. The bank of tests/tasks would support teachers' summative assessment and could be used in conjunction with AfL. This would, in the longer term, promote a far closer relationship between formative and summative assessment than exists currently.

18. The current assessment system does not sit well with the local cross-institutional collaborative approach required to give all 14–19 year olds the right to study the new diplomas. Their introduction provides the opportunity to begin the process of moving away from an assessment system dominated by the purposes of quality control and accountability and assessment of learning towards a more balanced model with a greater element of diagnostic and formative assessment for learning.

19. The GTC believes that there is a tension between the Government's commitment to personalised learning in 14–19 education, and more localised and responsive structures to support it, and the current emphasis on national external examinations and national performance tables as currently configured. The 14–19 phase should be established as a continuum for learners to move away from the break at 16, which performance tables encourage.

20. The GTC supports more localised 14–19 performance information that reflects area-based collaborative provision and area-based inspection involving institutional self-evaluation where appropriate.

Summary of the GTC's assessment proposals

21. The GTC supports a comprehensive review of the purposes of assessment, the type of information it generates and who uses it and how. Assessment, testing and coursework are means to an end, not ends in themselves. A review needs to start with identifying the key purposes that we need the future assessment system to support and the most effective ways of achieving them.

22. In the GTC's view, the design of a future assessment system should therefore be underpinned by three core purposes that focus on providing information about the learner, the school and the system. These are:

— supporting teaching and learning;
— providing public information and accountability; and
— monitoring national standards.

Principles

23. The GTC's proposals on assessment around the three purposes are based on the following key principles:

— a commitment to using teacher professional judgement in the assessment system to better effect than the current arrangements permit;
— enabling teachers to carry out assessment processes more effectively, so that the quality of pupil learning is further enhanced and standards of achievement are improved;
— separating the purposes of assessment so that the use of assessment for accountability no longer takes precedence over assessment for developing learning; and
— creating an assessment model for the future that involves robust and transparent processes in order to withstand public scrutiny.

GTC Proposals

Assessment to support learning

> *"Assessment should be about finding out what the children know so we can move them forward. It's not about a single test result, so stop national testing and trust in teacher assessment".*
>
> Teacher, GTC event, Manchester

Assessment for Learning (AfL)

24. The Government should continue to invest in AfL through the National Strategies working with the Assessment Reform Group (ARG) and ensure that AfL approaches are better embedded in the culture of schools. Local authorities and other local networks should support schools' and teachers' capacity to conduct assessment effectively, and aim to revitalise pupil and teacher learning in the process.

Bank of tests/tasks

> *"If I could change on thing about assessment it would be to abolish KS2 tests. This would alleviate pressure on children and staff. It would also allow upper KS2 children to enjoy learning and explore their natural curiosity".*
>
> Teacher, GTC event, Bristol

25. The Government should introduce a range of nationally devised tests/tasks which individual teachers could use with their pupils in the classroom when the teacher judges that the pupil(s) is/are ready. The tests/tasks would initially be used for summative purposes. Over time the range of tests/tasks would be expanded

so that teachers would effectively choose from a bank of resources that would be used to confirm or challenge their own summative judgements. The tests would replace the current universal end of KS tests. The information generated during and at the end of the key stage would be used by the school, local authority, parents and pupils themselves to move learning forward.

Teacher assessment

26. A teacher assessment model should be implemented incrementally. In the immediate term, AfL would be used for formative purposes and the bank of tests/tasks would be used summatively. Longer term, teachers would be working towards a position where all forms of pupil assessment, whatever their purpose, involve an increasing degree of teacher professional judgement.

Teacher learning

27. The GTC's proposals should be supported by increased Government investment in teachers' assessment skills. These include better support for all teachers during initial training and continued professional learning, including professional/peer moderation activities, and more specialist assessment career paths for teachers to lead assessment processes across schools and localities.

Assessment for school improvement and accountability

> *"Teachers feel they are sufficiently accountable in terms of quantity of information; it is the quality and nature of information that needs to be addressed".*
>
> Teacher, GTC event, Manchester

28. The increased investment in AfL, the use of an increasing range of assessment tests/tasks by teachers and the development of moderation processes in schools would provide the means for teachers to develop a relationship with parents based on a richer and better informed dialogue.

29. As part of the school's accountability to its stakeholders, parents and pupils should be entitled to be fully and regularly informed about progress and attainment, with information being wider than a report of levels and grades. Information must be provided in a timely way so that it can be used as the basis for any improvement strategy. Entitlement to better information would be a better basis for engagement in school evaluation and improvement processes.

30. As part of the New Relationship with Schools (NRwS), the GTC believes that the Government should endow schools with greater responsibility for communicating their accountability information to parents via the school profile on individual and collective pupil progress. This would include assessment information and draw on school self-evaluation and inspection findings. The GTC is committed to this school based model of accountability and believes that it has more valuable information to offer parents than the de-contextualised and incomplete comparisons between schools published in performance tables.

Assessment for monitoring national standards

> *"We are enthusiastic that this system may enable a broader, more accurate assessment base across the whole curriculum".*
>
> Teacher, GTC event, London

31. The Government should introduce a system of cohort sampling as the most cost effective and efficient way to monitor national standards. A limited number of pupils should be tested in a limited number of schools. Different pupils could be given different tests in order to cover a broad range across the curriculum. No pupil would take more than one test. Tests would contain common questions that allowed all pupils in the sample to be placed on a common scale. In the longer term, such a system for national monitoring should replace the use of the present universal testing model. In the shorter term a cohort sampling system should be trialled.

GTC Proposals: Detailed Discussion

Assessment to support learning

> *"The teacher and the school are best placed to known their children through using their professional judgement. Children are not commodities or 'raw material'".*
>
> Teacher, GTC event, London

Assessment for Learning (AfL)

> *The Government should continue to invest in Assessment for Learning (AfL) through the National Strategies working with the Assessment Reform Group (ARG) and ensure that AfL approaches are better embedded in the culture of schools. Local authorities and other local networks should support schools' and teachers' capacity to conduct assessment effectively, and aim to revitalise pupil and teacher learning in the process.* (para 24)

32. The Government is already investing in AfL through the national strategies. Helping to develop whole school approaches. AfL is critical because it offers the potential for radically changing the way that teachers and pupils interact. More than simply one of a number of strands of school strategy, it lies at the heart of any personalised learning vision for the classroom.

33. The Government should consider how best to continue developing AfL best practice and maintain its momentum. The GTC endorses the recommendation in the Gilbert Review report *Teaching and Learning 2020 Vision* that AfL should be a priority for teaching and learning and that resources should be put in place to ensure that it is better embedded in schools.

34. Local authorities and the many collaborative partnerships and federations of schools and other institutions working together should continue to have a key role in supporting schools to develop AfL and assessment communities across schools and their localities. This is particularly important because GTC dialogue with teachers indicates that even in those areas that have been involved in AfL action research there remains considerable diversity in school approaches to AfL.

Bank of tests/tasks

> *The Government should introduce a range of nationally devised tests/tasks which individual teachers could use with their pupils in the classroom when the teacher judges that the pupil(s) is/are ready. The tests/tasks would be used for summative purposes. Over time the range of tests/tasks would be expanded so that teachers would effectively choose from a bank of resources that would be used to confirm their own summative judgements. The tests would replace the current universal end of Key Stage tests. The information generated during and at the end of the key stage would be used by the school, local authority, parents and pupils themselves to move learning forward.* (para 25)

35. Over time and as teachers become familiar with and skilled in using a range of nationally devised tests/tasks, they should increasingly be used to confirm teachers' existing summative assessments of pupil progress and achievement.

36. In the early stages, schools should have the option of tests/tasks being externally marked. Teachers would become increasingly involved in the analysis of outcomes so that the tests/tasks would play a key role in teaching and learning. Teachers' involvement in marking and analysis should also be linked to professional development opportunities.

37. Teachers would be encouraged to use the evidence on which they based their judgement about the timing of the test as part of professional/peer moderation activities in preparation for working towards a more integrated model of teacher assessment. Teachers would collect evidence from test/task outcomes during and at the end of each KS so that the information can be used formatively to adapt teaching as well as the basis for summative decision-making. The evidence derived from the test/tasks would be subject to assessment moderation processes.

38. The range of materials would increase in the longer term into a bank of varied assessment materials on which the teacher could draw. The teaching profession, particularly those with assessment expertise, should be involved in the development of test/task materials, including on-line materials. The information generated would be used by the school, local authority, parents and pupils themselves to move learning forward.

Teacher assessment

> *A teacher assessment model should be implemented incrementally. In the immediate term, AfL would be used for formative purposes and the bank of tests/tasks would be used summatively. Longer term, teachers would be working towards a situation where all forms of pupil assessment, whatever their purpose, involve an increasing degree of teacher professional judgement.* (para 26)

39. The GTC supports the findings of the systematic review (Harlen, 2004) that assessment by teachers has the potential to provide summative information about students' achievements because teachers can build up a picture of individual students' attainment across a range of activities and goals. However, teacher evidence to the GTC over time suggests that many teachers do not feel they are currently in a position, or work within a structure, that would allow them to undertake all forms of pupil assessment. It is for this reason that the GTC recommends an incremental approach to implementation using AfL and a bank of tests/tasks.

40. The GTC remains convinced that in the longer term all forms of pupil assessment should involve an increasing degree of teacher professional judgement. This would create a richer educational experience for all pupils, with assessment integrated better into other elements of teaching and learning, especially for those pupils at risk of under achievement whose interests are not being served by the current use of statutory tests and external public examinations.

Teacher learning

> *The GTC's proposals should be supported by increased Government investment in teachers' assessment skills. These include better support for all teachers during initial training and continued professional learning, including professional/peer moderation activites, and more specialist assessment career paths for teachers to lead assessment processes across schools and localities.* (para 27)

41. There needs to be further investment in teachers' assessment skills along with and inside the AfL framework. Assessment needs to be a stronger element of the professional standards framework, including qualified teacher status (QTS) and induction standards. Managing assessment across a subject area, a department or faculty and as part of a whole school approach should also be a critical component of professional standards for leadership.

42. It is also vital that specialist assessment roles are created in every school. The Chartered Examiner route developed by the National Assessment Agency (NAA) to revitalise the teaching profession's involvement in public examiner roles must be extended to roles in National Curriculum assessment at all KS and in leading AfL in individual schools. The Government needs to invest in training and support for teachers to undertake these roles.

43. The priority for embedding AfL in schools as recommended by the Gilbert Review must be underpinned by making it a priority focus of teacher learning, as the Review Group also indicated.

Assessment for school improvement and accountability

> *The increased investment in AfL, the use of an increasing range of assessment tests/tasks by teacher and the development of moderation processes in schools would provide the means for teachers to develop a relationship with parents based on a richer and better informed dialogue than currently.* (para 28)
>
> *As part of the school's accountability to its key stakeholders, pupils and parents should be entitled to be fully and regularly informed about progress and attainment, with information being wider than level and grades and provided in a timely way so that the information can be used as the basis for any improvement strategy. Entitlement to better information would be a better basis for engagement in school evaluation and improvement processes.* (para 29)

44. Research (Black *et al*, 2003) into AfL comment only feedback to pupils found "*the provision of comments to students helps parents to focus on and support the student's learning rather than focus on uninformed efforts to interpret a mark or grade and/or simply urge their child to work harder*". The GTC believes that AfL approaches have the potential for providing pupils and parents with a source of information on progress that involves them as partners.

45. The GTC's proposals to replace KS testing with a bank of tests/tasks also add new opportunities better to involve individual pupils and their parents in a continuing and well-informed dialogue with teachers about learning and progress. These proposals would support better and more timely information that focuses as much on ongoing progress as the review of summative outcomes.

46. The importance of more timely information for parents based on progress was theme emerging from some recent focus groups of parents commissioned from BRMB Social Research by the GTC. The parents in the study were concerned that information given to them should represent a "*call to action*" if necessary rather than a retrospective summary based on assessment levels on which parents were unable to act. (BRMB, 2007) A report of the findings of this study is at Appendix 2. (*not printed*).

47. A MORI poll of parents commissioned by the GTC (2005) showed how much value parents placed on their communication with schools and their children's teachers. 97% of the sample appreciated verbal feedback with a further 71% finding written feedback in the form of a regular report very useful. These views were confirmed by the findings of a parent focus group carried out by NOP (2005) for the GTC. Here "*there was a strong desire for more written information to complement the academic results received*" and more frequent verbal information as it "*was considered to be more tailored to the individual pupil and offered the opportunity for discussion with parents*".

48. The BRMB study revealed not only that parents wanted increased information on progress and a greater range of information mechanisms. The findings reflected "*that parents generally did not understand or were confused about how their child was assessed at school, particularly during the primary school years*". Besides the confusion about what the assessment levels really meant, parents "*had little recall of when*

teacher-led assessment would take place, the range of methods that were likely to be used, nor the role of the assessment methods being used". The GTC believes that an enhanced dialogue between schools and parents must start with more information and explanation about the components of the assessment system.

49. The GTC broadly supports the NRwS framework developments that include:

- the greater weight given to school self-evaluation and schools managing their own data;
- the new School Improvement Partner (SIP) role working to support self- evaluation and improvement processes in all schools;
- a more differentiated model of shorter, sharper Ofsted inspections resulting in shorter and more accessible reports; and
- the introduction of the School Profile.

50. Research (Rudduck 2004) suggests that schools effectively involving pupils in shaping the way that teaching and learning is organised could have benefits for school improvement. Evidence collected by the GTC also reflects increasing efforts by schools to integrate parental consultation into school self-evaluation.

51. As part of the NRwS, the GTC believes that the Government should endow schools with greater responsibility for communicating their accountability information to parents via the school profile on individual and collective pupil progress, including assessment information and drawing on school self-evaluation and inspection findings. The GTC is committed to this school based model of accountability and believes that it has more valuable information to offer parents than the de-contextualised and incomplete comparisons between schools as published in performance tables.

52. Professional learning for teachers is the key to preparing schools to take on more responsibility for collecting, using and interpreting performance data as part of their accountability to their stakeholders. The Government should enable schools to focus on professional learning that is based on combining quantitative and qualitative pupil level data and using it, in partnership with pupils and parents, to plan the personalised learning of children and young people. *2020 Vision* indicates that the analysis and use of data—with a specific focus on AfL—is an important skill for the school workforce. Teacher learning therefore must meet the challenge of ensuring data is used properly and coherently in schools.

Assessment for monitoring national standards

> *The Government should introduce a system of cohort sampling as the most cost effective and efficient way to monitor national standards. A limited number of pupils should be tested in a limited number of schools. Different pupils could be given different tests in order to cover a broad range across the curriculum. No pupil would take more than one test. Test would contain common questions that allowed all pupils in the sample to be placed on a common scale. In the longer term, such a system for national monitoring should replace the use of the present universal testing model. In the shorter term a cohort sampling system should be trialled.* (para 31)
>
> *"This potentially sounds better, cheaper, less stressful".*
>
> Teacher, GTC event, London

53. The use of end of KS test outcomes to monitor standards over time may be flawed because there are a number of technical issues about the tests.

54. There are problems of scaling with, for example, a pupil assessed at the bottom of Level 4 being nearer in terms of marks to the top of Level 3 than the top of Level 4. There is weak criterion referencing involved in the system of testing. There is also a problem with the public demand that tests maintain consistent standards over time. In order to achieve this it would require everything related to the tests to remain exactly the same. In fact the tests are curriculum linked and the context on which they are based has been subject to constant change and even if that had not been the case, students have become better at taking the tests themselves (Oates 2004, 2005).

55. The GTC, therefore, proposes a system of cohort sampling involving a limited number of pupils in a limited number of schools and utilising a matrix test structure. This would mean that numerous tests could be used across the sample, thus widening the breadth of the curriculum that is being tested. Common questions would appear in any two or more tests by which pupils in the sample who take different tests could be put onto a common scale. No pupil would be required to take more than one test. The tests would be administered by teachers though external support could be called upon in relation to conducting practical tests. A detailed explanation of how monitoring by cohort sampling works is at Appendix 3.

56. This system would be relatively inexpensive as test items can be used repeatedly over time and questions can be replaced without the need to develop whole new tests.

57. The current testing burden placed on schools and students would be greatly reduced because the cohort would be made up of a light sampling of schools and a light sampling of students within those schools. The distortions of the curriculum and pressures on pupils, parents and teachers of high stakes testing would be removed.

58. The GTC proposes an initial pilot of the cohort sampling system by QCA, perhaps at a particular KS. In the long term we propose that the current universal testing model for national monitoring be replaced by cohort sampling.

Conclusion

> *"These proposals would lead to a more creative curriculum—wonderful idea—this would make teaching more enjoyable. How can we persuade the Government?"*
>
> Teachers, GTC event, Bristol

59. The GTC anticipates that the ESSC will receive very few submissions to this inquiry from the education community arguing that the assessment system should remain and continue in its current form. We are convinced that arguments will centre not on whether the assessment system should be changed, but how. We hope that the Select Committee will urge the Government to undertake a measured and wide-ranging consultation involving parents, pupils, teachers and all those with an interest in education and assessment. We owe it to our pupils to replace the system we have with one that genuinely serves the interests of pupils, parents, schools and the public. Over-hasty change runs the risk of replacing it with a new but equally dysfunctional system.

60. The GTC believes that its proposals for change would be supported by the teaching profession, parents and others because they offer a route towards countering the measurement culture that has gained currency since the 1988 Education Reform Act. They also reaffirm the pre-eminence of using assessment to support teaching and learning, lift the burdens from pupils, teachers and schools that distort the curriculum while providing information to those to whom schools and teacher are accountable that is meaningful, timely and reliable.

June 2007

References

DfES, January 2007. *Making Good Progress* consultation document.

Skidmore, P, 2003 *Beyond Measure*. DEMOS.

GfK NOP Social Research, 2005. *Research among parents: a qualitative study*. Conducted on behalf of the GTC. Unpublished.

Harlen, Wynne and Deakin Crick, R, *Testing and Motivation for Learning* Assessment in Education Principles, Policy and Practice, 10:2, 169–207.

Harlen, W, and Deakin Crick, R, *A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning*. (EPPI-Centre Review); In Research Evidence in Education Library Issue 1 London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Black, P, Presentation to Institute of Educational Assessors National Conference. London 4 May 2007.

Black, P, Harrison, C, Lee, C, Marshall, B & Wiliam, D 2003. *Assessment for Learning: putting it into practice*. Buckingham, UK: Open University Press.

Oates, 2004 and 2005. Presentations to GTC seminars—18 May 2004 and 11 October 2005.

Rudduck, J, January 2004. Presentation to GTC Accountability Stakeholder Forum.

Teaching and Learning in 2020 Review Group, December 2006. *2020 Vision*.

Wiliam, D, March 2007. Presentation to GTC Assessment Conference. London.

APPENDIX 1

GTC CONSULTATION ON ASSESSMENT AND TESTING

Introduction

Over the academic year 2006–07 the GTC held a series of consultative seminars in order to present its proposals for changes to the assessment system to teachers, head teachers, parents and others and to receive feedback. Events were held in Leeds, Bristol, Manchester and London as well as a national conference in March 2007. This appendix is a summary of the views expressed.

Assessment to Support Learning

GTC's bank of tests/tasks

Teachers welcomed the bank of tests because they felt they would free up time for teachers to enjoy teaching and pupils to enjoy learning. They wanted to know in detail how and when they would be administered, and by whom. The new tests could eliminate "teaching to the test". They could be a spur to improving the quality of teacher assessment, particularly if teachers had input into their design. The tests at primary phase would need to be matched to specific learning objectives and primary strategy materials. Teachers felt that the tests should be able to define what pupils can and cannot do rather than assigning a level. The DfES and other bodies must take time to trial new initiatives and systems before they expect teachers and schools to put them into practice. Teachers were clear that any new system that resulted in more assessment and less teaching was suspect and they sought reassurance that this was not the case with the bank of tests.

GTC Bank of Tests/Tasks—Teachers' Questions

- Will teachers have enough time to deal with the extra workload?
- Who will be responsible for writing the tests?
- How will special needs be dealt with? If a special education needs (SEN) pupil is extremely slow to arrive at a position where they are ready to be tested they may emerge from Key Stage (KS) 4 with very few benchmarks, if any.
- Will the tests measure knowledge or skills?
- Will there be tests for the gifted and talented?
- Would the tests measure in levels or in standard scores?
- How different will the tests/tasks be to the current national tests? "We don't want SATs look-alikes".
- Is yearly testing effective? If more testing was more frequent it would not be realistic to test every pupil in every subject. Would a bank of tests not just allow some teachers to test even more?
- Can this system work in secondary education as effectively as it might in primary?

Assessment for Learning

Teachers had concerns over how well Assessment for Learning (AfL) is embedded in schools and how rigorously it is being used. Knowledge of AfL can be very mixed. It was felt that teachers' professional judgement was key to successfully implementing AfL because teachers would need a thorough and detailed knowledge of their pupils. A culture shift was necessary to dispel the sense of distrust of assessments at change of phase and explain how AfL fits into test culture. If schools adopted AfL they would need to demonstrate to pupils and parents that standards are being assured and that the process is transparent.

Teachers also felt there was insufficient knowledge and recognition of pupils' learning speeds and styles. Many questioned the wisdom of a knowledge-based, rather than learning-based, curriculum.

Teachers and teaching assistants need support and opportunities to raise their skills in using AfL and the GTC's bank of tests. A useful tool for teachers and others would be the networks of teachers within local authorities to further assessment and help develop teacher confidence. This would ensure that what is done in each school is comparable and consistent.

Formative assessment is the bedrock because it is the tool for taking pupil learning forward. The process should be about achievable tasks which show where pupils' difficulties are for their future benefit. Over-assessment will alienate certain pupils. We should not forget the case for creativity and children's enjoyment of learning. We should be aiming to make each pupil aware and responsible for their learning, for instance through self-assessment.

> *"Assessment should be about finding out what the children know so we can move them forward. It's not about a single test result, so stop national testing ands trust in teachers' assessment".*

Personalised learning

Teachers were curious about how personalised learning, with its emphasis on progression and assessment at the pupil's own pace could be translated to the secondary model. Assessment for pupils with SEN or disabilities should be a particular focus. Teachers expressed a wish for less emphasis on national testing and more faith in personalised learning and formative assessment.

The current testing regime

Teachers questioned the value of Primary KS tests which, they said, distort teaching and are not effective in helping pupils to learn. Teachers also thought the marking scheme for national KS tests is too prescriptive and narrowly based on key words, with no scope for pupil creativity. They considered the system impersonal and that it inhibits curriculum enrichment. The outcomes of the tests, particularly at KS3 carry so much weight that teachers are increasingly teaching to the test, making them feel like trainers, not teachers.

The end of KS2 and KS3 tests do not, the teachers said, provide realistic information about children. These tests are demoralising for children who do not have strengths in the core areas but may have skills and talents in non-tested curricular areas. For special schools, tests are a mere paper exercise and do not show multi-sensory, multiple intelligence learning.

Teachers supported the new style KS1 SAT/assessment system because of the greater emphasis on teacher assessment and would support its extension to other year groups.

> *"If I could change one thing about assessment it would be abolish KS2 tests. This would alleviate pressure on children and staff. It would also allow upper KS2 children to enjoy learning and explore their natural curiosity".*

Teacher professional judgement and resources for continuing professional development

It was felt that teacher confidence had been eroded over time and that the GTC's proposals would help to restore it. Trust in professionalism should extend to teaching assistants and other school staff as well as teachers. Teachers also pointed out that many teachers had only ever known national tests for teacher assessment, so training in new ways of working was crucial.

Teachers identified the need for moderation training, and the time to do it, as necessary for an effective rebalancing of assessment; along with more input on AfL in initial teacher training as newly qualified teachers' (NQTs) understanding can be limited.

Assessment for Accountability

School level accountability

The majority of teachers who attended the GTC events thought that league tables were flawed and that they distorted teaching. The levels within the key stages did not work. Without performance tables, teachers thought it would be possible to network more effectively between schools, eliminate divisive competition between schools and develop better collaboration between primary and secondary schools.

While teachers recognised that performance tables can have a role in focusing on school improvement and in parents' choice of school, it was also felt that a wider variety of information should be available. Ofsted inspections, local knowledge, cultural considerations and proximity are also factors influencing parents. Some teachers favoured the use of portfolios of evidence of pupils' achievements to confirm teacher judgements. The development of IT-based "Learning Platforms" would give parents the ability to access pupil reports and assessment records. The School Profile was seen as an effective accountability mechanism that should be more widely promoted.

Teachers felt that the system is currently very accountable but meaningful information is not communicated well to parents. A more personalised approach was needed. It was noted that the independent sector is skilled in communicating with parents and engaging them in dialogue and could offer lessons to be learnt by the maintained sectors.

> *"Parents should be entitled to access any information about the school when they need it".*
>
> *"Schools need to be able, without risk, to be honest about when things are not going well. They also need to have access to data that helps them work on a level playing field and fee part of a local community".*

Teacher Assessment and Feedback

Teachers felt that they are sufficiently accountable in terms of the quality of information; it is the quality and nature of the information that needs to be addressed. Teachers are adept at assessment because they know their pupils well, but are judged on external test results. It was suggested that schools could be paired nationally to moderate each other and exploit the on-line facilities available.

> *"The teacher and the school are best placed to know their children through using their professional judgement. Children are not commodities or 'raw material'".*

Ofsted and Accountability

Teachers thought a culture is growing in which all children are expected to be above average. This pressure is unrealistic and not helpful to pupils and teaching.

It was felt that an appropriate accountability model would be strengthened and supported school self-evaluation, validated by Ofsted. There was a general desire that Ofsted be more supportive and consistent but frustration that performance tables—offered little of practical diagnostic use to schools.

> *"Data to compare schools is valued by heads and local authorities in managing their own specific systems. This data should be preserved but not necessarily in the form of league tables".*

Assessment for Monitoring

Teachers were curious about how cohort sampling would work—how pupils would be selected, how frequently they would be sampled and on what criteria. It was observed that universities self-moderate and regulate the awarding of degrees and it could be that they offer elements which could be incorporated into the cohort model used.

Some teachers were concerned about the time and cost involved in cohort sampling. They wanted to know how school effectiveness would be judged in the absence of school standards data. They were also concerned about sampling pupils with special needs. Others welcomed the proposal on the basis that it could provide comparative data to moderate each pupil against a range of curricular areas, skills and understanding of aspects of development not currently able to be assessed such as pupil attitudes, aspects of *Every Child Matters* and citizenship.

> *"We are enthusiastic that this system may enable a broader, more accurate assessment base across the whole curriculum".*
>
> *"This potentially sounds better, cheaper, less stressful".*

General Views on the GTC's Proposals

> *"These proposals would lead to a more creative curriculum".*
>
> *"Wonderful idea—this will make teaching more enjoyable".*
>
> *"How can we persuade the Government?"*

APPENDIX 2

ENGAGING WITH PARENTS: PUPIL ASSESSMENT[2]

APPENDIX 3

NATIONAL MONITORING BY COHORT SAMPLING: HOW IT WORKS

An approach to national monitoring that uses cohort sampling has numerous advantages compared to the testing of whole cohorts of students. The techniques of cohort sampling are well established and are used in studies of international comparisons of student performance such as in the PISA and TIMSS projects. Cohort sampling has also been used in this country from the mid seventies through the eighties by the Assessment of Performance Unit (APU) within DfES. An explanation of the approach used by the APU will serve to illustrate the workings of national monitoring by cohort sampling.

The APU was set up within DfES in 1975. Its brief was to promote the development of assessing and monitoring the achievement of children at school and to identify the incidence of underachievement.

The actual monitoring was contracted out. The National Foundation for Educational Research (NFER) was contracted to carry out the monitoring of mathematics, language and foreign languages, a consortium from Leeds University and King's College London was contracted to monitor science, whilst Goldsmiths College was contracted to monitor technology. Surveys of samples of students aged 11 years old were started in 1978 and continued until 1988. Surveys of students aged 13 were started in 1980 and continued until 1985 and surveys of students aged 15 were started in 1978 and continued until 1988. Table 1 gives the subject details and the specific dates of the APU surveys.

[2] Research report prepared for the General Teaching Council for England by BMRB Social Research. Separate document, not printed. See http://www.gtce.org.uk/shared/contentlibs/126795/93128/120213/engaging_parents_report.pdf

Table 1

APU SURVEYS BY SUBJECT, DATE AND AGE OF STUDENTS

Subject	*Age 11*	*Age 13*	*Age 15*
Mathematics	1978–82, 1987		1978–82, 1987
Language	1979–83, 1988		1979–83, 1988
Science	1980–84	1980–84	1980–84
Foreign Languages		1983–85	
Design & Technology			1988

The approach of the APU was to have a light sampling of schools and a light sampling of pupils within schools. Thus, in the case of the mathematics surveys in England a sample of 10,000 students (about 1.5% of population) was used. Each student was given a written test (students did not all have the same written test) and sub-samples of 2–3,000 were also given other assessments such as attitude questionnaires or practical mathematics tests. A linking and scaling structure was built into the written tests so that students could all be placed on a common scale. The structure is a cartwheel design in which common items appeared in any two tests. Table 2 illustrates this structure.

Table 2

LINKING STRUCTURE OF WRITTEN TESTS

Group of Test items	*Test 1*	*Test 2*	*Test 3*	*Test 4*	*Test 5*	*Test 6*
A	A					A
B	B	B				
C		C	C			
D			D	D		
E				E	E	
F					F	F

With reference to Table 2, although each student takes just one of the tests, the common items that appear across any two tests means that the performance of students across the whole six tests can be put onto a common scale.

It is by this design that a wider coverage of the curriculum can be assessed than is possible from any single test and this can be achieved without putting undue burden on individual schools and students. Furthermore, this approach enables students' performance to be monitored in those areas of the curriculum that it is impracticable to test a whole cohort such as practical mathematics. This can be achieved by setting assessment in these areas for small sub-samples of students.

The Advantages

The approach of cohort sampling combined with a linking and scaling structure for the tests offers numerous advantages for national monitoring.

1. As the approach is a light sampling of schools and a light sampling of students within schools this reduces the testing burden on schools and students compared to the present regime.
2. Within this approach, schools and students have anonymity; the testing is low stakes and thus should have minimal adverse impact upon the curriculum.
3. It is possible to have a wide curriculum coverage that is tested.
4. It is possible to have a range of assessment formats, for example some assessment of practical aspect of the curriculum can be addressed.
5. Test items can be used repeatedly over time.
6. Items can be replaced without the need to develop whole new tests.
7. It is relatively inexpensive.
8. The outcomes give a good indication of trends in performance.
9. It is a tried and tested method that has been used in this country and is still being used in surveys of performance for international comparisons.

THE DISADVANTAGES

There are some limitations to this approach.

1. It does not give ratings for individual schools.
2. With light sampling of pupils, it is difficult to give feedback to individual schools.
3. The linking and scaling is based on Item Response Theory (IRT), the statistics of which can be difficult to interpret. A simple scale would need to be developed that is adhered to and understood by all. An example of how this might be achieved can be seen in the international assessment projects such as TIMSS, PISA and PIRLS.

June 2007

Witnesses: **Dr Mary Bousted,** General Secretary, Association of Teachers and Lecturers (ATL), **Mick Brookes,** General Secretary, National Association of Head Teachers (NAHT), **Brian Lightman,** President, Association of School and College Leaders (ASCL), and **Keith Bartley,** Chief Executive, General Teaching Council for England (GTC), gave evidence.

Q128 Chairman (Fiona Mactaggart): Good afternoon, everyone. We have a rather interesting situation here. We are missing our Chair, because this session coincides with the debate on the Education and Skills Bill in the House, so I have agreed to act as Chair. We are also missing a bunch of witnesses. I am afraid that Steve Sinnott of the National Union of Teachers and Chris Keates of the National Association of Schoolmasters Union of Women Teachers pulled out. They would have provided two thirds of the witnesses in the second part of this evidence session, so we have decided to put together all four witnesses who are here—thank you for agreeing to this, Mary—for one slightly truncated session. It is important that we speak to the NASUWT and the NUT, and it is striking that in its evidence the NUT specifically asked to come before the Committee to give evidence. We want to speak to senior officials, and not to junior substitutes, so we will arrange an alternative date for them to appear. In the meantime, in this session—I imagine that it will finish at about 5.15 pm if that is convenient for all our witnesses—we will look at testing and assessment as part of our inquiry. It is usual for the Chair to offer witnesses an opportunity to make brief preliminary remarks about the issues before them, which can help the Committee to zero in on its main concerns. If any of you would like to do that, I would welcome your contribution.

Dr Bousted: The key issue for this Committee is that proposed by Dylan Wiliam, who said that the challenge that we have as a country is to have tests that are worth teaching to. At present, the view of the Association of Teachers and Lecturers is that we do not have tests that are worth teaching to. The current testing system is highly unsatisfactory. Some 30% of pupils will be awarded the wrong level at Key Stage tests. That is an issue for standard assessment tests and GCSEs. Another issue is that, because of over-teaching to the tests, six months on from being tested at Key Stages 2 and 3, 25% of children do not maintain the same level. For a Government who are keenly interested in raising pupils' standards and system levels of attainment and achievement, that is not good. What is striking from the evidence that has been presented by the people representing our organisations is the degree of consensus in the submissions. There is consensus that tests are used for too many different purposes, and because of that their value is corrupted. There is consensus on the inadequate relationship between the national curriculum and the tests. In other words, the tests cover very narrow aspects of the national curriculum, which leads to worries about validity. There is also striking evidence that because we test seven out of 11 years of compulsory schooling, there is a demotivating impact on pupils, which leads to a very instrumental view of learning. I was interested to read in *The Times Higher Education Supplement* that this instrumental view of learning is even affecting the most academic pupils—those who go on to higher education. They arrive at university without the necessary research skills and skills for independent learning, which then have to be taught in the first year of university. Therefore, the tests have a severe effect on all children in the curriculum. Even in the Government's own terms, the tests do not do the job and, more significantly, they militate against assessment for learning, which we need to encourage. This is a highly significant inquiry. I am glad that the Select Committee wants to consider the matter. I know that you were going to do it and then the inquiry was halted and you will come back to it. It is highly significant, and we will await your final report with interest because you are commenting on something for which the public perception is now changing. We are coming to an interesting time in the assessment and testing debate. There is beginning to be more of a clamour to do things differently.

Mick Brookes: I am very pleased to be here as well. It is important that we get beneath the headlines of what all the associations have been saying. The impression that the teaching unions are against assessment is palpably not true. We are for assessment, but it has to be assessment for the right reasons and with the right instruments. If we do not have that, we end up, as Mary said, corrupting the curriculum. You should have received the book from the Commission on Testing by the National Association of Head Teachers.[1] In that Commission, views were gathered from the wider community, and not just the teaching community. It included views from the National Confederation of

[1] Commission of Inquiry on tables, targets and testing

Parent Teacher Associations and governing bodies. It was not just our association that was represented. Anthony Seldon stated: "Children are encouraged to develop an attitude that, if it is not in the exam, it doesn't matter. Intellectual curiosity is stifled and young people's deeper cultural, moral, sporting, social and spiritual faculties are marginalised by a system in which all must come second to delivering improving test and exam numbers." That is where we are. I know from my colleagues that it has altered the curriculum, particularly in the primary sector but also in the secondary sector. I am sure that Brian will say more about that in a minute. It is timely that we come to this now to look at where we go to continue to raise standards in education.

Brian Lightman: What is interesting is the degree of consensus that is already here. I have heard nothing that I disagree with and would not have wanted to say myself. That is a very important message: that we really do feel strongly about this. I want to home in on two things. The first is the examinations system, which has become so costly and complex and is at a point that is completely unsustainable in its current format. There seems to be an assumption that everything has to be externally assessed, which is having all kinds of implications in terms of what we are doing in school, what we are doing for the children, the pressure we are putting them under and the disaffection that we are causing as well as the unhappiness and stress of children. That sounds as if I am going to speak against assessment and I am certainly not going to do that. Like the NAHT, we are far from opposed to assessment. In fact, we are saying that assessment is an integral part of the teaching and learning process. It is absolutely the bread and butter of what every teacher does. It is strange that that aspect of our work has almost been taken away from the professional skills of teachers, because it has been handed over to people outside the classroom. I want to talk about the problem with that. True assessment for learning is something that we are genuinely excited and passionate about as school leaders because, when you introduce those types of technique, you can see immediately improvements in the motivation of the students and quite enormous improvements in the quality of the learning that goes on. That is a terribly important aspect of what we are doing and we need to re-professionalise teachers and train them so that they can use those methods in their teaching. That would have an enormous impact on things like low-level disruption in the classroom, the motivation of students and the progress that they make. ASCL is providing in its paper a proposal for chartered assessors that we see as a solution to the problem. We do not want just to talk about a problem. We are saying that we understand the need for assessment for accountability and we understand that assessment needs to be robust, valid, reliable and so on. Therefore, we propose a model whereby teachers can be trained in their skills and assessment and we can build that into our work. That would be much better value for money and a much more efficient system. I could say a lot more about that, but by way of introduction, that will do for the moment.

Keith Bartley: I became Chief Executive of the General Teaching Council in March this year and one of the first things that impressed me was the range and the nature of the evidence and research work with teachers and with parents that underpinned our submission to you. I hope that you have had access to that. The General Teaching Council was founded and exists in the public interest, and I share the consensus that you have heard about already. We feel strongly that this country needs an assessment system that more effectively supports learning and promotes higher achievement. We are very much here in terms of a statement of intent to help the Government to find a means by which benchmarked information about schools in the public domain is valid, reliable and illustrative of the progress made by children. I will not go through them now, but all our proposals were submitted with that objective in mind.

Q129 Chairman: Thank you all very much. It could be understood that the arguments that you have all made—I doubt that this is what you believe—are against all forms of externally moderated examinations. I would like you to talk about where you feel externally moderated examinations ought to fit into the system and why.

Mick Brookes: You are quite right; that is not that case. We think that we should place greater reliance on teacher assessments, as Mary said, but it would need to be moderated. I know, from my experiences as a head teacher, that you can have two teachers with a parallel year group, one of whose glass is half-full and the other whose glass is half-empty, and who might assess something such as writing, for instance, which has a degree of subjectivity about it, at different levels. There needs, therefore, to be something there. The Scottish system is worth looking at. They have a bank of benchmarked tests, from which schools can draw, in order to check on the validity of teacher assessment. We are not against external assessments; in fact, it is important to have some benchmarking. Nobody in our association wants to return to the 1970s when you did not know what the school up the road was doing, let alone a school at the other end of the country. There needs to be some benchmarking and an idea of where schools should be, but we are saying that we need to test for the right purpose. The current testing regime is used for far too many purposes.

Brian Lightman: I am sure that we would agree with that. There is certainly a place for external assessment, which will increase as you go higher up the age range. We do not suggest removing A-levels because they are important external benchmarks, but there should be an appropriate range of assessment methods. However, going further down the age range, we need to think whether it is really necessary for material at, say, Key Stage 3, to be marked externally, bearing in mind that it is marked by the same people as those in the school at the time. Does everything need to be externally arranged? Do we need a system by which we send things away? Given technology, should we not, as the NAHT suggested, adopt a system by which, for example,

you could download assessment material? We should use new technology to download new material and use it when we are genuinely ready. That does not mean that everybody does the same test on the same day and in the same room, but that when you are ready, you draw down those resources to a certain standard.

Q130 Chairman: Sorry to interrupt you, but is that not exactly what is proposed in the single-level test?
Brian Lightman: No, I do not think that it is. At the moment, everybody across the country is doing the same test on the same day, so it is still an external test. We suggest having a bank of assessments, which the professionals should be trusted to draw down and use. When a class, or group of students within a class, is ready to be assessed, they could draw down those materials and apply the assessments to those students. We would have to ensure the appropriate external moderation. That could be helped by the model that we put forward of chartered assessors, whereby qualified people moderate both within and outside the school.

Q131 Chairman: In St. Cyres school, which you headed, did you find that occasionally you would pick different examination boards for different subjects, because of questions about whether a board is easier in some subjects than in others?
Brian Lightman: I am sure that that has happened in every school in the country.

Q132 Chairman: I am not picking you out, but just asking for your personal experiences as a head teacher.
Brian Lightman: In my experience, in all of my schools, including when I was a head of department in Surrey, we would change our syllabus, partly depending on how we felt that we could get the children through exams. That is bound to happen in a culture in which everything is looked at and accountable. We should be choosing assessment materials that reflect the kind of teaching and learning that we want to have. Given that you mentioned St. Cyres, I should add that it is in Wales, where we do not have Key Stage 3 tests. Interestingly, given the changes there, there is now a genuine debate among heads of different subjects about what constitutes effective learning at a particular level within each subject. Heads of department of different schools are getting together and really thinking about that moderation process in a way that I have not seen in the past 15 years or so.

Q133 Chairman: It does not seem to be producing better results in Wales and the rest of the country, but nevertheless—
Brian Lightman: Well, I think it is.
Keith Bartley: Going back to your question about external moderation, our research tends to suggest two things. One is that we think that public exams should be more about learning, which means that what is examined needs to be broader, and that more account needs to be taken of how they represent what has been learned. The second is linked to that. At the moment, most of our public exams have extremely high stakes. They are used for many purposes—that is the point that Mary started with. An externally moderated examination tells us how well a young person has achieved, comparatively. The scores are then aggregated to give us a sense of how well a school has done, and aggregated further to give us an idea of how well young people of a particular age have done across the country. That multiplicity of uses to which a single examination is put represents stakes that are too high, and it tends to subvert part of the original purpose of evaluating learning.
Chairman: Andy, perhaps this is the moment at which you would like to come in.

Q134 Mr Slaughter: Yes, I was very interested in what was said at the beginning about the types of test and whether they are of a good standard. I shall come to that in a second, but first I shall return to an even earlier stage and see whether I have understood what you are saying about testing in general.From my lay understanding, you are essentially talking about two different types of testing. The first is testing that is internal to an institution—the type of testing that I remember from when I was at school, which is a tool for teachers to use on their pupils to determine whether they are progressing and learning according to the curriculum that they are being taught. I would have thought that it is also used to encourage them to learn, because it provides an incentive, rather than their staring at a blank piece of paper in a test. I assume that that still goes on; it went on a lot when I was at school. I assume that you do not object to it. National tests—I am not so much talking about exams such as A-levels, which have been mentioned—seem to perform a wholly different function: to test whether an institution and its teachers are performing. Do you see testing as I have just explained it, and, if not, what is your analysis? Do you think that the first type is good and the second bad?
Dr Bousted: Well—
Chairman: Sorry, do come in there.
Mick Brookes: We are very well behaved.
Dr Bousted: Yes, we are. We are not going to speak without the teacher letting us speak. If your statement was right and the national tests were used to decide how good an institution was, that would cut down their purpose. You might be able to look at national tests that decide how good an institution is, but that is not the case. The national tests are used to give the individual performance of each child. Our argument is that in looking at an assessment system, we must consider two things: is it valid—testing the key, essential core abilities in a subject that are defined in the national curriculum—and is it reliable? ATL's contention is that the current system is neither valid—it does not test the essential core attributes of a subject—nor reliable. I return to the fact that up to 25% of children, maybe more, get the wrong grade, which has profound consequences for them. Also, a child who only just gets a Level 4 and one who nearly gets a Level 5 might be at very different stages in their learning. It is a broad brush

stroke. The other problem is that although tests are meant to be used to give individual level data, school level data and so on, at Key Stage 2, which has one of the most pernicious stages of testing, the results are given far too late to be any good for a child. The child goes on to secondary school, and secondary schools do not believe the grades because they think that primary schools train children to take tests. Indeed, independent evidence from the QCA proves that to be true. The children are then retested at secondary school because there is no confidence in the grades given by primary schools. The idea that the national tests are used just for the national picture is not right. The problem for children is that if you are told at seven, at 11 and then at 14 that you are not very good, it is perfectly logical to say, "Well, if I'm not very good, I won't try." If individual children are told that they are not very good and not told why they are not very good—they might actually be quite good, but they might have been given the wrong grade—that will have a profoundly pernicious effect on lots of them as individuals. So, I would contend with your outline statement that the two types of testing are for two completely different purposes. I do not think that is the case.

Q135 Chairman: How much are the figures that you are quoting a reflection of the fact that the levels and the curriculum are different at Key Stage 3 and Key Stage 2, although the assessment levels are the same?
Dr Bousted: What, if you get a level—

Q136 Chairman: If you get a Level 4 in the Key Stage 2 test and a Level 4 in the Key Stage 3 test, you are being tested on a different curriculum, are you not?
Dr Bousted: Yes, you are, and it does not mean the same thing. Level 4 at Key Stage 2 and Level 4 at Key Stage 3 do not mean that progress has not been made. You are being tested against a different curriculum and a different assessment framework.

Q137 Chairman: Were you quoting those figures to suggest that the original test was wrong? I was wondering whether the figures that you were quoting were a reflection of the different tests, rather than the wrongness of the first result.
Dr Bousted: The figures that I am quoting are a measure of the confidence that you can have in the fact that pupils are getting the right grade in the tests at each level. The reason why there is such a problem in the level of confidence is this. At 11 and 14, we assess whether a child is proficient in English, maths and science through pencil-and-paper tests. That has particular problems for the science curriculum, because you can test only a very narrow element of the science curriculum with pencil and paper. There are also huge problems with the English tests at Key Stage 3. There have been problems with their validity and reliability since their inception, and there has been a lot of political fury about them. The issue in terms of validity is whether the test actually relates to key concepts in the national curriculum. There is an argument that it does not, because what you can test with the test items is so narrow. That means that although a child might get a certain level in a test, it might not be—our argument is that, too often, it is not—reflective of their ability. That is equally damaging regardless of whether that goes up or down—whether they are assessed at too high or too low a level.
Mick Brookes: To come back to your question, some proof of the value that the profession attaches to the results of testing can be seen in the number—QCA will provide this data—of year 3, year 4 and year 5 tests that are purchased by schools to check on progress and teacher assessments at the end of the year. The difficulty with testing is not so much with those things that are easy to test, such as mathematics and comprehension. To pick up what Mary was saying about the validity of the tests, there can be really interesting variations, with the same teacher at the end of primary school—and perhaps all the way through—scoring something like 85% with their children in a reading test, but only 75% or less for writing. The variation in the national levels achieved in reading and in writing, which are often misunderstood by the press, is huge. Why is it that these results are so different if the same teacher, with the same skills, is teaching the same children for all those tests? I think that it has something to do with the assessment of writing. To take just one example, I know of an extremely good school that had very good writers, but the whole year group misunderstood the genre of the writing that they were supposed to be producing for Key Stage 2 SATs and none of them achieved their levels. That meant that the school was in deep trouble with the inspection system—quite unnecessarily so, given that there is over-reliance on the results of testing and not enough attention given to teachers' assessment of the actual ability of children, who, in this case, just made a mistake on a particular day.
Keith Bartley: I would like to go back to the premise about the assessment that is undertaken to inform learning and the assessment that is undertaken perhaps to give information about the effectiveness of a school. Parents told us very clearly that they felt that the information that was published about tests in a school was about the school justifying itself publicly in terms of its place in the national pecking order. Actually, the information that they valued about how well their pupils were doing was that which teachers gave them, which was very largely drawn from the teachers' own interactions with the pupil, whether that information was in written format or—most particularly—they were given the opportunity to talk to teachers about how well their children were doing. As well as questioning validity, we would question the utility of those tests in terms of the audience.

Q138 Mr Slaughter: That was really where I wanted to go. I was a little concerned about a comment that was made about failure. Obviously, one does not want to label children as failures, but I assume that it is common ground that we regard testing as part of teaching and learning and an essential tool, and that therefore there are going to be people who succeed or fail—that is what happens as a

consequence. I want further comment on that, but I would have thought that it was a starting point. Accepting what you just said about the validity of different types of tests, do you think that there is any validity to national testing in that way, or can the positive aspects of testing simply be dealt with at school level, with the judgment with respect to institutions being dealt with in other ways, such as through Ofsted?

Brian Lightman: I think that one of the problems with national testing is that you are applying it across the board and taking a snapshot of everybody at the same time, and there are other ways of sampling what progress—what learning—has taken place, if we want to have those national benchmarks. That is one of the things that we proposed in our paper. We used to have the assessment of performance unit, which sampled children's progress in different areas and looked across the whole country at one aspect of their learning. By doing that you can really see what children have actually learned, rather than trying to test across the board by giving everybody the same test on the same day and trying to cover everything, which, of course, you cannot possibly do in an hour, or an hour and a half. The other point I want to make is that testing is only a small part of assessment. There are other very valid and effective—and, in fact, proven—methods of assessment, like approaches to assessment such as externally moderated portfolios. Within things like BTEC at the moment there are some very successful models. The people who are doing that assessment have to be trained and accredited, and they have to meet very rigorous standards. You are then able to assess the work of students over a longer period, within that very rigorous framework, to make sure that you are not just doing a snapshot on a particular day. There are all kinds of things that come up, and we have experienced them over the years. Every teacher will tell you about results of tests that they have been mystified about and when they just do not understand the result that a student got in the test, given what they have seen every day in the classroom, because obviously they see the child over a longer period of time. Children get nervous in tests and underperform in tests, and so on. I think that we have to be very careful about how much credence we attach to one method of assessment.

Mick Brookes: Just on the headlining of what happens, I think that children who have overcome significant special educational needs and have reached Level 2 or upper Level 3 at the end of Key Stage 2 are what we have called the invisible children. They do not appear. While people say, "Well, they are in the contextual value added tables", what newspaper picks those up? What is reported is simply those children who have achieved Level 4-plus at the end of Key Stage 2, which also gives a misleading view, so this is not just at a pupil basis, it is also at a schools basis. I have the permission of head teacher William Ball to tell you this: New Manton primary school in Worksop, Nottinghamshire, has always been down at the bottom end of the league tables because they are norm referenced. You nevertheless get ill-informed people saying that it and others like it are failing schools. Here are three sentences from New Manton school's Ofsted report: "The excellent leadership of the head teacher is largely responsible for the good level of improvement in all areas of school life . . . The staff show a strong commitment to the personal development of individual pupils and, as a result, most make good progress . . . The very effective governing body is showing an equal determination to bring about change for the benefit of all pupils." The school is good, but it is down at the bottom end of the league tables, so there has been a distortion of fact. There are very good schools that work against the odds to produce higher educational qualifications than they have ever had in their areas, but they are disabused of that excellent work on an annual basis.

Q139 Mr Slaughter: I have a lot of sympathy for what you are saying—I am sure that we all know of similar schools in our constituencies. In shorthand, there are good schools that provide a good level of education, but are not in the top quarter of the tables—they might even be in the bottom quarter. I am asking about very basic stuff, and I shall shut up after this, but I want some clarity. I felt that by dissing everything about the tests—their quality, reliability and so on—you were not confronting the question of whether we should get rid of them altogether. Obviously, there must be some way in which to assess institutions; there are many ways, but there must also be some oversight—I used the example of Ofsted. What would you like to see? I would like you to confirm that there is a positive role for testing in schools, including primary schools. Are you saying that testing should be conducted entirely within an institution? What would you like to see done about external accountability, including nationally?

Chairman: Each of our witnesses would like to respond to that question, so you should all be quite brisk.

Keith Bartley: There will never be a time at which information about testing in schools is not in the public domain and viewable. We accept that and, indeed, we support it, for comparability purposes. However, the information conveyed by tests should be accurate and valid in terms of what they measure, and tests should not distort the curriculum and learning. At present, the multiplicity of uses to which a single test is put narrows the curriculum and distorts the outcome. To pick up on the question that the Chairman asked Dr Bousted earlier, Dylan Wiliam's view is that some of the fall-off between Key Stages 2 and 3 occurs because most of year 6 is spent drilling youngsters for Key Stage 2 tests, and they forget completely over the summer because they were coached only to climb that hurdle on that day. Removing the high-stakes nature of the testing will be valuable in future.

Brian Lightman: We are not arguing that you should get rid of testing. As others here have said, we are concerned about how test results are used: they produce simplistic league tables and feed misunderstanding. That has been evident in the

coverage that we have seen in the past week. The publication of the league tables has been completely misleading as to the actual meaning of the tests. There should be testing. Using a model whereby people can download high-quality assessment materials and use them at the right time would measure different, important things.

Mick Brookes: I agree with my colleagues. There is already a company that does online testing. Children take the test online, and they are assessed not only at their own level, but that of the whole cohort. There is already a quick and easy expedient. There must be testing, and it must be nationally benchmarked—otherwise, schools will not know where they are—but I agree that the testing system goes wrong because of the multiplicity of purposes to which the tests are put.

Dr Bousted: Yes, assessment including testing is a key part of the repertoire that teachers must have at their disposal, but that does not go for the tests that we have at the moment. The tests must be as valid and reliable as they can be but, at the moment, our testing system is corrupting. It does not just corrupt the results; it corrupts all the other things that teachers are trying to achieve, like a broad and balanced curriculum, a varied menu for pupils, and valid assessments of where pupils are and what they can do. You said that pupils have to experience failure. At some point, yes, they do. At some point there has to be sifting—a proper sifting. Children and young people, no matter how they look on the outside, are fairly fragile on the inside, just like the rest of us. What is more important than failure—failure does nobody any good in the end—is that pupils need to know where they are now, and what they need to do to be better. They need to know where they are at. They do not need to know that they have failed. In the end, failure does not get anyone anywhere—they just fail. It does not teach them how to do better. What they have to know more of is why they are at a particular stage and what they need to do to get better. I would say that the over-emphasis on testing means that where a child is now and what has to be done to enable that child to learn better is the most undeveloped aspect of our education system. It is one of the reasons why, in the PISA league tables, we are not performing as we should. We have one of the most undeveloped systems of assessment for learning among developed countries. It is parlously poor in our country.

Chairman: On that note, I am going to ask David to speak.

Q140 Mr Chaytor: May I put my first question to Keith? The chief inspector may tell us in his annual reports that we have the best generation of teachers ever and that standards of teaching are rising year on year, but what is the relationship between that and the testing regime? Since the testing regime came in and the publication of league tables became the norm, standards of teaching appear to have risen. How do you explain that relationship?

Keith Bartley: First, I do not think that it is a causal relationship. Secondly, I observe that the implementation of the national curriculum in the late 1980s had a significant wash-back effect on initial teacher training. There is no doubt in my mind that teachers who are moving through into qualified teacher status now have a higher level of teaching proficiency within that curriculum because the curriculum is now much more closely defined than ever before. Year on year, we have seen a rise in the quality of the training experience. Running alongside that, we are also seeing a rise in the levels of qualification of teachers. Last year's cohort of qualified teachers was the highest qualified that we had ever seen in our schools. As for the question of why those higher standards of teaching and better qualified teachers are not manifesting themselves in improved levels of achievement, I would argue that during that period we saw a significant increase in the value placed on the tests. At the same time as we have a curriculum—indeed, there is a very high level of encouragement from Government through their curriculum policies to broaden and experiment with that curriculum—all the time, and particularly at the ages of 11 and 14, youngsters are being narrowed down by a very narrow system of testing. That is the bit where we are not unlocking the achievable.

Q141 Mr Chaytor: Why do you assume that an improved quality of teaching is not leading to improved levels of achievement? Formal evidence from league tables, Key Stage tests, GCSE and A-level results, and vocational qualifications suggests the opposite.

Keith Bartley: What we have seen, all the way through, is that shortly after testing has been introduced, there is an increase in standards. Typically, in the first few years after the implementation of testing for a key stage, we see a steady increase in achievement and standards, and then we see it tail off. That is very much the case with any performance measure, whether it is to do with schools, or with the punctuality of trains. What happens is that people work towards a performance measure, and it quickly plateaus. That is why you need to have a much broader band of assessments that can be used.

Mick Brookes: Performance at the end of Key Stage 2 has been largely stuck for about five years, and until we start to do things differently, it will remain stuck. There have been improvements, not only in children's knowledge and understanding, but in teaching. We must be careful not to polarise the issue. Going from the 1970s and 1980s, there was, indeed, a need for much greater veracity of thought and practice in our schools. Although that might have been a painful process for schools to go through, we have gone beyond it. We now have sophisticated systems of assessment, which we should use, and, according to Ofsted, we have better standards of teaching in our schools. That worked then, but we must consider where we go next.

Mr Chaytor: Mary wants to come in.

Dr Bousted: I should like to answer the question directly. Ofsted does not investigate the quality of the tests. It takes the test level data and says that they indicate a rise in standards. If you examine the recent research by Peter Tymms, which was in the Robin

Alexander review, he undertook a long inquiry into how difficult it is to measure standards longitudinally. In fact, two years ago, ATL published, *Standards in English Primary Schools: are they rising?* by Colin Richards, which also examined the issue. Tymms and Richards said that standards in reading had risen marginally, but nothing like what the tests state. Tymms makes it quite clear that when the Key Stage 2 reading tests started, for the first four years they found it very difficult to get a foundation level. The baseline against which they tested whether standards had gone up or down changed. It is also clear from a paper by Mary Hilton, on English in education, that three years into the Key Stage 2 tests, the level of questions that asked for an inferential response—not, "Did this happen in the passage?" but, "Why did it happen; what were the motivations?", which is a much harder question—went down. Tymms argues that only since 2000 has the baseline level against which you measure year on year, by each cohort, become steady. If you consider the rise since 2000, it is nothing like as great as in the first four years, so Ofsted will say that standards have risen dramatically because it does not question the evidence base. The question that we must ask is, "What is the validity of the evidence base?"

Q142 Mr Chaytor: May I link that to the point that you made earlier about 25% of pupils not getting the grade that their ability merits at the Key Stage test? Surely the question is, who defines what their ability merits? It must be true in any form of testing and assessment that the subjective judgment of the tester comes into it. There are bound to be errors, surely?
Dr Bousted: There are, but you must consider the level of confidence error that is acceptable. Part of the problem that we have with the tests is that they test a very narrow range, so it is particularly difficult in science to get a valid pencil and paper written test that tests the key concepts in science. They are taken on a certain day, at a certain time, and you can have only very narrow test items. You could have tests with different test items that tested other parts of the syllabus, and they would be equally valid but give a completely different result. In other words, our system has gone for high reliability: it is likely as not—we will put a lot of effort into it, and be as sure as we can—that the kids who take the test will get similar grades if they have a similar level of ability. We have problems with that, but they are also highly invalid.

Q143 Mr Chaytor: May I move on to the question of the role of teacher assessment? Perhaps the question is directed to Brian. What does the evidence say about the relationship between teachers' judgment of ability and the Key Stages 2 or 3 results? Would it be possible to abandon completely external testing and replace it with an entirely teacher-assessed system?
Mick Brookes: Certainly, by moderated or accredited teacher assessment, but I think using—

Q144 Mr Chaytor: Sorry, what do you mean by accredited?
Mick Brookes: Accredited would mean using a bank of tests, so that if a teacher said, "I think 80% of children in my class have got Level 4, but how come the test doesn't say the same thing?", you then have that professional conversation with your member of staff, not about those who scored the same, who are easy, but about those who scored differently. You must ask, "What's happened here? Is it your assessment? Is it that they were sitting next to Fiona, and that is why they got a higher mark?" It needs to be brought back into the school and discussed in an atmosphere where professional integrity is encouraged. If we could do that we would have a system that is fit for our children and does not completely subvert the year 6 curriculum, for example. I could take you to a school where they cancelled Christmas for year 6 kids because they were so worried about—

Q145 Mr Chaytor: You have not told the *Daily Mail* this?
Mick Brookes: They want to know where it is, but I have not told them.
Brian Lightman: I absolutely agree. You have to have a clear definition of the standard that makes a Level 4 or Level 5. That is a problem: there is a big discussion about what constitutes a Level 4. You can test a little bit of it or you can do an overall assessment. To do that, you need to be able to draw down appropriate, high-quality materials that can be used to assess reliably where the students are. That sort of thing needs to take place. That professional conversation has been missing over the years, because people have relied almost entirely on external tests, rather than sitting down together as groups of teachers and saying, "In geography, which is my subject, a Level 4 is x, y or z; this is what constitutes it, and we need this evidence and that evidence to see that." Part of that will be a test and other parts will involve looking at students' written work and, perhaps, using oral assessments. There will be a whole bank of things. All of that put together and properly moderated will lead to a much more robust assessment, which will go back to the formative side of things—we can then advise students about how to improve.
Chairman: David, a last one.

Q146 Mr Chaytor: Finally, in terms of high-stakes testing and the publication of league tables, if there were a movement towards the greater involvement of teacher assessment, should that information be published in the league table list, or are you saying that they should be taken out completely?
Brian Lightman: The results of all that assessment?
Mr Chaytor: Yes.
Brian Lightman: I think we have to ask how much we need to publish in the form of league tables.
Mr Chaytor: I am asking you how much you think we should be—
Brian Lightman: My argument would be that there is far too much going into the league table. We are adding more and more. Parents need and have a

right to an overall picture of how the school is doing and how their child is doing individually. The move towards real-time reporting, where parents can see how their child is doing individually, is exciting and positive. At the other end, the whole school accountability side of assessment—which is a completely different purpose—could be published in much more effective ways than league tables, which often do not compare like with like, talk about the context of the school or look at raw data.

Chairman: I know that you all want to answer, but if you have anything to say that is different from what Brian said, that would be the most useful.

Dr Bousted: I want to go back to the issue of teacher assessment. Part of this discussion is predicated—or, rather, most discussions are, but not this—on the idea that teacher assessment is unreliable and invalid and that test assessment is good. Currently, at Key Stages 1, 2 and 3, you have teacher assessment scores against test scores. The issue is that teacher assessment scores count for nothing: they are not used by Ofsted and are not in the performance league tables. Teacher assessment is, however, usually significantly and consistently lower by 2 or 3 percentage points, so there is no reason to believe that teachers would make things up in a properly trained and moderated system.

Mick Brookes: May I say an unequivocal no from the NAHT? That information should not be published in the league tables that we have now. It simply tells you where rich people live and, sadly, where poor people live as well. It is deeply undermining for those communities, teachers and children. I cannot understand why a Labour Government have allowed this to go on for 10 years.

Chairman: Adam.

Q147 Adam Afriyie: Thank you, Chairman—or acting Chairman. This is my first official outing on the DCSF Committee, and I am delighted to be here. My background is that I come from the Science and Technology Committee, so I have been fascinated to hear some of the comments about how assessment works, the baselines, the changes over time, and whether things can be compared. I want to concentrate my two questions on the single level tests, to examine them as briskly as possible, and to obtain your views on the proposed progression tests and the league tables that they will be placed in. Is it a good thing to use single level tests, or is it better to stick with the key stages or take another route for assessment?

Brian Lightman: The first thing is that placing single level tests in league tables implies that everyone will have to do them at the same time, which flies in the face of personalised learning. If we are talking about personalised learning, we must be able to apply those tests at the right time. Trying to force the entire cohort through that at a certain time makes nonsense of the idea. The other thing is that we can use new technology for genuine testing when ready, which we must do, and we then return to the idea of downloading a bank of tools.

Q148 Adam Afriyie: It is interesting that you have made the point about testing when ready several times, and I have heard it loud and clear. Would you propose something more along the lines of allowing pupils to progress when they are ready or hit a key stage, or move on fast when they are ready, or is that just purely the testing format rather than what happens with pupils in the year in which they sit in school?

Brian Lightman: The only thing I have an issue with in the way you expressed that would be if you are saying that they should be allowed to progress when ready. We, as teachers, want to ensure as much challenge as possible, so we do not want to tell children that they can take as long as they like. We want them to progress. That is one of our issues about two levels of progress within a key stage, because that is an arbitrary decision. Key stages are different lengths. For example, there is no comparison between Key Stages 2 and 3 in length and time, so why talk about two levels? Some students could progress three levels, but it would be unreasonable to expect others to do so. Also, the levels are not equal. The idea is not as simplistic as the music grade analogy that is sometimes used.

Dr Bousted: Briefly, we think that making good progress is fraught with difficulties, and we do not think that the Government have thought through the matter. The single level test, taken up to four times a year, could lead to a huge proliferation of the test, and all the issues arising from that—narrowing of the curriculum and teaching to the test—could be ratcheted up to four times the present level.

Q149 Adam Afriyie: So you think the Government are just plain wrong.

Dr Bousted: Yes, just plain wrong. There is also a question about the validity of the levels. If a single level test is shorter, how valid is it and how worth while is the information that it provides? We agree absolutely with the Association of School and College Leaders. We all know that a less able child will find it far more difficult than a more able child to move two levels. That has been shown already in the current progression rates. Schools where less able children congregate are often in more challenging areas or where children have English as their second language. They would be penalised by a progression measure that they could never achieve. In schools where children arrive with Level 5 they are much more likely to be able to move two levels than those in schools where the majority of children arrive with Level 3. It is much more difficult for them to make that progress because they learn more slowly. Yet that school might be doing an equally good job, but be penalised by a completely inappropriate progression measure. That has been scientifically determined by the levels, so why the Government have chosen this measure, I do not know. Finally, a system of repeated and proliferation of testing is like excellence and enjoyment. The Government think that you can put in excellence and enjoyment, and have a high testing regime and a broad curriculum, and that just by putting them in the same title it will happen. They

also think that as long as assessment for learning and progression level testing are in the same document, that is all right. You will not get assessment for learning in the system. They could have done so many things, but this is not the right thing for them to be doing.

Q150 Adam Afriyie: Thank you. And Mick, do you agree with that view?

Mick Brookes: I would like to encourage the Government to move on from where they are. The concept of having a bank of tests that are appropriate for a single level is right, but it is the way in which it is being rolled out. To do the right deed for the wrong reason is the greatest treason. The tests genuinely have to be for schools' use in assessing their children. That assessment is then reported to the right people: the parents, the children themselves and, of course, the governing body. The problem is trying to use the same tests to judge the performance of schools. I think that we have made that point.

Q151 Adam Afriyie: I hear your point. You are arguing for separate tests or assessments for teachers, for the school and for the pupil—they are not necessarily the same thing.

Keith Bartley: May I just put a slightly different take on that? At the moment, our reading of the *Making Good Progress* pilots is that they are still about testing when the system, rather than the child, is ready. Actually, if you were to place a greater emphasis on the validity of teacher assessment, and if you were to provide teachers with a bank of tests that they can draw on at an appropriate stage to confirm their judgment of where a child is in a particular aspect of his or her learning, you would change the system. The learners' needs would drive the assessment and, therefore, the subsequent teaching. *Making Good Progress* is a step in the right direction, but we would particularly like a robust evaluation of its impact and the possibilities that it is beginning to open up.

Q152 Adam Afriyie: I think we already had the answer to my final question in previous answers. Do you consider that it is in any way possible to disconnect summative assessment of the children from the monitoring and performance of the school or the teacher? Is that possible?

Mick Brookes: Yes, I think it is, by sampling. Of course, we must retain some idea of national benchmarking to see where things are. That is important because this is part of the public purse and therefore the public need to know how well education is faring across the country, but that could be done with sampling rather than going to every school. Going back to the previous question, I think an analogy with the driving test is a good one. Some people might need five lessons before they are ready to take the test; others will need 20 and some might need 120.

Brian Lightman: I would agree with that.

Q153 Lynda Waltho: My feeling is that testing and assessment is coming out as, "Yes, we need it, but perhaps not in the way we are doing it." I do not get the same glowing feeling about league tables. Without league tables and other value-added measures, can you suggest a measure of school performance that can be used by the Government for accountability but also by parents for their decision making? Does that exist? Is it possible?

Mick Brookes: I think that parents are being grossly misled by raw, or league table, data, and, indeed, by some Ofsted assessments that are driven by exactly the same set of figures. That creates polarisation. When it comes to admissions, people will travel and do all sorts of strange things to get their child into the school of their choice because of that polarisation. Actually, their local school may be very good, but the only way they can find that out is by going there and having a look. We think there should be an emphasis on parents, who should make sure that they visit their local school and see the work that it is doing. As an experienced head teacher, I could walk into any school and within five minutes tell you whether it is a good school. Parents then need to look at the results, but they need to look at them in the context of their community. That is not necessarily about poverty. It is also about the expectations of the community in terms of education qualifications, which is something that Ofsted does not take into account. The level of higher education qualifications in the community is a key factor in whether the children will expect to progress in education. We are doing the job of moving those expectations on, and we need to keep doing it. Parents must go and see.

Brian Lightman: What are the criteria for a good school, and what are the criteria that parents look for? I show parents around the school. I strongly encourage all parents to come round to our school before they make their decision about whether to send their child there. They want to know whether they are going to be happy, safe and encouraged to make the best progress they can. It is those types of things that they want to know. Over and above that, parents look at our school prospectus. They read the school prospectus, which contains, by statute, a whole range of indicators against the school's performance and they can see a detailed account of the school's performance. I know I am talking from a different context, where I work, but that is a much more valid approach than putting the information up, as it was done in the paper last week, where the top 10 schools in any authority—the best 10, they were called—were published, whether it was an authority with 100 schools or one with 10 schools. That is the way they published the league tables last week. That does not tell you which ones are the best schools. Looking at the broad picture of a school is the way in which you can judge its quality—and I think there are plenty. There is also the Ofsted report—I hope you know—and all the information

in the public domain. There is a vast amount of information enabling a parent to see what a school is like.

Q154 Lynda Waltho: So we are never going to be able to do it with a league table. Is that the feeling?
Brian Lightman: I do not think that we are going to have the sort of accurate information, which gives a true picture, with league tables the way they are.
Dr Bousted: The problem with league tables is that you can write whatever you want, but in the end the drop-down system for league tables is so pernicious in other respects. Let us consider a school in a challenging area that is working hard and making the good progress that we know a school can make. Variation between schools is much less than variation within schools. We know a school can make about 14% of the difference; the rest of the determining factors on a child's achievement come from their background, actually. And 14% is a lot; I am not undermining what a school can do. However, that means that schools in the most challenging areas have to work extremely hard to get the results that they do. It is up to Ofsted to go into that school. We have one of the most accountable systems in the world. Ofsted goes in and looks at whether the school has sufficiently challenging targets for pupils—whether it is sufficiently demanding of their progress, whether the behaviour in school is appropriate and whether the teachers are working well enough. Ofsted looks at a range of issues to do with the performance of that school. But if you are in that challenging school, doing very good work and adding clear value to the children's lives in that school and you are still, because of the nature of your intake, going to be at the bottom of your local authority league table, that is not a good thing. I will tell you another thing: it is not just not good for the teachers, but it is not good for the pupils going to that school, who need to feel good about themselves.
Keith Bartley: May I share some of the research that has been commissioned from parents? MORI, in 2005, discovered that parents attributed very low value to league tables in assisting them to determine their choice of school. That was partly because they felt that the information that they portrayed was to some extent confusing. For others, it did not match what parents saw as being the really important things about schools. The list that Brian gave captured those things. We undertook some further research through the British Market Research Bureau about what parents found most helpful in determining the quality of a school. They looked for a much broader range, particularly in respect of contextual factors: the nature of the young people that the school educated, the kinds of aspirations it set out for those young people and the way in which it delivered against those. That wider bank and portfolio is what parents said they would find most useful. League tables, parents tell us, are questionable in terms of their value. The further dimension is that as our 14–19 curriculum unfolds, and particularly as we seek to offer young people a much broader range of both education and training opportunities through to the age of 18 or 19, the influence of an individual institution is going to be much harder to measure through league tables, because young people will have been involved in several institutions—maybe a college, maybe two schools or maybe a training provider—so their validity will become even less as our system better meets the needs of that older group of pupils.

Q155 Lynda Waltho: In light of that, although it is slightly unfair to mention this because possibly only one of you may have read *The Western Mail* today—*[Interruption.]* Okay. Actually, it refers to Wales. It published comments by Professor David Hopkins, who has said that school tests and league tables should be brought back. In fact, it said that, "Statistics show children in England, where testing the two ages and league tables remain in place, performed better last summer in key exams like GCSEs and A-levels." Professor Hopkins said, "We very much know that the performance at seven correlates to success at 16 . . . The previous system was too harsh but Wales went too far the other way." That is also backed up, to a certain extent, by Professor David Reynolds, who described Wales's PISA—Programme for International Student Assessment—ratings as "awful" and said they are falling behind England. I am sorry to zero in on you, Brian, but obviously I know that you have a more direct experience of this subject. In the light of that report and bearing in mind your comments earlier, what would you say in response to that?
Brian Lightman: I have not seen that article this morning. However, one of the big discussions that I have had with the Welsh Assembly Government about the way that they have published the results this year is that I discovered a few weeks ago that what they understand as five As to C and what the English league tables describe as five As to C are completely different measures. That is an example of the type of thing that happens. In England, there are all kinds of qualifications that count towards the Level 2 qualification and in Wales they have not got round to including those qualifications. So, for example, my school does the DiDA qualification—Diploma in Digital Applications—that counts as four GCSEs here in England, but in Wales it does not count. So you have to be very careful about how you produce these figures and let them mean what you want them to mean after one year. This is one year's results and one indicator and we must be very careful about the way that we read things into those results. Obviously, in Wales as in England, we want our children to achieve the best possible standards. So we must look at a range of measures and not just one thing. At the moment, it is not possible to compare England and Wales. It is not possible to compare national curriculum tests or assessments in England and Wales, because the national curriculum for each country is different. So there is a real danger that things will be misinterpreted in order to put forward a particular point of view. I would be very careful about that.
Mick Brookes: Of course, the mechanics of this system are bound to want to defend it; David Hopkins was one of those. But I think that he is

absolutely wrong. I think that it is absolutely right to say that comparability between England and Wales is very difficult, but those mechanics need to look at other things. For instance, they need to look at the UNICEF report about the happiness of children in this country. They need to look at all the work that has been done by Robin Alexander on the primary curriculum. There is a lot of evidence now that says that we are simply focusing on too narrow an area and it is having an effect on children's lives and it is certainly having an effect on the curriculum. We must try to measure the things that we value, rather than valuing the things that we measure; that is an old cliché, but I think that it is true.

Q156 Lynda Waltho: You talked there about the comparison between England and Wales, but the Organisation for Economic Co-operation and Development tests—
Brian Lightman: Sorry, which tests?

Q157 Lynda Waltho: Some 400,000 15-year-olds across 57 countries were tested for the first time in 2007. Wales came bottom in the UK for maths, reading and science, and trailed at 33rd place for maths, beneath Latvia and Lithuania. That is obviously a wider comparison, across a wider set of nations, in which England was rated along the same lines. I am just wondering about that test.
Brian Lightman: But, you know, you get those results and you then have to start asking questions. At the moment, there has not been time in Wales for anybody to analyse the evidence base of that test. There will be questions asked about that test, but I cannot draw a conclusion about the results until we see the evidence base. It is the first time that those assessments were used in Wales and we need to look at the evidence base now and ask questions about it. I do not think that it would be fair to draw on the results of that one test for the first time and come to some conclusion that is perhaps making two and two into five before we have analysed what the evidence base is and what those results are saying.
Mick Brookes: And I could direct you to Jersey, which has no SATs, no league tables and a totally different inspection system and which would be top of that league. I do not know whether it is in the table, but the results from schools in Jersey and staying-on rates in Jersey are much, much higher and it does not have the same system as we have, so it depends where you look.
Chairman: Indeed, and Jersey does not have the variety of population. These tests are exactly the kind of tests that I have heard you arguing for. They are like the APU—Assessment of Performance Unit—tests. They are internationally moderated. Anyway, we will leave it there and I will give Sharon a chance to ask her questions.

Q158 Mrs. Hodgson: One question sprang to mind on the back of the decision-making that parents go through and how they use the league tables to make that decision when perhaps they should be visiting the school and making the decision on a whole host of other things, one of which I believe should be the distance that the child would have to travel. My daughter has just started secondary school and is within walking distance of the school. It is a good walk, but it helps to keep her fit and she and my son do the walk together. One of my daughter's friends is late about twice a week because she has a horrendous journey. She probably passes between six and 10 other high schools to go to that one. Obviously, her parents have chosen the school for a whole host of reasons, but the experience that that girl—and probably many other children throughout the country—is going through must be affecting her learning in negative ways. I am thinking of the stress of constantly worrying whether she will catch the bus and getting into trouble when she gets to school late. Eventually that might affect her whole learning experience at school. My children were in school in London for a while and some children commute to school from outside London. They have a commuting distance that an adult would consider a chore. What are your comments on that?
Dr Bousted: In London that is endemic. I get on the train every morning and my carriage is delightfully shared with schoolchildren commuting from one area to another. A long commute to and from school must affect the learning ability of children, notwithstanding what it is doing to the environment and everything else, and of course that relates to the school run as well. This is often about the parents' perceptions of a good school, and the perception of a good school is often based on the class of the intake.
Mick Brookes: I was going to say the same thing. Rather than being based on the quality of teaching and learning, the decision is sometimes based on the fact that the school is full of the children of "people like us".
Dr Bousted: It nearly always is.

Q159 Mrs. Hodgson: So there are parents who want to get their children into the school for that reason?
Dr Bousted: That is right.
Chairman: You wanted to ask about Making Good Progress.

Q160 Mrs. Hodgson: Yes. I am very interested in the move towards personalised learning and then equally, hopefully, towards specialist teachers, especially with regard to SEN. I know we are not talking specifically about SEN, but you might be aware that I have a Private Member's Bill about SEN statistics and information gathering that will be hitting the House on 1 February. The crux of what I want to get to is this. Witnesses have pointed to a contradiction between personalised learning, which recognises that all children learn in different ways and have different abilities and needs, and the systematic targets that assume that children develop at the same rate. The NUT, for instance, is critical of the current practice of diverting resources towards teaching children at the margins of the level. Again, that is to get the best league table results for the school. I want to talk about what I feel we should be doing in going down the route of a more personalised learning agenda. The assessment for learning is the crux of all this; what we need to be

getting to. That is what the test is all about and that would then identify the gifted child right the way down to the child with a special educational need. You use that as a tool for a teaching assessment or assessment for learning. That should be, in my opinion, the whole basis of these tests. I would imagine from things I have picked up today that you agree with that, rather than trying to produce a form of league tables that are then used for all sorts of other reasons. Is an increased emphasis on personalised learning, combined with single-level tests, likely to effect real change in the classroom? In particular, is it likely to lead to pupils being treated equally, so that each is enabled to achieve to the best of his or her ability?

Brian Lightman: Personalised learning will enable each child to be treated equally. However, the issue is about the targets set, not the personalised learning and single level testing. For example, if your target focuses on five grades A* to C, inevitably, the focus will be on those with four and who are nearly heading towards the fifth. You will concentrate on giving those children the extra help. If you are talking about children who have made two levels of progress through the national curriculum, you will focus on those heading towards that, but not quite there. The children who you are talking about—the others—who do not quite fit into those categories, will be left out. That has been one of the major shortcomings of this target-setting culture over many years. For example, the focus of GCSEs has been very heavily on the C-D border line, and not, for example, on students underachieving by getting a grade A, but who could hopefully get an A*, or on those getting a B, but who could be helped to get an A. Genuine personalised learning does not focus on such perverse indicators that make us concentrate on those who will help us meet the target, rather than on ensuring that all children in our schools learn effectively.

Dr Bousted: The question is very interesting. I return to my previous point: too often, the Government believe that you can place contradictory things in a policy document and that they will happen. Personalised learning will not take root in the current system and is unlikely to do so with single level testing on the same days. In order for it to take root, teachers must be confident in their ability to assess where a child is at. In our system, that is one of the things that they are weakest at and least confident on. I was at a conference at which two head teachers said to me, "The teachers in our schools do not have the confidence in their own assessments, because the system is so geared towards exams that the professional competence has not been built up." It needs to be rebuilt, because it is the essence of personalisation—you know where a child is at and what you need to do to take them further. We do not know enough, in anything like enough detail, about where children are at in the system. Interestingly, we surveyed our members on assessment recently and got quite a big response—from about 400 members. They were asked whether the national system of external assessments supports a range of things, one of which was personalised learning. Some 83% of the 400 correspondents either disagreed or strongly disagreed that the current system provides the bedrock and foundation in which personalised learning can take place. The danger for Governments of all persuasions is that, although a policy document might sound and look wonderful, and be full of high-minded ideas, it will have no real effect in the system. That is the danger with personalised learning.

Q161 Chairman: But, Keith, did you not say earlier that teachers are better at assessments now?

Keith Bartley: No, I said that teachers are more adept at teaching the curriculum, because of the way in which it has developed, which means that they are also trained in assessing within that curriculum. My reservation was about the narrowness of the elements in the curriculum that are tested. I was separating testing and assessment. May I respond more generally to the point about assessment for learning and personalisation? That is at the heart of the matter. Teachers and children exploring what they have learned, and what they need to learn next, is absolutely central to taking forward an examination system that examines us according to outcomes and products—if you like—including whether we use the OECD and other measures such as the programme for international student assessment. However, that requires considerable investment in teachers' continuing professional development, because of the issue about what they have been trained to do thus far in their teacher training. Teachers tell us that they would love to be able to explore more, with other teachers and their own pupils, ways in which they can better understand what pupils have learned, to be able to draw down tests to confirm that, and to help them set targets for what they need to learn next. On your starting point, about special educational needs, one of the greatest concerns that we have is that we are now losing that generation of teachers that were trained as specialists in special educational needs. We are also concerned that many of the skills and talents of our teachers who spend most of their time teaching children with special educational needs—things like the use of P levels and very fine graduations of understanding of learning—are in danger of being lost to the training element of the system and being compressed into a narrow population, when actually they are skills that all teachers need.

Mrs. Hodgson: Mick wanted to speak.

Chairman: I am worried about the time; I know that some of our witnesses have to leave and I am trying to get us going.

Mick Brookes: I would like to say a little about target-setting and how important that is, and what a precise science it is for an individual child. If you set targets too high, the child cannot do it, becomes frustrated and disconnects. If you set that target too low, the child becomes bored and disconnects; they then leave school as soon as they can—24% of them. So target-setting is a very individual and personalised event. I would suggest that it cannot be done from the building just down the road here.

Q162 Mrs. Hodgson: The Government signalled a formal change in approach, which is what we are discussing here, such as personal learning and teacher assessment. Without some form of league table, how can the standards of teaching effectively be monitored? What, in your opinion, can be done to monitor the effectiveness? You have already sort of answered that question because you have said that we cannot do personalised learning under the current regime with league tables.
Dr Bousted: We have one of the most monitored systems in the world, but we do not monitor in very clever ways. Going back to what all of the witnesses said, we need far more about cohort sampling. If you do enough cohort sampling in the key subjects, you can test much more of the curriculum because not every child needs to do the same test. They can do the tests at the same level, but they can test different items. If the tests are statistically significant, you can get a much wider range of test items, which is much more valid. We need to do much more of that. We have been very poor in doing cohort sampling which monitors standards over time. In fact, when we moved to national curriculum tests, we packed up the assessment of performance unit and cohort monitoring over periods of time. We have lost a rich vein of data that used to give us really interesting reports, such as the National Foundation for Educational Research survey about standards in reading over 20 years and whether they had risen or not, with really fine, detailed information about in what types of school standards of reading had risen and in what types of school they had not. We have lost the ability to make that fine, detailed monitoring of the system.

Q163 Mrs. Hodgson: My last question about that matter is, to touch on the stress and demotivation that Mick mentioned, will the single-level tests address the problems experienced by pupils under the current regime? To give one example—I am terrible for giving personal, real-life examples—my daughter has just gone to high school. She got very good SATs and now that she is in high school she is in all the top sets. She never was stressed going through her SATs and I kept saying to her every day, "Are you okay? Are you okay?" Every day now she comes in stressed because, "I can't cope, Mum—I'm in the top set, I don't think I should be in the top sets." Now she is stressed, and I think that she was obviously hot-housed, got through the SATs, got really good results and has now been thrown in the deep end in this high school where she feels that she cannot cope.
Mick Brookes: That is a good example.
Dr Bousted: Yes.

Q164 Mrs. Hodgson: So, do you think that level tests will help?
Dr Bousted: Not necessarily, they might compound the problem. What you would then get at Key Stage 2 is a quadrupling—of four times a year going in for your test again. Because your test is coming up and tests are coming up so much more often, all the other aspects of the curriculum may be equally neglected. With single-level tests four times a year, in the way that they are currently being done in the pilots, I would be interested to see whether there is a narrowing of the curriculum and whether it compounds the problem of teaching to the test.
Mr Heppell: Very quickly, are we sticking to the time?
Chairman: I am trying to, but failing. This is an inexperienced Chair not managing.

Q165 Mr Heppell: I have a question about the new approach in schools. Like you, I have been around a long time, and I was a school governor for nearly 30 years—indeed, I was chairman of the school governors—so I feel that I can walk into a school and see whether it is a good school in half an hour, and I do that on visits now. Sometimes, schools will have had good SATs, sometimes Ofsted and others will have got it wrong and sometimes the SATs will have been wrong. The real problem, however, is that schools change; schools are not static—they can better and they can get worse. I recently went to a school with bad SATs, and the school next door, with pupils from the same sort of area, had good SATs, so it is not about the rich in one area and the poor in another area. When I questioned the school about its SATs, the answer I got was, "We're not interested in SATs. We don't really bother with them at all." In fact, I got so anxious, I actually went and told local councillors about it. The school also told me, "We're involved in making sure that we have personalised learning for everybody and we're assessing everybody's progress as they go along." That actually sounded like a great idea, until I asked to see some of the assessments for the children, but there were none; in fact, I am fairly certain that the bloke I was talking to was giving me a load of bull. That is my worry. If there is no outside testing, how do I know whether things are going wrong, as they clearly were in that school? How do I know that there is not a problem that has not been identified?
Mick Brookes: That is why, in response to David, I said that there needs to be accreditation and moderation, but not what we have at the minute, so that there is some external view of how accurate assessments are. It is a question of adjusting the system, not throwing it out.
Keith Bartley: There are two aspects to that for us. One is that what we measure needs to be more valid—in other words, we need to measure things that tell us something about what children are doing—and the way in which we measure things has to illustrate the progress that they have made. We are not in any sense saying that there should be no form of public accountability; it is just that the measures used need to be much more informed than the ones we have at the moment.

Q166 Mr Heppell: Okay. What progress have schools made already in terms of personalised learning? Who is actually doing things now, unlike the school that I mentioned? How far have people got in terms of assessment for learning? Can we point to excellence or good examples in schools?

Dr Bousted: We ran our fringes on the curriculum at the party conferences last year and we got in head teachers who were adopting innovative approaches to the curriculum. We have just done a book on the curriculum and we have clear examples of schools that are starting to integrate subjects, to use literacy and numeracy across the curriculum, and to integrate curriculum development and assessment. There is beginning to be more confidence in schools that that is a legitimate thing for them to do. We went through a period when the national strategy was so rigid—you teach reading like this and numeracy like that—that schools lost the confidence to think that they had any professional expertise to bring to the party. The Government have moved from the idea that everything can be done from Whitehall, and that is a significant shift. They have a lot further to go, but my experience is that schools are beginning to re-engage with the issues of what is an appropriate curriculum, what decisions should be made at school level, how they can more effectively assess their pupils, how they offer curriculum that meets pupil needs, and what forms of pedagogy are most suitable for pupils. However, the profession needs more support in that, and that is really key. Over the past 10 years, nearly all the CPD has been offered through the strategies. Subject-specific CPD has virtually withered on the vine, and teachers regularly report that the CPD they are offered is not suitable for them or for what they want. Teachers are moving down the long road towards regaining control of those aspects of the learning and teaching process that they should be in control of, but we need to go a lot further and we need support to do so.

Q167 Chairman: By CPD, you mean Continuous Professional Development?
Dr Bousted: Yes.
Brian Lightman: Our association has been doing a great deal of work with the Specialist Schools and Academies Trust on this, and we have seen an enormous amount of good practice developing over the country. It is not embedded yet and it is not everywhere, because it is in a stage of development. We are seeing real enthusiasm because people are seeing the benefits of these approaches in the classroom. We need to continue to support and motivate people and encourage them to develop it further. There is a vast amount of good practice going on around the country. We need to build on that and encourage it to develop further.
Mick Brookes: It goes back to this: people will do what they are expected to do. If they are told what they shortly will do with Fischer Family Trust and other measures, why will they do it themselves? We need to remove that, encourage schools to develop their own systems within the national framework, and acknowledge that the vast majority of head teachers, teachers and all the people who turn up to school have a passion for children's learning. We need to harness that passion, rather than dumbing it down.

Q168 Mr Heppell: You have already described some of your reservations about the single-level test. You might have some more. We were talking about staggering them even more than having a situation of individuals being able to take individual tests. Does that then start to create a problem with resources because you are not doing one test a year, but are having to organise various tests?
Dr Bousted: That is the system in Scotland at the moment, and it works.
Keith Bartley: When that kind of testing, which is about confirming teachers' assessments, and those assessments, which are more and more being built around assessment for learning practices, become more mainstream, we get very much back to the kind of situation that was described earlier. Testing was a regular part of my primary and secondary schooling, on a daily and weekly basis. It is about putting those tests back in a functional, useful way into schools. The level of resource would be different. The amount of money that is spent nationally on the external administration and validation of our current testing system nowhere near justifies some of the benefits and disbenefits that it generates.
Chairman: I see nodding heads from your colleagues, so I will not go to them, if that is all right. I will invite Annette to ask the last group of questions.

Q169 Annette Brooke: I shall be very brief. You mentioned bringing back passion into teaching. Is it impossible to do that within the present system? Is it inevitable that there will be teaching to the test and narrowing on the national curriculum unless we scrap the current system?
Mick Brookes: Yes, I believe that that is absolutely true. I do not want to overstate the case, but a system of fear has been inculcated throughout education, particularly if you are a young head, or a young deputy, with a young family and a mortgage. You do not want to go to headship because you know that you will carry the can. Unfair and unfounded decisions are made on the performance of schools because Ofsted is now relying far too heavily on the data that we have discredited during this presentation. It is having a profound effect not only on the curriculum, but on the recruitment and retention of head teachers, in particular, who carry this can and do not survive being put into a category, on many occasions quite unfairly.

Q170 Annette Brooke: Can I follow up on the personalised learning test? Personalised learning seems to equate with goodness, but under the current system will it just be booster classes and personal intervention plans? Will it be centrally directed, almost?
Dr Bousted: It is likely to be highly bureaucratic and it should be very simple. The issue about personalised learning should be at the heart of good teaching. What is it that I know a child can do? What is it with which they need my help, or the help of other pupils in the class—their more able peers? What help can I use from another pedagogue that will enable them to learn? The danger is that at the

moment something that should be right at the heart of teachers' instinctive professional practice is being formalised into a whole range of other structures.

Mick Brookes: The other problem is that we need to move from norm-referencing to criterion-referencing pupils' progress, and that makes it individual. It is not necessarily how Sally compares with Fred, but the progress that Sally has made, and has made despite the fact that she might have quite severe special educational needs.

Q171 Annette Brooke: Could that be introduced alongside the present system, or do we really need to scrap that totally? Obviously, it is of great importance to value every child's achievement, which we are not doing at the moment.

Mick Brookes: Our clear view is that the current system is not helpful to children or to the curriculum, and it is certainly not helpful to my colleagues in schools.

Annette Brooke: May I just run through—

Chairman: Keith is desperate to answer these questions, and I want to give him a chance to do so.

Annette Brooke: I am sorry.

Keith Bartley: I want to come back to your original question. You asked whether it was impossible. I do not think that it is impossible; I just think that teachers have to be absolutely exceptional to be able to flourish in our current system. There can be an effect on motivation and retention. However, equally, I want to be clear that we see, through our teacher learning academy, some amazingly innovative examples of teachers innovating in their own classrooms and feeling that they have permission to do so. That is not impossible; it is just very difficult.

Q172 Annette Brooke: Finally, throughout this sitting, we have had the impression that the current system is demotivating for children, teachers and, as we have just heard, head teachers. Could each person give me what to them is an important factor around the demotivation of children and teachers—just one?

Dr Bousted: For children it is if they do not get their Level 4. That is hugely demotivating if you are going into secondary school. You feel yourself a failure. It is interesting to note that that is particularly a problem for boys. Boys react very badly, gentlemen. It happens to you throughout life—you react very badly to failure.

Adam Afriyie: How dare you say that? *[Laughter.]*

Mick Brookes: I would say it is the idea that we will continue having meetings until morale improves. If we continue doing things the way we are doing, we will not improve the morale of our children or teaching staff in schools, so we have to do things differently. We have to broaden the curriculum and give children those experiences of delight in other areas. We can teach English and literacy through other curriculum areas. Schools need to be encouraged and empowered to do so.

Brian Lightman: I think that the biggest thing is the fear of failure, coupled with not knowing what to do about it. We know that there are things that students will find difficult—that is part of learning. It is an important part of the learning process that we get things wrong and have to correct them, but it is not being able to do anything about it that causes problems. Students think, "Oh dear, I am going to take this test and I am going to fail and be labelled a failure," at the age of 11, 14 or whatever, and then we wonder why we have problems with motivation.

Keith Bartley: We need to give our schools and our teachers permission to innovate and permission to fail. They need to be confident about giving children permission to explore by learning.

Chairman: As we have overrun our time, that was a good line to end on. You have all had your opportunity to summarise what you feel about the matter. I am sorry that I did not manage the time as well as I had hoped. Next time, if I am ever in this Chair again, I will. Thank you very much for your time.

Monday 21 January 2008

Members present:

Mr. Barry Sheerman, in the Chair

Annette Brooke	Mrs. Sharon Hodgson
Ms Dawn Butler	Fiona Mactaggart
Mr. David Chaytor	Mr. Graham Stuart
Mr. John Heppell	Lynda Waltho

Memorandum submitted by The Assessment and Qualifications Alliance (AQA)

SUMMARY

This memorandum is AQA's response to the invitation from the Select Committee to submit evidence in connection with the Select Committee's inquiry into Testing and Assessment. The memorandum offers background information on AQA and then focuses upon those of the questions posed by the Select Committee in its published terms of reference on which AQA is particularly well-placed to express a view. In doing this, issues raised by some of the other questions posed by the Committee are also addressed.

The essence of this submission is an argument for choice in Key Stage 4 and post-16 qualifications offered by competing Awarding Bodies. There should be choice between modular and non-modular examinations; choice as to the use of properly controlled teacher-assessed coursework; choice between GCSEs, A-levels, Diplomas and Baccalaureates. Effective regulation is required in such a system to maintain public confidence and ensure that young people are treated fairly. However, there are no insoluble technical problems to achieving this aim.

As in other areas, competition between Awarding Bodies drives both technical and educational innovation and helps to reduce the burdens and costs of the assessment system as a whole. Again, there are no insoluble technical problems about ensuring comparability of standards in such a system and no evidence of any consistent problem in this respect at present.

Most importantly, choice in qualifications is consistent with the widespread desire to move learning in a more personalised direction and what we know about the varying ways in which different young people learn and are best motivated.

THE ASSESSMENT AND QUALIFICATIONS ALLIANCE

1. AQA is the UK's leading Awarding Body and, as a long-standing provider of high quality general qualifications at GCSE and A-level, the awarding body of choice for schools. We are a non-profit making educational charity so all our income from examination fees goes into running and developing our examinations and other services to schools and colleges. We place great emphasis on engagement with our stakeholders in educational centres to ensure we are fully meeting their needs. As the UK's main Awarding Body for general qualifications, one of our primary roles is to engage with our regulators and policymakers on issues of curriculum design and wider educational and assessment policy, utilising our educational research department which has a considerable international reputation. One of our priorities is the effective use of innovative technology to facilitate and modernise assessment techniques. AQA is pioneering the introduction of new methods of electronic assessment and marking that increase accuracy and reliability while maintaining and enhancing the integrity of the examination system.

GENERAL ISSUES

Why do we have a centrally run system of testing and assessment?

2. The National Curriculum tests at Key Stages 2 and 3 are centrally run by QCA but at Key Stage 4, and thereafter, competing Awarding Bodies provide a choice of assessments and qualifications. Including AQA, there are three General Qualification Awarding Bodies in England, one in Wales and one in Northern Ireland. All five of them are able to offer qualifications in England, Wales or Northern Ireland.

3. In the broader sense of assessment in support of the teaching and learning process, we do not have a centrally run system, of course, and there is much good practice carried out by teachers in their own classrooms, some involving material or tests purchased from publishers and other suppliers, some constructed by themselves. Assessment by teachers is the essence of good teaching which requires constant evaluation of the specific strengths and weaknesses of each pupil and planning of the best way to develop their knowledge and understanding further. More formal assessment for learning has a significant part to play and can build upon and illuminate the teacher's informal knowledge of their pupils.

4. Testing can act as a motivator for students and teachers but, in general, it can only improve attainment directly if the outcomes of the tests are used to guide subsequent teaching and learning. To be effective in this regard, assessment instruments need to be designed primarily with this purpose in mind and to provide more than a general snapshot of achievement. Although it is possible to extract some formative information from summative assessments like the National Curriculum tests by looking in detail at the pattern of an individual pupil's performance on the different questions in the test, their use for this purpose is inevitably limited.

5. However, the primary use of the National Curriculum tests is accountability for schools and monitoring of the performance of the education system as a whole. They have a second main purpose of providing standardised information for parents and others about the progress of individual pupils. It is sensible for them to be centrally run for both these purposes. Although an arrangement in which there is local control of progress assessment (say, at Local Authority level) with some system for ensuring that the results can be aggregated to form a coherent national picture for accountability purposes are possible, they are likely to be more expensive and burdensome than a single national system because of the need to establish comparability of standards across many different local assessments.

6. At Key Stage 4 and beyond, the need for accountability measures is met by collating outcomes from the qualifications which the Awarding Bodies already provide for individual students. This ensures that no additional assessment burden is placed upon schools and students for accountability reasons. The primary purpose of the work of the Awarding Bodies is certification of individual achievement. The credibility of the qualifications is enhanced because much of the assessment is external to the school or college and centrally regulated, albeit not centrally administered.

7. The Awarding Bodies operate in competition with each other and, as in other industries, competition encourages them to be innovative and to operate to high standards of service and offer high quality examinations. Modernization and innovation in assessment is one of the benefits of competition. For example, in large scale GCSE and A-level examinations, AQA is leading the way in the use of electronic assessment, where candidates answer on computers rather than on paper. A great deal of development work is also underway to improve the quality, security and control of the marking process by the use of modern technology. Quality of marking is clearly an issue of great concern to the schools which are our customers and it is unlikely to be a coincidence that these important developments are progressing much faster for qualifications where there are competing Awarding Bodies than for National Curriculum Tests where there is no competitive pressure.

8. Even in an era when the content of many examination requirements is largely determined by the regulator QCA, there is scope for significant curriculum innovation which is encouraged by competition between the Awarding Bodies. For example AQA's recently developed Foundation Certificate in Secondary Education in Modern Languages was welcomed by Lord Dearing in his recent report on Language teaching and is intended to help reverse the decline in the take up of Modern Languages in secondary schools. We are also developing an AQA baccalaureate qualification which will enable A-level students to have the breadth of their educational experience, including enrichment activities such as work experience or significant contributions in the community or to their school or college, to be formally certificated alongside their A-level results.

9. There is significant scope for competition in the level of service and support for teaching and learning that an Awarding Body provides. AQA offers extensive programmes of teacher support at the beginning of a new syllabus and regular review opportunities as the syllabus becomes operational. High quality and rapid feedback comes from examiner reports after each examination and we ensure that comprehensive, innovative and motivating teaching and learning materials are ready for use as soon as a school has selected AQA as its examination provider. The provision of teaching and learning support is a major, and growing, part of the work which, as an educational charity, AQA does and is one of the things which differentiates us from competitors.

10. Some critics believe that the existence of several Awarding Bodies is bound to lower standards because they will compete by awarding more high grades. The evidence does not support this. There is little correlation between market share and pass rates and thus little incentive to compete on standards but, in any case, the General Qualification Awarding Bodies all see their role as providing educational services for the benefit of young people. This, and the credibility of the sector as a whole, clearly depends upon comparability of standards between the Awarding Bodies and we therefore work together on research to ensure that our standards remain comparable, publishing the reports of that work for public scrutiny. As our regulator, QCA also does research in this area and has found no evidence of significant and consistent differences of standard between the GCSE and A-level awarding bodies.

Does a focus on national testing and assessment reduce the scope of creativity in the curriculum?

11. The crucial issue here is the status given to the results of testing and assessment and the extent to which they become the primary focus of teaching and learning. There is a clear tension between the need to ensure that schools are properly accountable on the one hand and the need to allow room for curriculum innovation and inspired teaching on the other.

12. There is no reason why a suitably structured and restricted range of national testing and assessment should swamp the curriculum and constrain creativity. Clarity about the purpose of national testing and assessment is crucial to its design. If the intention was only to measure the performance of schools, then there would be no need to assess every pupil across the entire curriculum at each key stage—a sampling approach within each school would be adequate. In the same way, if the intention was only to monitor the system as a whole then sampling of schools themselves would be possible within a system which would still provide effective national accountability. If, on the other hand, the intention is to provide information about the progress of individual pupils, then a more comprehensive assessment regime, like the present one, is necessary.

13. In this connection, we welcome the work which QCA is proposing on the construction of diagnostic National Curriculum assessment material for use by teachers wishing to assess their pupil's progress when they are ready for it. Once established, these assessments could replace the use of Key Stage tests for individual purposes, allowing those tests to be scaled down, reducing the burden of testing for accountability purposes upon pupils and improving the cost-effectiveness of the accountability system.

14. Underpinning the testing and assessment regime is, of course, the National Curriculum itself. The latest revisions to the programmes of study (in Science, for example) appear to be taking an approach based more on key concepts and requirements, rather than specifying all that must be taught. This is to be welcomed as widening the scope for creativity in the classroom.

Who is the QCA accountable to and is this accountability effective?

15. QCA operates to remits from the Secretary of State in relation to testing, assessment, qualifications development and operation. From its behaviours, it appears that QCA sees its main accountability as lying there.

16. There is therefore considerable potential for conflict of interest between QCA's twin roles of developer and regulator. The clearest example is in National Curriculum Tests which QCA simultaneously operates (through its NAA division) and regulates. However, there are many other instances. For example, QCA produces the current Basic and Key Skills tests and requires the Awarding Bodies to use them, rather than develop their own. As a result, regulation is weak and there are serious problems about accountability for the quality of the test material and assessment processes used.

17. However, the most important problem arising from the lack of clarity about QCA's accountability is encapsulated by its recent decision to regulate the entry fees for new A-levels and Diplomas. This is within QCA's statutory powers and entry fees are clearly a legitimate area of engagement for a qualifications regulator. But there was no visible attention paid to operational efficiency or cost-effectiveness during QCA's work on developing the structures and operating rules for these qualifications, and nor is there reference to these matters in the criteria for their accreditation. There is a need for clear separation between the qualifications regulator—with real independence from Government—and the agency working to government remits to develop new families of qualifications in furtherance of policy. Such a separation could make a major contribution to the proper consideration of the impacts, both educational and financial, of new qualifications that are being considered and developed as part of national policy. At present these important considerations, which have major implications for the administrative burdens on schools and significant impacts upon school budgets, are given little or no priority until the design phase for new qualifications is over—a highly unsatisfactory, if long-standing, state of affairs.

What role should exam boards have in testing and assessment?

18. Exam boards (Awarding Bodies) are concerned primarily with the provision of qualifications to enable individual students to progress to subsequent stages of education or employment. The use of qualifications for this purpose has the effect of ensuring that, as far as possible, those who engage in further study or particular jobs are equipped to do so. In this way, qualifications contribute to greater efficiency in the employment of the nation's human capital than the alternatives (essentially influence, patronage and random selection). Equally important, qualifications contribute to social cohesion by providing a widely accepted and essentially meritocratic basis for sharing educational and vocational resources.

19. As a result, the Awarding Bodies are the organisations in the UK with the greatest knowledge and most practical experience of educational assessment. AQA, for example, prepares assessment materials and organises examinations on a national scale every year (we set, mark and process approximately 13 million individual student assessments every summer). AQA also has the largest research facility of any of the exam boards in the UK, making our Research Division the largest group of professional researchers anywhere in the UK who are working solely on educational assessment.

20. With this background, it would be sensible for Awarding Bodies like AQA to play a central role in National Curriculum testing and assessment within the UK. However, there is little incentive for an organisation like AQA to involve itself in current National Curriculum testing because the educational work is inevitably tightly specified, leaving no scope for curriculum development and, historically, development of the assessment processes is marked by an aversion to innovation which makes the use of modern

technology and the delivery of cost-efficiencies difficult. The result is that contracts to operate National Curriculum tests are essentially about delivery of routine administration on behalf of QCA, rather than educational or assessment development work.

21. This may be about to change to some extent with the proposed development by QCA of National Curriculum assessments which pupils will take when they are ready, with the intention of supporting personalised learning backed by diagnostic assessment in classrooms. It will be crucial for this work to be approached as an opportunity to develop National Curriculum assessments which are innovative in terms of assessment processes if the expertise of the Awarding Bodies is to be engaged, as it should be, in these important national developments.

Testing and Assessment at 16 and After

Is the testing and assessment in "summative" tests (for example, GCSE, AS, A2) fit for purpose?

22. The purpose of GCSEs and A-levels is to identify within reasonable parameters of confidence, and certificate, a student's performance in a particular subject at a particular time. The assessment arrangements are fit for this purpose and based upon educationally valuable specifications which are only accredited following rigorous consideration by QCA. Specifications have to meet national qualification and subject criteria and assessment processes have to comply with a statutory Code of Practice. As a result, there is no evidence that standards of demand and reward are not broadly consistent across Awarding Bodies and over time so the confidence that end users of the qualifications have in them is not misplaced.

23. The standards of attainment which GCSE and A-levels represent are widely recognised, understood and valued by Further and Higher Education and employers. The qualifications are highly valued by young people and their parents.

Are the changes to GCSE coursework due to come into effect in 2009 reasonable? What alternative forms of assessment might be used?

24. In 2006, examination coursework in GCSE and A-level examinations was the subject of an important report from QCA. The report confirmed the value of coursework in many subjects but the issues of candidates receiving assistance with their coursework and of plagiarism, especially involving the use of the Internet, were highlighted. QCA outlined a number of initiatives to address these issues, such as improving the understanding of what is acceptable in terms of assistance. AQA is fully committed to these initiatives and has been actively working on them with QCA and other awarding bodies.

25. QCA has also reduced considerably the amount of coursework in examination requirements, particularly for the new GCSEs. There has also been a reduction in the amount of coursework involved in A-level for courses starting in 2008, as part of the change in the number of assessment units in most subjects from 6 to 4. Is the reduction in the amount of coursework in GCSE and A-level examinations the best policy to pursue? To answer this question, we need to understand how we have arrived at the current position.

26. Although coursework is the term that everybody uses, teacher-assessment might be a better one. Most GCSE and A-level coursework components consist of a well-defined piece (or pieces) of work which students complete during their course and which is then marked by their own teacher. Rigorous external moderation procedures are applied by the Awarding Bodies after the assessments have been made, in which external examiners re-mark a sample of work from each school to ensure that every teacher's marking is done to the same standard.

27. However, examination coursework was originally intended to be work carried out during the course, not an additional examination requirement such as an extended essay or project. For example, the early requirements for examination coursework in GCSE English Literature were for assessments by teachers of pieces of work produced during the course of study across a range of genres, periods, and so on—work which arose naturally as part of the study of literature over two years. This approach leads to a wide range of work being produced and assessed and requires significant professional participation in standardising and moderating the work. Consequently, moderation of this sort of coursework focussed upon training and professional development in meetings of teachers organised by Awarding Bodies, as well as on external checking after the event.

28. Between the 1980s and the present day, coursework changed in its nature and the perception of it changed equally. Concerns expressed, but never justified with substantial evidence, about the extent to which the original approach involved trusting in the professionalism of teachers led policy makers to seek increasing amounts of control over the nature of the work assessed and direct moderation of the marks awarded. The consequence is the situation we now have where a more formulaic and controlled approach leads to less motivation for students and more of a sense of burden for teachers. The very tight definition of the coursework which candidates have to do facilitates plagiarism and other practices which now cause concern.

29. In essence, the historical attempt to reduce risks relating to teacher professionalism by increasing amounts of control, has created different risks relating to the authentication of coursework as the work of the students themselves. The question is whether it is now possible to move back towards a system in which examination coursework arises more naturally as part of the student's learning and is assessed by teachers working under a framework of quality assurance, as well as quality control.

30. Ironically, at the very time when QCA is reducing coursework in many GCSE and A-level examinations, the new Diplomas will involve subjects and units which, by their very nature, require assessment by teachers of vocationally related activities, many of which involve hands-on practical work. The resources required to externally examine these activities would be prohibitive and they do not lend themselves to re-marking after the event by an external moderator. Similarly, Applied GCSEs are already fully operational and feature a pattern of two internally assessed portfolio units and one external assessment. Teaching for Applied GCE A-levels started in September 2005 and this summer (2007) will see students being awarded the first full Applied A-level results. These are a key part of the strategy to broaden the range of learning available to 14 to 19 year olds and, in general, consist of 2/3 coursework and 1/3 external assessment. They involve an approach which puts major emphasis on the accreditation of individual teachers, who retain their accreditation even if they move centre, subject to light moderation.

31. With appropriate structures of support and moderation, teacher assessment can provide valid and reliable results within an environment that is encouraging to students, rather than daunting. The concerns that have been voiced about possible plagiarism through the use of the Internet and malpractice through the input of others (fellow students, parents, and so on) are valid, but appropriate action can be taken to ensure fairness and that appropriate grades are awarded. Work which arises naturally during the course is much more varied, enabling teachers more easily to detect plagiarism, especially if the work is done, at least in part, in the classroom under controlled conditions. Discussions over what represent adequately controlled conditions are currently taking place to define criteria for teacher-assessed components and AQA will put those outcomes into action. It is possible to set assessment pieces that are of a task-like nature but which can be completed in the classroom, under direct supervision and within a restricted timeframe. In these sorts of ways, plagiarism can be made much more difficult to do and easier to detect so that, if it does occur, it can be identified and penalised.

32. In the world outside education it is relatively unusual for anyone to sit down and think and write continuously without the opportunity to explore additional resources or check points with colleagues. Such opportunities reinforce the value of collaborative and team building skills. Many modern coursework requirements are artificial to the extent that help from colleagues—in this case teachers and parents—is artificially restricted. If we are preparing our young people for the world beyond school, then activities which require elements of independent study and research but also sharing and cross referencing with others must be appropriate.

33. These arguments show that alternatives to reducing the amount of coursework in examinations are possible and consistent with current policy developments. They also reflect the fact that, on a large scale, many practical skills cannot be assessed externally in a valid and cost-effective way. As is implicitly acknowledged in the new Applied GCSEs, Applied GCEs and Diplomas, some assessment is best done by teachers—a further example would be oral skills. QCA has therefore tried to identify those academic subjects which require coursework assessment and has prohibited its use in others. This is a restrictive and narrow approach which unnecessarily limits the choice available to teachers and students. If teacher assessment can provide good quality assessment when it is essential, there is little logic in prohibiting it as an option in any subject.

What are the benefits of exams and coursework? How should they work together? What should the balance between them be?

34. Balance is everything here. Exams ensure a level playing field, but can distort teaching and learning. Coursework, where it is well designed and well implemented, supports learning but, as noted above, makes it more difficult to ensure a level playing field. Each form of assessment can assess different aspects of knowledge, skills and understanding and they work well together where they do not assess the same things but are used in a complementary way.

35. There is no single answer to the question about balance. One balance is not appropriate for all subjects. A judgement must be made as to which assessment strategies best fit the particular subject and course specification and which strategies will provide the most valid and robust assessment of the student's abilities in each area of learning involved. By this means it is possible to produce an assessment scheme that is supportive of good teaching and facilitates learning. This is the aspiration which should determine the choice of balance between exams and coursework, or any other form of assessment.

Will the ways in which the new 14–19 diplomas are to be assessed impact on other qualifications, such as GCSE?

36. The multi-component nature of the Diplomas will impose pressures on students that are different from those which arise from the separate demands of a number of GCSE or A-level subjects. This will be particularly true at Level 2 where there is strong evidence that the requirement to pass the functional skills at that level before a Diploma can be awarded will be a severe challenge for many students. If a student is finding it difficult to maintain the requisite level of success across the range of demands of the Diploma it is likely that they will seek to reduce the pressures from elsewhere to enable them to focus on their Diploma work. However, this may well be a desirable trade-off if it replaces poor results in a range of GCSEs with success in a coherent Diploma course. The number of GCSEs taken by many young people is, in any case, an issue which is worthy of review and pressures to reduce it somewhat are not self-evidently disadvantageous, given that there is a requirement for the functional skills and other broader activities alongside the vocational cores of the new Diplomas.

37. There is, however, a related matter of major concern to AQA which arises from the proposed relationship between Functional Skills and GCSE qualifications in English Mathematics and ICT. For GCSE courses in these subjects starting in 2010 it is presently policy that students will be required to achieve Level 2 in the relevant Functional Skill to be eligible for the award of a Grade C. Our research into the potential impact of this 'hurdle' suggests a major risk that there will be a significant consequential reduction in the number of students achieving success at Grade C or better in these GCSEs, particularly in English.

38. Of course, it could be that the situation will be different in practice when the new GCSEs in English, Mathematics and ICT are first certificated in 2012. Levels of achievement in the functional skills will, hopefully, have improved significantly as a result of a strong emphasis on their teaching and learning in the intervening years. This is clearly the intention of the policy involved. However, the policy will require significant investment and careful monitoring of its success in order to ensure that GCSE standards can be maintained when the functional skills "hurdles" begin to operate in 2012. The consequences of a major change in GCSE standards and outcomes in English and Mathematics in that year would be problematic because of the compulsory inclusion of these subjects within school performance measures at Key Stage 4. More important, however, is the potential for very grave injustice to be done to the young people affected, as they compete with those from the year before for the same jobs and places in Further and Higher Education.

Is holding formal summative tests at ages 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?

39. The burden is considerable at age 16 because of the range of subjects commonly taken. Where students are clearly going on to post-16 study, what purpose do 8–10 GCSE grades serve? Where students are clear which subjects they want to take through to post-16 education, it is entirely possible to have a system where they do not take the 'summative' GCSE examination in those subjects at 16. This was an aspect of the old O-level and A-level examinations which operated successfully in many schools. Clearly, in the current world where there is more mobility at age 16 there is a greater need for certification of achievement at 16 but for many students a record of subjects studied would serve the purpose adequately in subjects which they plan to study further—certainly when that further study is to be at the same school.

40. Similarly, there is little point in AS certification (and thus examination) when a student plans to progress to A2. If AS and A2 were decoupled, so that AS was no longer formally part of A-level examinations, such students could bypass assessment at AS and proceed straight to A-level. For the rest, AS could continue to act as a useful qualification, certificating their work in the first post-16 year and enabling them to choose their full A-levels informed by their progress so far. Such a decoupling arrangement would need the support of Higher Education as it would impact upon admissions arrangements but it would offer the opportunity for the A-level assessment to be more holistic in nature which is something for which many HE institutions have expressed support.

To what extent is frequent, modular assessment altering both the scope of teaching and the style of teaching?

41. Modular assessment arrangements can give rise to a perception that the subject as a whole is not addressed as it would be with a wholly end of course assessment. If the modular assessment structure is accompanied by a modular approach to teaching and learning, then opportunities to explore broader issues arising from the studies are reduced. In most subjects, particularly at A-level, these sorts of issues require a more comprehensive understanding of different aspects of study. It takes time to accumulate the knowledge and skills required to address real issues rather than ones that have been contrived to fit within the confines of the limited areas of study currently in focus. In order to attempt to address such concerns a number of artifices have been added to A-level, such as synoptic assessment and stretch and challenge questions, but with varying degrees of success.

42. On the other hand, modular examinations and courses enable greater flexibility of provision and can improve achievement through better structuring of courses, increased accessibility of course content to students, valuable feedback about progress and continuing motivation. They also help to ensure that

students who do not successfully complete entire courses nonetheless receive proper recognition and certification of what they have achieved. Some institutions—particularly in the FE sector—use module results as part of their accountability arrangements for individual teachers.

43. It is for these sorts of reasons that many Higher Education courses of study are now modular in nature, despite precisely the same implications for holistic understanding of the subject being an issue at that level. It is also important to note that unitisation (modules) is a design feature of the new Diplomas and vocational qualifications generally.

44. The right response to this sort of situation in which there are clear strengths and weaknesses in a particular assessment model is to provide choice for schools, colleges and therefore students. One student's motivation from regular module tests is another student's stressful burden and a choice should be provided, rather than a single model to fit all. This was the case prior to the introduction of the Curriculum 2000 A-levels and could be so again if QCA were to develop appropriately flexible criteria for A-levels.

How does the national assessment system interact with university entrance? What does it mean for a national system of testing and assessment that universities are setting entrance tests as individual institutions?

45. University resources are not unlimited so we have to choose which of our young people we are going to make them available to. Putting this selection process on a sound basis is the key purpose of A-level examinations but there is currently a significant problem in using A-levels for selection to some university courses: a minority of University selectors find themselves faced with more applicants than they can accommodate who have straight A grades from their A-levels.

46. Does this mean that A-levels are too easy? The average Grade A pass rate for A-levels in 2006 was about 24%—meaning that almost a quarter of the candidates for a typical A-level got a grade A. There are five grades altogether, so having a quarter of candidates in the top grade means that the grade scale is not providing as much differentiation between the candidates as it could. This is the result of maintaining the standard represented by the grades over a period when young people's achievements improve significantly. Judged educationally, it is cause for celebration. But judged from the standpoint of the effectiveness of A-level examinations for university selection purposes it is a problem.

47. On the other hand, if we look at the complete results of individual candidates, most of whom take several A-levels, we do not see a general problem of differentiation:

	4 As	*3 As*	*2 As*	*1 A*
Cumulative % of candidates	2.5%	9.2%	17.2%	31.9%
Cumulative % of 18 year-olds	0.8%	3.1%	5.8%	10.8%

48. Only just over 9% of candidates get 3 or more Grade As—and that is just 3% of all the 18 year-olds in the country in any one year because only about a third of 18 year-olds take any A-levels at all. Only about 11% of our 18 year-olds get one or more Grade As at A-level. From these figures, it is very hard to argue a case for making it harder to get a Grade A at A-level, but there is a case for introducing a Grade A* to differentiate within the top 25% of candidates in each subject and to ensure that A-levels remain fit for their purpose of providing the basis for fair selection for university courses of all types.

49. AQA is therefore strongly supportive of the decision to introduce a new A* grade at A-level. There are to be revised A-level courses starting in 2008 which include some more demanding and open-ended questions and the present policy is to bring in the new A* when the first awards are made for those new A-levels in 2010. However, there is enough information in the marks of candidates taking current A-levels to enable us to differentiate effectively between candidates within the present Grade A and we believe that we could, and should be permitted to, introduce the new A* two years earlier than that, in 2008. We would urge that this be given serious consideration.

50. Of course, there is detail to be worked through about the precise mechanism for awarding Grade A* but AQA has already done a substantial amount of research on this topic. There are other important matters to be decided as well—not least the question of how many UCAS points the new A* grade will be given. However, none of these technical matters is difficult or complex enough to prevent A* grades being awarded and used in university selection from Summer 2008.

51. There are two issues of concern about the introduction of an A* grade. One is the increased pressure which it will exert on young people. Some of those who currently aspire to achieve straight Grade As will now aspire to straight Grade A*s. For some, that will be an appropriate aspiration—and it is not the business of educators to persuade their students out of having high expectations of themselves—but for others it will not be an appropriate target and it will be vital for teachers to provide wise counsel to students about how best to spend their time and energy during their A-level studies. It will be essential for everyone to understand that the standard of Grade A has not changed—that a Grade A at A-level remains an excellent result which only 11% of our 18 year-olds achieve.

52. The other concern which has sometimes been expressed about the introduction of an A* grade at A-level relates to widening participation and, particularly, the issue of inclusiveness. In particular, independent school students are over-represented among those awarded a grade A, given the proportion of all candidates

educated in independent schools. Independent school students make up 14% of all A-level candidates but 28% of those with grade A. Will they predominate even more in the new Grade A*? Preliminary research carried out by AQA across a range of subjects, including Art, Sciences, English Literature, Foreign Languages and Social Sciences suggests that the proportion of independent school pupils in Grade A* is likely to be a little higher than it is for Grade A, at around 33%, with 18% coming from FE and 6th form colleges and 47% of candidates with A* grades coming from maintained schools. There is a challenge here for the maintained sector as a whole but these are only average figures. The reality is that there is a range of achievement in all school types, with many maintained schools and colleges more than matching their independent colleagues in terms of examination results. It can only be good if the introduction of a new Grade A* at A-level serves as a spur to further improvement in schools of all types.

53. Certainly, the introduction of specific university entrance tests is a backward step and not conducive to an education system that seeks to give the same opportunities to all. Such tests bring significant risks of curriculum distortion and problems of social inclusion as a result of differential availability and level of preparation for candidates from different backgrounds and in different types of schools. A national assessment system which meets the selection requirements of all universities is a much more equitable approach and is the pattern in use generally around the world.

54. One of the ways of selecting people for University which is sometimes suggested as being better than A-levels, precisely from a social inclusion perspective, is the use of aptitude, or reasoning, tests. Those who propose aptitude—or reasoning—tests for university selection usually make two, related claims:

— They claim that assessing aptitude or reasoning ability, rather than class-room learning, removes the effects of schooling. So they claim it is a fairer way of selecting which will help to widen participation and promote social inclusion.

— They also often claim that it is better because it provides better predictions of success at university.

55. Neither of these claims is consistent with the evidence. In the USA aptitude tests have been used to select students for university for years, with the consequence that a great deal of time is spent practising for the tests. And in this country when selection at 11 was common primary school children spent many hours practising for their 11 plus. The fact is that scores can be significantly improved on any sort of test by practice and preparation. All the evidence is that teaching improves aptitude test scores—they are not "school proof" and are not, therefore, inherently better than examinations in terms of ensuring that selection is fair or socially inclusive.

56. Nor are aptitude or reasoning tests any better predictors of success in university than examinations. In fact, in this country, studies have shown A-level to be the best single predictor of success at university, albeit not a very good one. Around the world, there are countries which use tests of various kinds, countries which use examinations and plenty of combined approaches. Overwhelmingly, the evidence is that none of these selection methods provides very reliable predictions of students' performance at university. And nor is it surprising that it is difficult to predict the future 3 years hence—a future which involves a quite different approach to education and some of the most significant developments in many students' personal lives.

57. So aptitude tests are neither fairer nor better predictors of success than exams like A-levels. But their downside is that time spent practising and preparing for them could be spent on learning which has real benefit. Preparing for exams like A-levels which are embedded in the curriculum is about learning which has purpose. The use of tests which supposedly assess reasoning ability or aptitude, independent of schooling, can—and usually does—distort the curriculum significantly.

58. But if it is to be examinations, rather than aptitude tests, which we use to select young people for university, would baccalaureate style examinations be better than A-levels? Baccalaureates compel each student to follow a broad range of study. This may mean including mathematics, their mother tongue, a foreign language, perhaps something about the theory of knowledge in their course—as well as the subjects in which they are specialising. Choice for students is therefore constrained. Such constraints are not self-evidently desirable, especially at a time when individualised learning is seen as an important tool in the drive to encourage more young people to stay in education after the age of 16, and to go to University. For example, when there is concern about the decline in the number of people studying physical sciences, we would surely not want to prevent a talented scientist from going to University because they either cannot, or will not, study a modern foreign language.

59. A-levels, unlike baccalaureates, offer a wide choice for students in which broad study or specialisation is possible, depending upon their own interests and enthusiasms. And it is interest and enthusiasm which leads to learning which lasts, rather than learning which is done purely to get a qualification and which is then forgotten afterwards.

60. A-levels are also available in ones and twos. It is often not appreciated that 25% of our 18 year old A-level candidates take only 1 or 2 of them. This means that students who have a particular talent, but who would struggle to succeed in all the elements of a baccalaureate, can obtain valued qualifications which fully recognise their achievement. We should not deny those students that possibility, perhaps forcing them, instead, to take a lower level baccalaureate qualification which does not do justice to their particular talents in Art, Sport, Music or, indeed, in English Literature or Mathematics, if that is where their abilities lie. Such an approach would be inconsistent with the aim of developing each individual to the full extent of their capabilities, both for their own benefit, and that of society at large.

61. Baccalaureates should be offered as a choice for those whose educational needs are best met that way and AQA is currently developing its own baccalaureate qualification which will certificate student's A-level results alongside the completion of a personal in-depth study, learning about thinking or citizenship and their wider activities such as work experience, contributions in the local community or personal development programmes such as the Duke of Edinburgh's Award scheme.

IN CONCLUSION

AQA's Director General, Mike Cresswell, would be delighted to give oral evidence to the Select Committee if required and would welcome the opportunity for further discussion and clarification of any of the points raised in this memorandum.

As well as being Director General of AQA, Mike Cresswell is a Visiting Professor at London University's Institute of Education. He has worked on national and international surveys of students' mathematical attainment at the National Foundation for Educational Research and was an active researcher on assessment matters for many years before the formation of AQA for which he was Director of Examinations Administration before taking up his present post in 2003.

He has a national and international reputation as an expert on assessment—particularly on the topic of setting examination standards. He has published many papers, research reports and books on assessment. He served on Tomlinson's Assessment Group and continues to be closely involved in national discussions on the introduction of the new Diplomas, contributing especially to the recent work on methods of awarding candidates' grades.

Mike believes the major challenge of the next few years in the assessment field is the search for ways of bringing technology to bear on assessment to reduce costs and burden on schools and students in a way which both improves assessment quality and retains fitness for educational purpose.

June 2007

Memorandum submitted by City & Guilds

1. City & Guilds approaches the discussion from a vocational qualification standpoint.

2. Assessment should be seen as a rare event that put demands and responsibilities on both the designer and the learner.

3. The use of awarding bodies within the English/UK system is untypical in comparison with European countires. However, they provide a valued assurance of independence and professional expertise to the consumer.

4. Through a pair of professional associations awarding bodies are successfully reducing examination bureaucracy.

5. The twin requirements of validity and reliability should always govern the choice of assessment methods.

6. While formative and summative assessments have different purposes there is potentially valuable feedback available from both which can aid learning for all parties.

7. Coursework should not be abandoned in favour of examination. The design of coursework should be improved.

8. To reduce the examination burden between 16 and 18 teachers should be given a greater role in summative assessment of performance.

9. Using multiple approaches to assessment increases the reliability and accuracy of the assessment of the learner's knowledge and skills.

10. There is too little time for there to be the development of innovative approaches to assessment for the new aspects of the Diploma.

11. While employers should be encouraged to train to awards or units on the national qualifications framework, involving awarding bodies can ensure that bespoke qualifications are of a high standard and are portable.

12. Introduction

12.1 It must be made clear at the outset that City & Guilds is a vocational awarding body that has as its primary focus the assessment and certification of vocational knowledge and skills. Our market is generally 16+ and the average age of our candidature is around 30.

12.2 We have over 500 qualifications on offer and deliver to around 6500 centres in the UK. A centre can be anything from a FTSE 100 employer, to a College of Further Education, a Sixth Form college, a private training provider, to small employers. We issue around 1.5 million certificates a year. We believe that about 1 in 5 adults within the UK hold a City & Guilds certificate.

12.3 While our history would associate us strongly with traditional craft skills like agriculture and horticulture or construction and building services our broad portfolio of products reaches to retail, care, IT and ESOL (English for Speakers of Other Languages) and much beyond. The City & Guilds Group also includes the Institute of Leadership and Management offering awards in over 2,200 centres. Across the Group the range of awards extends from Entry Level to the equivalent of Level 8 of the QCA National Qualification Framework. Our portfolio is also delivered in about 100 countries internationally through some 3500 centres worldwide.

12.4 City & Guilds has been an examining body since 1878 and was awarded a Royal Charter in 1900. It has necessarily acquired considerable skills in curriculum and assessment design and delivery.

12.5 With regard to the interests of the Select Committee we seek to offer some general observations on the principles and purposes of testing and assessment, and on the role of awarding bodies.

13. General Issues

13.1 The UK stands out in comparison with both its local and more distant neighbours in that independent awarding bodies carry out the process of examining and certification for national awards. These bodies have become centres of considerable expertise in these processes and carry a major financial and moral burden of expectation in terms of accuracy and prompt delivery. This situation is a consequence of historical decisions and a presumed wish by past governments not to take on the administration and associated significant costs of running this national system.

13.2 The management of this system through regulation provides a reassuring degree of independence to the system that we believe is valued by the consumer. There is more accountability in that the awarding body's reputation depends upon their ability to deliver and market forces exert continued pressure on the system to ensure high levels of quality assurance. As opposed to being run by what could so easily be perceived as a bureaucratic government department.

13.3 Recent initiatives by awarding body consortia, the Joint Council for Qualifications (JCQ) and the Federation of Awarding Bodies (FAB), are making progress in the fight against bureaucracy. In particular the JCQ (comprising Edexcel, AQA, OCR, City and Guilds, WJEC, CCEA and SQA) supports an initiative it has called The Eight Pledges. These are principally eight areas where the awarding bodies will collaborate to reduce the complexity in the qualifications system by reducing the administrative burden caused by assessment and quality assurance with the purpose of simplifying the relationships within and between awarding bodies and centres. FAB is a consortium of over 30 awarding bodies concerned with vocational qualifications.

13.4 In the awarding body system we have collaboration and competition both have roles to play. In that much of the marketplace is buying national awards then collaboration is critically important to retain customer confidence and economy within the system. The opportunity for competition enables awarding bodies to fine-tune their approaches to satisfy a wide marketplace and provide specific and differentiating customer benefits.

13.5 The English regulator, QCA has contributed to this situation. On the one hand when discussing major initiatives it ensures wide representation by awarding bodies but its supervision of the awarding process remains specific to each awarding body. Where there has been confusion in recent times it has been when QCA has adopted roles already performed by awarding bodies. For example, in the development of curriculum for say GNVQ or more recently the management of question banks for Key Skills.

13.6 Awarding bodies are centres of expertise. The larger awarding bodies, in particular those in membership to the JCQ, carry specific technical expertise in curriculum design and assessment practice. The JCQ itself through its committee structure also promotes the further development of these skills and addresses the technical issues of assessment and standards setting for general qualifications.

14. Testing and Assessment at 16 and After

14.1 Assessment is about the collection and validation of specific evidence from or about a learner. Assessment in any of its forms is intrusive and for many unwelcome, un-looked for and unpleasant. Consequently, there is considerable obligation on the designer of tests or assessments to make them as efficient and meaningful as possible. Assessment opportunities should be seen as rare events during which the assessment tool must be finely tuned, accurate and incisive. To conduct a test that is inaccurate,

excessive, unreliable or inappropriate is unpardonable. Moreover, it is an insult to the hard work and anxiety of the learner to waste their time or be needlessly over demanding. Economy it time, effort and cost is imperative.

14.2 Assessment can be put to two principal purposes; namely formative and summative. These have different roles in the learning process and in the ultimate recognition of achievement. Put briefly, formative assessment provides feedback to the learning process by identifying both progress and gaps in learning. Used in this sense it is often regarded as diagnostic. Summative assessment is usually conducted at the end of a learning process and is focussed on assessing learning against a known standard for the purposes of certification.

14.3 There is a strong argument to suggest that all assessment should support or promote learning. This will depend upon the opportunity for feedback post assessment. Currently, this is the strongest division between the two types of assessment. Formative assessment generally operates at a local level and is built into a learning programme. The outcomes of assessment are not high stakes but provide staging posts to further or remedial learning. It is assumed that summative assessment only has value in terms of the final result (pass, merit, distinction or grades A-E etc). Given suitable analysis or interpretation much value can be extracted from a candidate's examination paper or final practical assessment. Considering the significant effort that goes into the final examining process by all parties the current under-use of this data is a travesty. Some awarding bodies are now developing analytical software associated with on-line access to enable some benefit to be drawn from this available data. This information can be of use to the awarding body, the examiner, the centre as well as the candidate.

14.4 For summative assessments like GCSE, AS or A2 enabling the outcome of assessment to support learning would make demands on both the examination structure and the marking process. However, the potential benefits for the candidate are significant. It would also exert additional rigour on the assessment process to achieve greater detail and accuracy. Some attempt has been made to compensate for the scant regard given final assessment through the provision of the opportunity for centres and learners to review examination scripts for general qualifications.

14.5 The discussion of examination and coursework may benefit from a brief consideration of two of the technical issues within assessment, namely validity and reliability. Put simply, the 'test' of validity seeks to confirm that the form of assessment used adequately reflects or accesses the skills to be measured. You will have a more appropriate assessment of the skills of baking by setting the task of baking rather than setting an essay question on how to bake. The "test" of reliability requires the assessment designer to show that assessment will repeatedly produce the same outcome, that there is no inherent bias or variability in the assessment instrument. Examination and Coursework should be regarded as two separate assessment instruments that reside in the assessment designer's toolbox which carry different degrees of validity and reliability.

14.6 In considering the balance between examination and coursework it may be interesting to reflect upon the current situation in vocational qualifications. In particular the National Vocational Qualification (NVQ) as regulated by QCA. The NVQ is a performance-based qualification underpinned by specific occupational standards. The assessment is almost totally locally supervised against assessment schedules prepared by awarding bodies in association with sector skills councils (SSCs) and accredited by QCA. Some NVQ do involve additional knowledge tests issued by the awarding body. The coursework is evidenced through a portfolio, a physical or electronic document that maps the learner's progress of performance/skills demonstration through the various units of the award. A locally based, occupationally competent assessor who has the opportunity to ask questions, challenge and reconfirm the performance carefully monitors and confirms the learner's achievements.

14.7 It is interesting to note that were workplace qualifications are concerned, where the country's economic performance is essential and of keen government interest the primary assessment decisions are made at local (supervisor) level. Whereas for school based qualifications which may lead to initial employment, further education and training or higher education an elaborate system of external examinations and near total independence from the local centre of learning is required. Many other countries invest considerable importance in the professional judgement of their trained teachers with regard to summative assessments of achievement.

14.8 It has been noted in some studies of the vocational education and training settings that a proportion of trainees experience difficulties with the development of their NVQ portfolio because of the tight examination based schooling they have received which has failed to inculcate the independence of thought and action needed within vocational education and training and employment.

14.9 The recent anxieties expressed over coursework and the opportunities for plagiarism are not unknown in vocational awards (VQs) though the types of incidence are different. The quality assurance system for VQs depends upon occasional visits to centres to essentially undertake an audit of local assessment practice. This process supported through regulation is called external verification. The process seeks to ensure that centres follow the required guidance and maintain the performance standards and criteria.

14.10 One particular issue to be considered with coursework is the nature of the task being demanded. If the task is one which can be easily downloaded from the Internet and passed off as the new learner's work then one can rightly suggest that the original task was poorly conceived or set out. Assessment techniques must of necessity move with the times. If new technologies make accepted assessment practices redundant then new ways of accessing the required skills must be devised. This takes time but the solution is not to rely upon one form of assessment, as this should be regarded as poor practice.

14.11 It is important to recognise that learners differ from each other and it is likely that acquiring curriculum content or specific skills demands a range of skills on behalf of the learner. Consequently using only one or at best two assessment techniques limits the type and value of the evidence one is collecting. There is a risk that the process will not do the learner justice. Over reliance on one or other forms of assessment cannot be regarded as good practice. However, it must be conceded that getting the appropriate weighting between assessment methods is also a difficult process. Dispensing with coursework is not the answer to plagiarism, as this would over-focus the teaching programme on the final examination to the diminution of those auxiliary skills the curriculum sought to develop. This also returns us to the question of the validity of the chosen assessment tool.

14.12 The weighting issue has been of particular concern in the development of the new Diploma, in that the final award has to be derived from performance in a range of specified elements. What has been more problematic has been the difficulty in securing clarity over the purpose of the qualification and its underpinning curriculum. The tension between the general and vocational themes will also be played out in terms of assessment regimes. The vocational trend would be for more emphasis on performance evidence (ie 'can do') whereas the general trend goes for knowledge-based evidence (ie 'knows and understands that'). The vocational aim for the Diploma was to get learners out of the classroom or at least to make learning relevant beyond the classroom. It is still too early to say how the Diploma will work out in practice. A full and energetic evaluation of the pilot programmes is essential.

14.13 It is difficult to say whether or not Diploma assessment methods will have a consequential effect on GCSE. This will in part depend upon the innovation awarding bodies are able to bring to the assessment of the new elements in the Diploma. Given current timescales there has been insufficient time for Component Awarding Bodies to research and develop new assessment approaches. It can easily be argued that for a new qualification which is to be the standard bearer for a subtle blend of general and vocational skills that new and innovative approaches would be needed in both curriculum and assessment delivery. Accepting some of the earlier points in this submission some of this innovation should pay attention to improving feedback after assessment following the principles of assessment for learning.

14.14 There can be little doubt that the years 16,17 and 18 are a great challenge and pose a significant demand on all learners as they face an intense 3 year examination period. It is a shame that the regulated examining systems are unable to make greater use of the expert judgement of teachers and tutors and that increasing reliance is put upon single shot examinations. One of the benefits of the NVQ assessment system is the requirement to observe performance over time. Consequently random poor performances can be weighed against evidence of more consistent performance. Repeated demonstrations of good performance are also required rather than a single inspired one.

14.15 It must be recognised that the general qualifications process remains a highly competitive one and a learner's success depends heavily upon the school or college they attend and the resources that school or college is able to secure. In this sense it is not a 'fair' system. It will not be 'fair' until all learners are able to access equivalent resources delivered to a common standard. While the same criticism can be made of vocational qualifications the system has embedded procedures to reduce the variability. All occupationally competent assessors must also hold a nationally approved assessors qualification. Every centre delivering an NVQ must meet regulated approval criteria. Each centre is regularly visited by a representative of the awarding body to ensure that the centre is maintaining the occupational competence standards in its assessment practices.

14.16 The same is true even where employers choose to deliver NVQ within their staff training programme. The benefit being that staff trainees will ultimately receive a nationally recognised qualification which will be truly portable rather than a training package locally conceived by the employer which may have no relevance to any other (future) employer. A training programme endorsed by the employer alone, except where that employer has achieved considerable brand credibility, will have only limited value.

14.17 Where awarding bodies collaborate with major employers to produce bespoke qualifications the staff trainees have the benefit of knowing that they will receive a properly validated qualification based on the experience, expertise and professionalism of the awarding body. The portability is based upon the recognised professionalism of the awarding body. It is also most likely that the awarding body will retain, in a lasting archive, full records of past achievements. City & Guilds, for example, goes back 100 years. Few commercial businesses last that long or would wish to retain records of long departed employees.

May 2007

Memorandum submitted by Edexcel

Introduction

This response is in two parts:

Part 1 describes who we are and what we do.

Part 2 provides responses to those specific questions posed by the Committee for which we have most direct experience. We have addressed issues regarding testing and assessment pre and post 16 together.

Part 1: Who we are and what we do

1.1 Edexcel is one of the largest awarding bodies in the UK and a Pearson company. It offers a wide range of academic and vocational qualifications, testing and assessment services and associated products and support aimed at helping teachers to teach and students of all ages to learn and get on in their lives.

1.2 Qualifications offered by Edexcel include GCSEs and A levels, Key and Basic Skills, NVQs, professional qualifications and the BTEC qualification suite. In the UK, Edexcel qualifications are taken by over 4,200 secondary schools, 450 colleges, 80 Higher Education (HE) institutions, 800 public and private sector employers and a number of e-learning providers. Internationally, Edexcel operates in over 100 countries.

Part 2: Edexcel Responses

2.1 *Why do we have a centrally run system of testing and assessment?*

2.1.1 Since the establishment of the National Curriculum, testing has been a key central mechanism for driving up standards in schools.

2.1.2 Testing was the mechanism for making schools accountable in the drive to raise standards.

2.1.3 The need for improvement when compared to standards elsewhere in the world was high.

2.1.4 The issue is that forms of assessment that raise standards are not the same as those that are right for accountability.

2.1.5 Comparability became possible as all pupils were doing the same test at the same time.

2.2 *What other systems of assessment are in place both internationally and across the UK?*

2.2.1 Within the UK, it would be useful to examine the approach to testing in Scotland.

2.2.2 International comparison studies are available, particularly TIMSS and PISA.

2.2.3 The approach in Norway is to make assessment a major part of professional development. This approach differs from England and Wales in that national assessment is an event that is external to the school and hence leads to being something done to schools as apposed to the school being a part of the process.

2.2.4 The USA makes extensive use of testing. The difference between the UK and the USA is that of validity and reliability. In the USA the emphasis is on high reliability whereas our emphasis is on high validity. Whilst this is a generalisation and there are notable exceptions in both countries, the assertion is general true.

2.2.5 There are many commercial assessment instruments available to schools. The oldest and one of the most respected are the NFER standardised tests. Durham University has over the last few years established itself as a major provider of school tests.

2.3 *Does a focus on national testing and assessment reduce the scope for creativity in the curriculum?*

2.3.1 In the main, yes; this is a function of the nature of the National Curriculum. How restrictive a curriculum do we want? The more prescriptive the curriculum, the more restrictive will be the assessment.

2.3.2 There is a cost/benefit issue here; creativity is the cost of a prescriptive National Curriculum which has the benefit of being an effective driver for accountability.

2.3.3 Too much weight on the outcomes of assessment can damage creativity. The emphasis for the school can become understandingly, the test. Creativity, by its nature is not known to flourish during a timed test.

2.4 *Who is the QCA accountable to and is this accountability effective?*

2.4.1 The QCA is accountable to the DfES. There is an argument that testing and assessment should also be subject to independent scrutiny.

2.5 *What role should exam boards have in testing and assessment?*

2.5.1 The key difference between national tests and GCSE, GCE, diplomas, BTEC is that the latter are qualifications. Exam boards have a wealth of assessment expertise and could have a role in formative assessment. They have a global view of assessment over a large range of qualifications; they are well placed to position national testing into the gamete of predictive assessment and comparability.

2.6 *How effective are the current Key Stage tests?*

2.6.1 There is no single answer to this question.

2.6.2 They are fairly effective as curriculum drivers.

2.6.3 They are very effective at assessing part of the curriculum but there are aspects that are not susceptible to a timed test of short questions.

2.6.4 As a measure of progress for a school they can mask the many unexplained variables which may be making a significant contribution to a schools performance.

2.6.5 The introduction of national tests improved standards. It is less clear as to whether a plateau has been reached as to the contribution that national tests can make to further improvement.

2.6.6 The tests have been effective at raising standards through accountability. The future for raising standards may be a combination of standard tests and an assessment for learning approach.

2.6.7 As valid and reliable tests, they are the best that can be achieved in their current format.

2.6.8 The tests have been effective in bringing a common expectation of teacher performance. The questions remains, is it the right expectation?

2.7 *Do they adequately reflect levels of performance of children and schools, and changes in performance over time?*

2.7.1 They adequately reflect performance on short written tests; to what extent this reflects levels of performance for all children could be questioned.

2.7.2 They are good at giving a level across each curriculum subject. They do not give enough detail of each particular part of the curriculum.

2.7.3 They are good at reflecting the performance of schools.

2.7.4 There are too many expectations for one test. They do not support the subdivision of attainment into smaller levels. It would be useful for the Committee to revisit the Task Group for Assessment and Testing Report which gives a different complexion to the use of levels.

2.7.5 Whilst the tests give a useful indication of changes in performance over time, they are not suited to influencing major decision making. The curriculum has changed over time, new elements have been introduced and different approaches rewarded. To accurately measure such progress, the curriculum would need to be stable and the same test used each year.

2.8 *Do they provide assessment for learning (enabling teachers to concentrate on areas of a pupil's performance that needs improvement)?*

2.8.1 Sometimes they are good indicators of areas that a teacher should concentrate on. However, they are not diagnostic and many areas of understanding are not covered in a particular test. So they are not sufficient as the sole indicator of pupil performance.

2.9 *Does testing help to improve levels of attainment?*

2.9.1 Yes, indirectly through accountability.

2.9.2 Measurement on its own is not sufficient, levels of attainment improve when performance is targeted by teachers.

2.9.3 A consequence of testing is to narrow progression for schools that over prepare children for a particular level on a test.

2.10 *Are they effective in holding schools accountable for their performance?*

2.10.1 To some extent but there are possibilities for schools to use the value added indicator tactically by ensuring that there is adequate attainment in early tests and maximum attainment in a final years test. This practice is not widespread but is a reaction to what is perceived to be a strategic response to high stakes testing.

2.10.2 The tests report performance around the average; the extremes of performance are not obvious. 70% at Level 4 and above could mean both 70% at Level 4 or 70% at Level 5.

2.10.3 The tests, by their nature, provide a crude measure and it is easy for schools to find themselves in a comfort zone.

2.10.4 They only hold schools accountable for English, mathematics and science.

2.11 *How effective are performance measures such as value-added scores for schools?*

2.11.1 Value added is a useful measure.

2.11.2 There are ceiling effects for some schools; can you make significant improvement for ever? Low intakes make consistent performance a factor of the ability of a particular cohort of children.

2.11.3 The value added measure is not a transparent process; it is not easy to judge the validity of the variables.

2.12 *Are league tables based on test results an accurate reflection of how well schools are performing?*

2.12.1 They do not present the full picture. They provide a crude score of particular areas of a school's provision. That is not to say that they are not a useful measure; however, it would be misleading to use them as the single predictor of performance.

2.13 *To what extent is there "teaching to the test"?*

2.13.1 Everywhere, but that can be a positive thing. The test assesses the curriculum and so teaching to the test is teaching to the curriculum.

2.13.2 The problem is that short answer tests do not assess all parts of the curriculum. So excessive teaching to the test narrows the curriculum experience for the pupils.

2.13.3 Teaching to the test distorts the curriculum when taken to the extreme.

2.13.4 It is acknowledged that tests can define a curriculum but they are not the most appropriate driver for ensuring a comprehensive teaching approach.

2.14 *How much of a factor is "hot-housing" in the fall-off in pupil performance from Year 6 to Year 7?*

2.14.1 Many pupils are looked after at the end of Key Stage 2 with one to one attention to address areas in which they are having difficulties; so hot-housing is a factor.

2.14.2 It is not the only factor and may not be the most significant. Other to consider are: lack of expectation by year 7 teachers; lack of use of the information from the Key Stage 2 school; and, very importantly, the change of social situation for the pupils. In addition, pupils themselves are beginning to change.

2.15 *Does the importance given to test results mean that teaching generally is narrowly focused?*

2.15.1 On balance, yes for the core subjects; however, the introduction of end of Key Stage tests widened the curriculum for many pupils.

2.16 *What role does assessment by teachers have in teaching and learning?*

2.16.1 What role does teacher assessment play or what role should teacher assessment play? There is considerable variance within classrooms.

2.16.2 Ofsted says that teacher assessment does not play a significant role at present. This may be because assessment has been taken out of the hands of teachers; it is something that is done to the pupils from outside.

2.16.3 For some, teacher assessment is undertaken by mimicking the national tests; this is not the most productive way of using the opportunities that teachers have in the classroom.

2.16.4 The link between teacher assessment and learning needs to be strengthened; this will not be the case unless teacher expectation is that they are in control of formative assessment.

2.17 *Should the system of national tests be changed?*

2.17.1 On balance: yes.

2.18 *If so, should the tests be modified or abolished?*

2.18.1 National testing has too many purposes attributed on one test experience.

2.18.2 A national picture of standards could be found by sampling pupils.

2.18.3 For formative assessment; instruments should be provided that help teachers address the different aspects of the curriculum. Good formative assessment which influences learning will raise standards.

2.18.4 Teachers need to be trained in assessment techniques and interpreting assessment outcomes. Teachers should be doing assessment not administering an external test.

2.18.5 Tests which have been standardised should be an important addition to teacher assessment. The administration of such tests should be in the hands of the school.

2.18.6 Schools should be accountable. Assessments should be moderated and schools should be able to demonstrate progress and that they are raising standards. For standards to rise within a school, there needs to be attention to assessment outcomes, appropriate teaching, well developed curriculum guidelines and social structures, such as behaviour. All these aspects should be monitored by Ofsted.

2.19 *The Secretary of State has suggested that there should be a move to more personalised assessment to measure how a pupil's level of attainment has improved over time. Pilot areas to test proposals have just been announced. Would the introduction of this kind of assessment make it possible to make an overall judgment on a school's performance?*

2.19.1 Could be, it depends on the nature of the personalised assessment. Single level progress tests will not be sufficient to judge personal progress. Such assessments could lead to a distortion of the curriculum as schools focus on a competency approach to pupil performance.

2.19.2 We would support measures that incorporated a tool kit of assessment opportunities for teachers. These would include standardised tests and assessments when ready.

2.20 *Would it be possible to make meaningful comparisons between different schools?*

2.20.1 Yes, but a meaningful comparison would be more than performance tables of attainment on single level tests.

2.21 *What effect would testing at different times have on pupils and schools? Would it create pressure on schools to push pupils to take tests earlier?*

2.21.1 More than likely it would increase the testing burden.

2.21.2 At the end of a key stage the focus of the curriculum becomes narrowed as pupils are prepared for the test. This will be compounded by more frequent test exposure.

2.21.3 This can be ameliorated by a test design that complements teacher assessment.

2.22 *If Key Stage tests remain, what should they be seeking to measure?*

2.22.1 They should measure pupil attainment at the end of the key stage across as much of the programme of study as is appropriate for the test structure. As such they will give a national picture of standards.

2.22.2 Sampling would be sufficient and could identify trends and patterns.

2.22.3 School accountability should be by more intensive assessment measures as described above and moderated by Ofsted.

2.23 *If, for example, performance at Level 4 is the average level of attainment for an eleven year old, what proportion of children is it reasonable to expect to achieve at or above that level?*

2.23.1 It depends on how wide the band for average is to be. As with all things in this area, policy will dictate not pupil performance.

2.23.2 The level descriptions are meaningful, but their interpretation has been narrowed down to match expectations in the tests. It would be useful for the Committee to refer back to the Task Group on Assessment and Testing report which established the original 10 level scale.

2.24 *How are the different levels of performance expected at each age decided on? Is there broad agreement that the levels are appropriate and meaningful?*

2.24.1 They have become, de facto, the accepted levels as policy documents repeatedly stated the level of attainment for the average pupil.

2.24.2 They were originally standardised using teacher judgement and, once established, have to be maintained if comparison over time is to be meaningful.

2.24.3 The issue is, if standards rise and Level 4 remains average performance, then the difficulty of Level 4 needs to increase accordingly.

Introduction to ResultsPlus

In 2007, Edexcel will roll out a programme to help schools raise exam attainment and meet the personalised learning agenda. This may be of interest to the Committee as part of its inquiry into testing and assessment.

The programme, called ResultsPlus, will provide personalised information on exam performance to GCSE and A Level students, and to their teachers and head teachers. This has major implications as it will empower students and teachers with a new range of transparent and accessible information.

ResultsPlus represents a leap forward in personalised learning in the UK. This is made possible because Edexcel's digital ePen technology, which allows completed exam papers to be marked by trained markers on screen, is also able to produce a range of data based on exam performance.

ResultsPlus comprises four IT products:

- ResultsPlus Direct;
- ResultsPlus Analysis;
- ResultsPlus Skills; and
- ResultsPlus Progress.

ResultsPlus Direct

In summer 2007, all students of Edexcel GCSE and A Levels will be able to receive their results online for the first time via ResultsPlus Direct.

The results will feature a Gradeometer which will show students how close they were to the next grade up or down. This information provides transparency in the exam process and allows students and their parents to make informed choices about applying for the exam to be re-marked, or re-sitting.

This system was successfully piloted in 2006, when Edexcel provided 2,000 GCSE Maths students with their results online.

ResultsPlus Direct will allow students to go online from wherever they are in the world on results day and access their results using a unique PIN number.

In the traditional process, schools and colleges post lists of results on a notice board. With a secure online system, each student will see only their own results. Market research shows that 74% of people think that exam results should be available via the Internet.

ResultsPlus Analysis

In summer 2007, Edexcel will offer head teachers and school management teams a new resource, ResultsPlus Analysis.

It will provide analysis of results and performance at a cohort and individual student level. It will allow teachers to produce comprehensive reports to ascertain how the syllabus is being delivered and achieved against. If a group of students have not performed well in an area of the syllabus, ResultsPlus Analysis will highlight the problem and teachers will be able to adjust their teaching accordingly.

Edexcel will provide access to results information down to individual question level, as well as providing links to the examination papers, mark schemes and chief examiners' reports. This enables centres to compare their results against the national average, compare results by type of centre, download results data onto a spreadsheet and sort results by teaching group or gender and make detailed observations about students' performance.

This builds on Edexcel's Results Analysis service (RAS), which already allows schools and colleges to access their results at question level online.

ResultsPlus Skills

In addition to the performance information offered in ResultsPlus Analysis, ResultsPlus Skills will provide skills maps, so teachers will be able to see at a glance which topics and skills are causing their students problems.

By putting performance data into context, the skills maps will enable teachers to alter teaching programmes to raise attainment. For students who need to re-sit exams, their skills map can form the basis of a revision plan.

ResultsPlus Skills will be available when Edexcel's GCSE Maths and Science results are delivered in August 2007.

ResultsPlus Progress

ResultsPlus Progress will be introduced in autumn 2007 as online tests that will allow teachers to check the progress of their students' learning and identify areas of weakness that may require further teaching or revision.

Test results will be provided with skills maps for each candidate, tailored to identify their own strengths and weaknesses. This will help students plan their own revision and help teachers plan lessons more effectively and concentrate on weak areas. Using individual performance information to guide individual progress is at the heart of the personalised learning agenda.

ResultsPlus Progress will be available for Edexcel's Key Stage 3 Mathematics, GCSE Mathematics and 360 Science subjects from the start of the 2007 academic year.

June 2007

Memorandum submitted by OCR

1. Introduction

1. Assessment is a complex, long established, yet ever-developing science which is used throughout education and working life in a myriad of ways. In giving evidence, we have sought to confine ourselves to the context of the 14–19 qualifications system. In doing so we have sought to avoid highly technical issues, and have concentrated on systemic issues, seeking to point up some of the many strengths in the system as well as highlighting the tensions and weaknesses which need to be addressed.

2. In setting the context, it is important to establish the purpose of assessment. It can be used to provide ongoing feedback to the learner on their performance; it can be used to monitor progress and for diagnostic purposes, highlighting learners' strengths and where further work is needed.

3. A key purpose of formal assessment is to be able to issue qualifications. Qualifications provide structure to programmes of learning, ensuring coverage of the curriculum, and can influence the style and approach to delivery. Above all a qualification serves to confirm and celebrate the achievement of an individual. Qualifications must have value and recognition in society which requires that they are robustly assessed to ensure a comparable national standard across location, institutions and time, despite a constantly changing and evolving curriculum.

Important measures of success are that there is general trust in the system and that qualifications are valued by HE institutions and employers for recruitment and selection purposes. This requires a powerful but delicately balanced collaboration between teachers, employers, higher education and assessment experts whenever new qualifications are developed. The extent to which this delicate balance is achieved can be assisted or inhibited by the involvement of governments and their agencies.

The UK 14–19 Qualification System—A Success Story

4. It is important at the outset to emphasise the many strengths of the existing system. The system is well established and widely, if not perfectly, understood. In our experience it takes at least 10 years for a new qualification to become accepted and take root, so "heritage" is an essential feature of a qualifications system. It is no coincidence that GCSEs, GCEs and their equivalent forbears are the most widely recognised qualifications in the UK—they have been around for generations of learners. Nor is their success limited to the UK; in an increasingly knowledge-based global economy, it is no trivial mater that the international GCSEs and GCEs are offered in excess of 150 other countries.[1]

[1] Source: Cambridge International Assessment Submission to the Committee.

5. But the system is not stuck in the past. Over the years it has proven remarkably adaptable to social and economic change. At A level, it has moved from a 1950s model of providing a service to a tiny minority of aspiring university entrants, to a mainstream, backbone of a comprehensive education system where a university place is a realistic aspiration for all. It has adapted to accommodate the increase in school leaving age from 14 to 16, it has reflected every change in teaching styles and pedagogy, and most remarkable of all, it has accommodated massive changes to the curriculum.

6. The system remains highly adaptable, constantly introducing incremental change, such as the revisions to new A Levels, changes to GCSE coursework, the introduction of cutting edge technologies. At the same time, exam boards have quietly introduced new choices and flexibility within the curriculum, leading to highly successful new qualifications such as OCR Nationals.

7. Despite the difficulties associated with the introduction of Curriculum 2000 and a growing culture of general mistrust in public services, trust in the exams system remains remarkably robust:

> "The level of support for the A level qualification remains high and unchanged since March 2003 . . .. Indeed among A level students, there has been a significant decline in the proportion who believe A levels should be abolished (falling from 14% in 2004 to just 3% now) . . .. There has been an increase in confidence in the GCSE . . .".[2]

8. The regulatory bodies have played a significant role in providing a sound regulatory framework and responding to public concerns. It is still the case that, in the UK, the most critical question asked by employers and HE is not where did you learn, but what did you achieve? Nor is this just a matter of public perception. The independent review of standards commissioned by QCA concluded:

> "It is our considered judgement that QCA has done a commendable job in its effort to assure quality of the A level examinations, especially as QCA is a developing organisation. In addition, it must contend with a raft of notable changes: in curriculum, examination practices, consolidation of awarding bodies, policies seeking to expand upper secondary and university enrolment, and increased school accountability, among others".[3]

9. The system is supported by a range of independent exam boards, each with their distinctive strengths and heritage, competing to deliver efficiencies and to modernise infrastructures. The accuracy and precision of the examination system is one that would be the envy of almost any other industry:

		Performance in 2006		
	QCA expectation	*AQA*	*Edexcel*	*OCR*
% of question papers dispatched to centres on time	100	100 (100)	100 (100)	100 (100)
% of question papers without errors	100	99.1 (99.0)	98.7 (99.1*)	98.6 (99.2)
% of examination results issued to centres on time	100	100 (100)	100 (100)	99.9 (100)
% of priority enquiries about examination results completed within 20 days	100	100 (100)	100 (100)	100 (100)
% of examination papers copied and sent out at least 10 days before the deadline for enquiries about results	100	100 (100)	100 (99.8)	99.7 (100)

Equivalent figures for June 2005 are provided in brackets. Percentages shown to nearest 0.1%.

* The 2005 figure for Edexcel has been revised from 98.0% to 99.1% to enable like for like comparisons between awarding bodies and years.[4]

10. The examination system has seen unprecedented investment in technological advances estimated at around £150 million over the last 10 years with relatively little direct support from the taxpayer.

11. The exam boards are now well-placed to take on the challenges and opportunities that new technologies will provide. Many innovations are already in place, such as on-demand electronic tests; adaptive tests, e-portfolios, and industrial scale electronic script management. It is important that a measured and long term view is taken with technology which isn't just about relatively simple on-screen objective tests or replicating paper-based conventions in an electronic format. Eye-catching initiatives, such as the Adult Basic Skills tests, require greater scrutiny and analysis. The real opportunities for assessment lie in the creation of new virtual environments and interactive processes. A glimpse of what is possible is the virtual geography field trip created by Cambridge International Assessment, which makes it possible for students to interact with a simulated tropical rain forest, enabling their skills to be tracked as well as their answers logged—a significant enhancement of pen and paper work.

[2] GCSEs and A levels: the experiences of teachers, students, parents and the general public, QCA February 2006 (results of an independent survey commissioned with MORI).
[3] Maintaining GCE A Level Standards, QCA October 2001.
[4] Report on the performance of awarding bodies for general qualifications in 2006, QCA, 2007.

A Far from Perfect System

12. Despite its many clear benefits, there is growing unease from many stakeholders that there is something not quite right with the system. The sources of these tensions are not straight forward but we believe the system is most dysfunctional when policy makers and regulators begin to intervene too closely in assessment design, set prescriptive and unnecessary requirements, or actively participate in the development and implementation of qualifications.

This dissatisfaction tends to cluster around the following issues:

"The system is too burdensome"

13. The system is sometimes caricatured as being heavily bureaucratic and expensive and it cannot be denied that the administration of valid and reliable examinations comes with an associated amount of bureaucracy and cost. The submitting of accurate candidate data, the administering and timetabling of exams, the collection and distribution of results will never be entirely free from bureaucracy.

14. There have been huge advances in recent years such as:

— on-line entries and results services;
— the expansion of Exam board customer support services;
— the NAA's programme to "professionalis" exams officers;
— direct capital spending on exams offices;
— completion of the JCQ initiative "Eight Pledges to reduce Bureaucracy"; and
— the centralisation and consolidation of exams offices within large institutions.

15. Yet there is still a feeling that more can be done to take pressure out of the system and the Joint Council for Qualifications (JCQ) and the Federation of Awarding Bodies (FAB) remain committed to working together wherever possible to ensure that the systems and processes we require customers to use are aligned.

16. To some extent, the level of bureaucracy is determined by the assessment model used for any particular qualification, along with its structure and design. It should be noted that, as a rule, locally assessed, centrally moderated models (such as coursework) tend to be more bureaucratic than examinations. Yet often Government requirements or regulatory mandates have been at one remove from the impact on organisations delivering qualifications. It was the regulator, supporting government policy, that determined in its 2002 GCSE criteria that each GCSE must contain coursework. This was driven by a perfectly valid view that coursework supported the important practice of "learning by doing" or "applied learning", but the mandatory requirement across all subjects was disproportionate to the gain, made the volume of coursework unmanageable, and brought a perfectly valid form of assessment into disrepute.

17. Too often policy has driven solutions which take no account of their cost, manageability and impact. Key Skills, the first qualification to be designed and run by the regulator, became the only single unit qualification that had both an external test and a locally assessed portfolio, both designed to assess exactly the same thing. This belt and braces approach derived from a ministerial view that external tests were "rigorous", but with an acceptance that local assessment was the best way of confirming that learners could actually apply their Key Skills to real tasks.

18. The blueprint for "specialised" Diplomas, as laid down in the 14–19 Education and Skills White Paper, appears not to have taken into account cost and complexity or to have acknowledged at all the bureaucracy that would be needed to underpin it. A delegate at a recent OCR conference described the Diplomas as "every exam officer's worst nightmare". The requirement to provide an overall grade for the Diplomas has been additional decision which does not seem to take into account the additional burden on schools and colleges.

19. Once initiatives have been put in place, the regulator and the system in general is capable of learning from, and correcting, mistakes. The reduction of A levels from six units to four and revisions to GCSE criteria are examples of this. But QCA has been less successful in trying to reduce the burden at the implementation end. The English regulator's project on common centre approval (widely believed to have cost £1 million pounds) seems unlikely, after more than two years, to deliver any tangible benefits. It is just one of a series of misguided attempts to find quick fixes and magic bullets where there are none.

20. Whilst much has been made of the costs of assessment in our system, and there are undoubtedly further efficiency savings to be made, more attention needs to be paid to evaluating the alternatives. A substantial case could be made to show that qualifications developed for large national uptake, in a competitive market, will always generate greater efficiencies than a more local, devolved model (ie teacher assessment). Although we will highlight the risks of over-dependence on exams as an assessment model, they are undoubtedly the most cost efficient way of assessing large numbers of people. Discussions with the Learning and Skills Council have led us to believe that the spend on exam board fees is a miniscule proportion of their overall spend. When looking at cost, we would recommend that some consideration is given to the costs of alternative systems operated in other countries.

"There is too much assessment"

21. This is intrinsically linked to the theme of burden and bureaucracy but refers to the view that learners themselves are subject to too much assessment. Again, much of this is a consequence of policy decisions or too inflexible, one-size-fits all, regulatory requirements. As already stated, many of the forthcoming reforms will address this issue. Whilst we are likely to see increased modularisation in the new GCSEs, which will lead to more ongoing assessment, this will be balanced against a reduction in coursework.

22. Moreover, we are developing a view that the spreading out of assessment over a longer period of time is less stressful than a concentrated period of assessment at the end of a two year period. In addition, our conversations with teachers lead us to believe that it is the sustained, unnecessary and inappropriate mass testing of very young people through the key stage national tests that is the single biggest cause of the view that there is too much assessment. (This is discussed at greater length in Cambridge Assessment's submission to the Committee.)

23. The amount of qualifications taken and achieved, and therefore the amount of assessment undertaken, has risen significantly in recent years.[5] A greater number of more qualified people must largely be a good thing, although there are some behaviours, such as the amount of re-sits taken, which need further analysis also there is a tendency for some candidates to take increasing numbers of GCE and GCSEs, when higher level qualifications, requiring greater stretch and challenge, or different more skills- based qualifications, might be more appropriate.

The Range of Qualifications is too Narrow, Stifling Innovation and Choice

24. This phenomenon of learners being entered for increasingly large numbers of GCSE/GCEs may be a symptom of the view that the current curriculum is too narrow.

25. In truth, the curriculum has been getting broader and richer exponentially over the last 10 years. With pathfinder projects to 'flex' the curriculum and the introduction of increasing numbers of alternative qualifications on section 96 (the list of qualifications approved for use in schools), we have seen a transformation of what many schools and colleges now offer. The 'Entitlement Framework' in Northern Ireland and the Welsh Baccalaureate are commendable examples of proportionate initiatives to broaden the curriculum offer.

26. However, many traditional institutions have yet to embrace these new opportunities; the dominance of General Qualifications in terms of recognition and the value placed on them, combined with poor levels of advice and guidance, means that some learners are still condemned to an inappropriate diet solely of General Qualifications. This may be why some employers and HE institutions still complain that current qualifications do not provide rounded, multi skilled, motivated young people.

27. Government policy on assessment has tended to reflect a nervousness of any form of assessment other than a formal examination. A belief that this is the only 'rigorous' way of assessing achievement has led to many alternative qualifications which assess the practical application of skills with suspicion. This was one of the key reasons for endless tinkering with the GNVQ and its final withdrawal and why we ended up with Vocational GCSEs were developed in such a way as to look the same as any other GCSE. Some have accused the current Diploma developments of showing signs of 'academic drift'.

28. Tomlinson recommended the withdrawal of A Levels and GCSEs as the most certain way of resolving the domination of these qualifications over the rest. OCR's view has always been that this solution is too drastic, and that it is possible to work towards a fuller curriculum by gradual and careful enhancement of the provision. Single, one hit initiatives, such as GNVQ, or possibly even Diplomas, are not easy or guaranteed solutions. There are other ways of implementing policy, as Cambridge International Examination's submission highlights.

29. Nor should we assume that general qualifications need to be wholly knowledge based, purely academic qualifications. Once again, the system continues to adapt and improve in response to change and to learn from its mistakes. The last round of GCSE developments took place against fairly prescriptive regulatory criteria which, narrowed the opportunity to develop stimulating and imaginative qualifications that would engage learners and allow teachers to bring a wide range of teaching styles to the classroom. We believe the new criteria will allow us to develop far more engaging new GCSEs and that their introduction in 2009 will be something of a watershed. A precursor has been the new GCSE in Science suite 'Twenty First Century Science', introduced in September 2006, ahead of the other new GCSEs. This highly practical suite, rooted in contemporary, relevant contexts has proven a runaway success with teachers and learners alike at a time when interest in Science appeared to be in terminal decline.

[5] "The success rate for 16–18 year-olds taking full Level 2 qualifications has improved by 9 percentage points over a two year period to 67% in 2005–06, with the success rate for adults following the same programmes increasing by 11 percentage points to 66% over the same 2 year period".
Further Education and work-based learning for young people—learner outcomes in England 2005–06, Learning and Skills Council, April 2007.

Using assessment to measure the wrong thing

30. The purpose of qualifications has already been set out. Problems arise when they are used as a proxy measure for completely different things. The use of qualifications in school performance tables, national targets, OECD comparisons etc leads to misinformation and drives undesirable behaviours. The use of performance tables in particular, leads to cynicism and a devaluing of qualifications and learning. Assessment is designed around that which it is intended to assess—it is no doubt wholly possible to devise mechanisms (many of which already exist) to assess the performance of schools and colleges, but qualifications are the wrong instrument. Cambridge Assessment's submission deals with this area in more detail.

2. Section 2: Questions

Is the testing and assessment in "summative" tests (for example, GCSE, AS, A2) fit for purpose?

31. By and large the assessments used in GCSE, AS and A2 are wholly fit for purpose and are adapted to suit the different nature of what is being assessed, albeit within a framework of regulatory criteria. Where wider skills and different learning styles come into play, other assessment models and qualifications are available outside of those offered by general qualifications.

Are the changes to GCSE coursework due to come into effect in 2009 reasonable? What alternative forms of assessment might be used?

32. It is a shame that coursework has been dropped across many subjects. All assessment requires trade-offs between validity (its ability to reflect the genuine level of achievement of a learner) and its reliability (its ability to produce the same outcome for learners who reach the same level of performance). Well designed coursework can sometimes deliver validity which sit-down examinations do not because of the inherent artificiality of the examination environment. Much of the criticism of coursework relates to scope for plagiarism; this is only a risk under the increasingly prescriptive regulation applied to coursework which means that all candidates undertake similar tasks, rather than the highly personalised work which characterised the original introduction of coursework. It is also disappointing that coursework has been scaled back at a time when technology, in the form of e-portfolios which OCR has already trialled in several subjects, offers the opportunity for learners' work to be monitored and audited in ways which were not previously possible.

What are the benefits of exams and coursework? How should they work together? What should the balance between them be?

33. In reality, there are a great many assessment models—examinations and tests come in many forms and can involve controlled assignments, practical activities, orals etc. Coursework might include the creation of artefacts, musical compositions, major project management activities and research projects. Often assessment takes place whilst a learner is performing a task, say in a dramatic performance, conducting a live experiment during a field trip, dealing with customers during work experience and so forth. It is necessary to design assessment around what is to be assessed.

Will the ways in which the new 14–19 diplomas are to be assessed impact on other qualifications, such as GCSE?

34. So far it has been the other way round. In seeking parity with GCSE and GCE, the main parts of the Diplomas have increasingly adopted models which mirror the models for GCSE/GCE laid out in the regulatory codes of practice. The grading structures have also been adopted to mirror GCSE/GCE scales. The diplomas already have a minimum of four different assessment models within them. The increased modularity of GCSEs from 2009 should encourage greater take up of components of GCSEs within Diplomas. With careful management and a vision that goes beyond 2013, it should be possible to see components of GCSEs, Diplomas and other qualifications being combined to create new choices which are not as narrow as the prescribed Diploma routes, and not as unvaried as a diet entirely of General qualifications.

Is holding formal summative tests at ages 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?

35. Evidence suggests that the majority of learners prefer assessment to be spread out over time and closer to the learning experience. OCR has supported QCA in developing guidance to teachers on synopticity at GCE to ensure that candidates are able to demonstrate a full grasp of a subject across and between topics.

How does the national assessment system interact with university entrance? What does it mean for a national system of testing and assessment that universities are setting entrance tests as individual institutions?

36. It has never been claimed that A levels should be the only tool used to determine university entrance (or recruitment to employment for that matter). The Cambridge Assessment response provides information on how university tests provide additional information about potential undergraduates, and where they seem to merely replicate existing 14-19 assessments (as with SAT 1)

General Questions

Who is the QCA accountable to and is this accountability effective?

37. We understand that QCA is accountable to the Secretary of State, who is accountable to Parliament. What is less clear is precisely what QCA is accountable for. This is a consequence its somewhat open-ended statutory remit as defined in the 1998 Education Act, a tendency for ministers to add to QCA's remit over time through ad hoc remits and funding, and a tendency over time for QCA to shed its independence from Government and to act as the Government's implementation arm for qualifications policy.

What role should exam boards have in testing and assessment?

38. We believe our evidence sets out clearly the key role of examination boards in developing and delivering qualifications for 14–19 year olds. We would like to emphasise again the delicate balance of stakeholder interests that must be obtained to ensure a valued qualification system, and the unique position of examination board, which possess deep understanding of assessment and qualifications, and sit independently of government and between the many stakeholders enables them to play the role of consensus builder, provided that regulation does not prevent them from doing so.

June 2007

Witnesses: **Dr Andrew Bird,** Deputy Director General, Assessment and Qualifications Alliance (AQA), **Murray Butcher,** Director, Assessment and Quality, City & Guilds, **Jerry Jarvis,** Managing Director, Edexcel, and **Greg Watson,** Chief Executive, Oxford, Cambridge and RSA Examinations (OCR), gave evidence.

Q173 Chairman: May I welcome our witnesses to this session of the Children, Schools and Families Committee? We are very pleased to have such a talented group of experts with us this afternoon and we hope to learn a lot from them. As we have at least six sections to cover, I hope that you will not mind if we cut a section to move on to the next. We really could spend a couple of hours on each section. Sometimes I will rather rudely say, "Quick questions and quick answers." Do not get upset about that. Will you all introduce yourselves? We have your CVs, so there is no need for you to say anything about them. Starting with Andrew, have you any one thought that you would like to raise before we start the questions and answers?
Dr. Bird: I am Andrew Bird from the Assessment and Qualifications Alliance. I take it that you are looking for an opening statement from us?
Chairman: It depends how long your opening statement is.
Dr. Bird: A minute.
Chairman: You can have a minute. Before you came in, I was saying to my colleagues that you used to do a really useful job in a fantastic chemical company in Huddersfield.
Dr. Bird: That was a few years ago.
Chairman: They did not believe that I was going to say that. Andrew, you are very welcome. Please give me your minute.
Dr. Bird: First, AQA is an independent charity. The board of trustees is drawn from the teaching profession, higher education and business. Our one purpose is to "do good in education". We aim to discharge that by giving high quality qualifications in respect of teachers, parents, employers and HE, by delivering new qualifications and modes of assessment that meet the needs of today's and tomorrow's learners, by providing the best level of training support to teachers who deliver our specifications, and by carrying out and publishing research into educational assessment. May I draw your attention to a couple of points that we raised in our written evidence, and then two that have arisen since? The first one, drawn from our written evidence, concerns functional skills and hurdles for GCSE. The policy intention is to impose a functional skills hurdle at Level 2 of GCSE on English, maths and ICT. On considering our research, we are concerned that when it is introduced, it will de facto be a change of standard. From our modelling work, it will suppress the pass rate for A to C at GCSE. The policy position is that making such things explicit will lead to more discreet, direct teaching of those skills and, hence, a rise in performance. That might be true, but we need to consider the policy implications of that. We have no problem with raising the standard. In fact, we think that that is a good idea, providing that we understand it and we all know what will happen as a consequence of that at the transition point. Secondly, throughout our evidence, we are quite keen on diversity of provision: giving the choice to teachers and advisers to give students the widest

range of curriculum opportunities. It is important to remember that only a small number of our students do three A-levels. Many do one or two A-levels, and they would find the Diploma to be too much at Level 3. Diversity of providers—meaning people such as ourselves—drive competition in service delivery and support, which, we believe, helps innovation. Evidence from contractual models suggests that, in so doing, the Qualifications and Curriculum Authority drives out innovation. I also want to raise two items that are drawn from the more recent past. Clearly, we welcome the Government's intention to separate the regulator. As we mentioned that in our written evidence, we cannot do anything else at this stage. However, we want to draw two points to your attention. One is the need for a willingness to co-ordinate and integrate regulation across the three countries of England, Wales and Northern Ireland. There is a de facto single market in qualifications but, as you will be aware, the policy position in those countries is diverging. Therefore, from a regulatory point of view, there needs to be a bringing together of regulation. We are also concerned that as the shadow regulator—sorry, it is the interim regulator at this stage, I am told. As that process has got under way, the focus has seemed to be mainly on picking a new name for it, rather than considering the technical capability and capacity it requires to be an effective regulator. We are concerned that the result will be a stifling, box-ticking bureaucracy, rather than a strategic regulator of our activities. Finally, we think there is an emerging dilemma between two terms that we hear a lot from the Government and regulator. This is the whole high-stakes environment versus light-touch regulation. We obviously welcome appropriate and sensible regulation that aligns to the five principles, but we see the intention to extend the availability of qualifications from colleges and workplaces and, in those situations, encourage light-touch regulation—which one can understand to help people enter the market and to ensure that those qualifications are acquired and certificated—as working against those qualifications being portable and having utility. If regulation is only light touch, it is in danger of not meeting the standard of regulation that the high-stakes qualifications are put under. Qualifications need to command respect, and not just from the initial provider of those qualifications. Thank you, Chairman.

Chairman: I let you get away with that even though it took more than a minute.

Murray Butcher: Good afternoon. Thank you for inviting me. I am Murray Butcher, director of assessment and quality at City & Guilds, which is the UK's largest vocational awarding body. Established in 1878, it received a royal charter from Queen Victoria in 1900. We provide about 500 vocational qualifications in diverse occupational areas, ranging from agriculture to zoo keeping. City & Guilds currently comprises four qualification brands: the City & Guilds, which is the wide range of vocational qualifications; the Institute of Leadership and Management, which covers first-line management and beyond; the Hospitality Awarding Body, which relates primarily to hotel and catering qualifications; and the National Proficiency Tests Council, which covers all our land-based awards. We operate in the UK and internationally, covering about 100 countries and working through about 8,000 centres. A centre can be a school, college, university, training provider or employer. I will seek to make any other points I can during the general questioning.

Chairman: Fine.

Jerry Jarvis: Good afternoon, ladies and gentlemen. I will be brief. I am Managing Director of Edexcel, which is a Pearson company. We are known principally for our technology. The only thing I would like to say in an opening statement is how struck I was by what Ken Boston said in his evidence on 17 December. He picked up three key issues that he said were critical in improving attainment in this country. First was the provision of personalised learning; the second was the provision of continuous analytical testing and evaluation; and the third was the professional training of teachers. We strongly endorse that view. We also believe that those three factors are the issues that would most quickly improve attainment in this country, and we have invested massively in the provision of a framework to do that. Like my colleagues, I am very pleased to participate in this inquiry and I look forward to the recommendations and the outcome.

Chairman: Thank you.

Greg Watson: Good afternoon. I am Greg Watson, the Chief Executive of OCR—Oxford, Cambridge and RSA Examinations, to give it its full title. We are a major UK awarding body that principally makes qualification awards to 14 to 19-year-olds in schools and colleges. We are exactly 150 years old and a not-for-profit social enterprise. For a century and a half, we have developed assessments of various types to structure, motivate and reward learning. We are a member of the Cambridge Assessment Group, which is a well known international education group that operates in more than 150 countries. In many of those countries, it offers assessments similar to the style of assessment that we have here in the UK. In the past 10 years there have been three developments of note, which perhaps we shall have an opportunity to explore in this inquiry. The first is a growing use of qualifications as a public policy lever, and with that a widening of the uses to which assessment is put beyond the original purpose of structuring, motivating and rewarding learning. I am thinking of uses such as measuring school performance. Secondly, there has been more frequent change at both the system-wide level and that of individual qualifications, and some short-circuiting of long-established disciplines of evaluation and research based on hard evidence. Thirdly, and connected with the previous two, there has been an imperceptible but worrying loss of public confidence and a feeling that somehow things are not quite what they used to be. That concern has become harder to deal with because of the many uses to which assessment and qualifications have been put and the difficulty of explaining and assessing the impact of change. We very much welcome the Secretary of State's announcement before Christmas that a new

independent exams regulator will be created. We see in that a once-in-a-generation opportunity to deal with the three issues that I have mentioned and to put the exams system in a position of being seen to be sufficiently independent while commanding public confidence, and for the regulator to have a key role in balancing the desire to innovate and keep pace with society with a desire to maintain stability and integrity over time.

Q174 Chairman: Thank you very much for those openers. May I open the questioning by asking Jerry Jarvis something? I shall start with him, as he is in the middle. People used to say that the trouble with our examinations was that we had a number of boards, and that what we needed was one big board that did everything. That would stop competition and prevent people from switching from one examining board to another, and everything would be a lot tidier if one board did the job that the four of you do. Is that not an unanswerable proposition?
Jerry Jarvis: Inevitably, I have a personal view. I spent a long career outside education before coming into it, and I am used to competition being used to drive up standards and reduce costs. My observation is that we benefit massively from having competition in the marketplace. The huge majority of teachers who choose the specifications of examination systems tell us that they value the choice that they have. That choice certainly makes me compete strongly with the colleagues who are sat beside me. Without it, we would not have the degree of ingenuity, purpose and lead that we have in this country, nor the stability and reliability. Competition has been very good for education.

Q175 Chairman: What if a teacher or head teacher said to you that the danger of competition between examination boards is that everybody knows that if you are being pushed and pushed to raise standards and raise the number of people getting grades A to C and so on at GCSE, A-level and other levels, people go for the easiest pass? Reputations go around, and people say, "It is easier to take English at GCSE or A-level with that board," and they switch around. If you are really going to compete, you will just become known as the easiest board from which to get qualifications so that you can wipe out the other three.
Jerry Jarvis: That is a popular view, but it is generally not held up by fact. I am one of three accountable officers in this country. I am responsible for ensuring that each award made by Edexcel is made under strict scrutiny and that the standard is maintained across time and in comparison with other awards. I do not have the ability to interfere with that standard. If you look at the appearance of so-called easy qualifications, the arguments tend to break down when you get into some of the detail. For example, the pass rate in GCSE maths is higher than the pass rate in media. Does that mean that maths is easier than media? Because of choice, these days, students will take the qualifications that they enjoy and are good at. Ken Boston put it eloquently when he drew the difference between the standard that is the hurdle that students must achieve, and the standard that is expressed as the number of students who have actually achieved that standard. We do not and cannot compete by producing easy qualifications.

Q176 Chairman: Why then, Greg, are so many people and parents out there, let alone the poor old editor of the *Daily Mail*, unhappy and feeling that standards have gone down and that kids do not work as hard or get qualifications of the same standard as when they were at school? Why is there a general feeling that things ain't what they used to be?
Greg Watson: Let me offer two possibilities. First, qualifications have changed and evolved. The A-levels that young people sit today are not the same as those I sat, with good reason. The need for the routine replaying of a large body of knowledge has probably weakened slightly as access to information has become easier. On the other hand, industry says that it wants people who are more skilled in using that information—in applying it and being able to think for themselves. In A-level, we have seen a shift over time so that the body of knowledge in any given subject is probably a bit smaller, but the skills needed to apply that knowledge have moved to a slightly higher level of demand. Some of the commentary is simply an unfamiliarity with how the qualifications have changed—they do not feel the same. Secondly, I think that there is a misunderstanding about the nature of competition among the people sat at this table. Ours is a competition of not standards, but ideas. Because we are all independent organisations, because we are all close to the business of teaching and learning, and because we find ourselves between schools and colleges on one hand and universities and employers on the other, I think that we feel driven to look for new approaches. Look at what has happened with GCSE science recently. There has been a real rejuvenation of science in the classroom because of a particularly innovative programme that we at OCR have developed in partnership with the Nuffield Foundation. Look at what is happening with geography at the moment. We are running a groundbreaking pilot with a different approach to geography that reconnects the concepts of geography with a study of the real world. That helps young people to make more sense of the subject. One reason why we have a 150-year tradition in this country that so many other countries overseas want to buy into is that we have had the power to innovate and a competition of ideas and subjects. They have helped us to keep subjects fresh and interesting and to offer different approaches to different young people who want to learn in different styles.

Q177 Chairman: Andrew, you are a scientist by training, are you not? Why do people say that people are shifting to what are perceived as easier subjects? We are still having difficulty in attracting enough people to carry on with maths, physics and the sciences in general. Even geography seems to be losing students, despite the new course that Greg Watson has just described to us. With what you are

providing, are you not colluding to stimulate movement away from the hard scientific subjects to subjects that are perceived to be easier?

Dr. Bird: Perception is everything, is it not? We are trying to reflect those things that students want to study that are relevant to commerce and work today. Media studies qualifications meet a student need, and teachers feel that students would enjoy learning it. Through it, students can collect important basic study skills and skills for future employment. Is that easier than science? I did French and science at school and found the former incredibly hard. Was the French exam easier than the science exam? It was much harder for me, because of my ability. People find what they enjoy easier, so I found maths quite straightforward, whereas other people find it very difficult. A lot of this is about perception. We work extraordinarily hard to ensure that the level of demand among subjects is maintained over time, and we do that by using experts in the classroom and expert examiners—people who are knowledgeable about their subject. We cannot undo the perception of, "Well, it is not what I was taught at school"—nor should it be, because times, demand and needs have moved on—"and I cannot connect with or understand it, so I do not appreciate that it is as difficult as, say, physics."

Chairman: Right, I have warmed you up. Murray, you will have to wait for a moment, because people will get testy if I carry on. Now we will drill down a bit, and John Heppell will lead us.

Mr. Heppell: Sharon can go first because she has to leave.

Chairman: Sorry, Sharon, you are next. They have switched; they have done a secret deal.

Q178 Mrs. Hodgson: I want to ask a couple of quick questions and then go, because I am due in the Chamber on Bench duty. I want to talk about evidence that we received from Professor Dylan Wiliam, who argues that although A-levels have not necessarily become easier, examinations no longer measure what they used to. From that, he infers that a pupil achieving a top grade does not necessarily have the same skills as a pupil who achieved a top grade years ago. How are the gatekeepers to further and higher education, and employers, to compare students in similar subjects, but from different years, given the changes in qualifications?

Jerry Jarvis: Our examination system is complicated and driven by populism. It is actually very difficult to compare an A-level taken in 2007 with one taken pre-Curriculum 2000. The structure is different, and we are examining different things. Access to A-level education was different some 10, 15 and 20 years ago, so the cohort taking those examinations was also different. However, we can perhaps see a continuing thread through the regulator's work in attempting to maintain a standard in A-levels over the years. Truly speaking, however, we can compare precisely only A-levels that were taken since the introduction of Curriculum 2000. I shall return to the mantra that I am sure that you will hear time and again when speaking to anyone from an examination board or awarding body: we have attempted to maintain the hurdle at the same height, even though the features that we are examining are different. What has changed quite dramatically is access. There is far more choice, so, for example, students can take a number of AS-level examinations and continue the AS-level studies that they are best at. There is multiple access to resits, modular variants, and so on, which have increased the probability of students attaining that same fixed standard. There is a very difficult notion to get across, so it is easy to say, "When I did A-levels, they were much harder." Our students work very hard for A-levels today, and something in the region of only 3% of 18-year-olds achieve three A grades.

Q179 Mrs. Hodgson: If you are saying that standards have been maintained and the hurdle is still at the same height, how can you counter the claims by some universities that school leavers entering the first year do not have the same depth of knowledge that students with the same grades had years ago? The universities are saying that.

Greg Watson: I think that you have to bear in mind that the role of A-levels in the education system has changed over time. There was a time when A-levels were purely for those entering higher education and they were actually offered to a pretty small part of the 17 and 18-year-old age group. A-levels moved over time to become the standard school-leaving qualification, in many ways, and that will be even more the case if the rate of those staying on to 18 continues to rise. As A-levels have evolved, there has inevitably been a trade-off between ensuring that the qualification is suitably motivating and providing the right structure for learning for a wide range of young people, and making sure that it is a good basis for university entrance. I would recognise that, in the drive to widen the use of A-levels, we have lost a little, and that is why we have come back and started to look at the stretching of the upper end of A-levels to make sure that we reintroduce a little more stretch for the most able exam candidates and give some universities more of an ability to see who the most able are. We should also draw a distinction between the year on year reliability of standards and the long progress of the history of standards. Year on year, all of us at this table go to great lengths—in fact, the QCA's independent review last year said that we go to greater lengths in this country than anywhere—to guard standards in our subjects. What Jerry said was right. Over the decades, we have seen a number of structural changes, which have all been there for a good reason, but we have lacked an independent assessment of the impact of those changes. As I said in my opening remarks, I see it as a positive development that the Secretary of State wants to put the regulator some distance from the Government of the day and Ministers. One role that the regulator will be able to play will be to look at change in the system, to evaluate the impact that any change might have on standards and public perceptions, and to see whether we are happy to make a trade-off for the benefit that we get from making that change. That will be very welcome.

Chairman: Sharon, have you finished?
Mrs. Hodgson: Yes. That is great. Thank you.

Q180 Chairman: Before you move on, Greg, can you tell us whether you share Andrew Bird's view? He said in his introduction that he was worried about the regulator, who might be a tick-boxing person with no strategic role at all. You are very complacent about that.
Greg Watson: No, I think that is absolutely the right move to make. There is a lot in the detail, and we are all involved in a consultation process with the Department. There are some important nuances about where the line should be drawn between what remains under a Minister's direct control and what is the responsibility of the independent regulator. I have some views on that, and I guess that we do not have enough time to explore them in detail here, but, yes, the fundamental point in terms of getting the regulator right is to put the right responsibilities and activities on the other side of that independence line so that we get the building of confidence that we are looking for.
Chairman: We will come back to that. Sharon is off to some important duty on the Front Bench, so I call John Heppell.

Q181 Mr. Heppell: I am left feeling that there is no proof that standards have dropped, but also that there is no proof that they have been maintained. You said that it was difficult to see what you could do about that, and I find it difficult to see what measurement you could make, given the change in the system. However, you may be right to suggest that the regulator will ease some people's difficulties. But what about the implications for the validity and reliability of the work you all do in providing sample questions and answers and targeted syllabus training for teachers, including comprehensive teaching and learning materials about what exams will be about? Does that not encourage people to get children to study those things that are relevant to the test, rather than to the broader education that all of us are looking for?
Dr. Bird: As you rightly said, we all provide curriculum support materials, training for teachers and such like. The publishers of books in support of our products also include sample questions, tests and so forth. I do not detect that that is very different from what it has ever been. Past exam textbooks have always included past questions to help people prepare effectively and it is surely better that people understand the style of questioning that they are going to face than that they do not. Our challenge in assessment terms is to make sure that we do not become formulaic, that we cover, over time, the whole curriculum— although we do not cover the whole curriculum necessarily every time we set a particular test—so that it does not become predictable and that it discourages people from things like question-spotting, which is not a new phenomenon either. I think teachers have always done an element of that. What we are trying to do is provide the very best access for teachers and the very best guidance and support, so that teaching experiences in the classroom can be exciting and informative and can carry students' interest forward, so that they are successful. We are all interested in children being successful, but we separate that very carefully; we do that through separate parts of the organisation. The setting of the exam is an independent operation.
Chairman: John?
Mr. Heppell: I thought Greg wanted to say something.
Greg Watson: Just to add to that—
Chairman: Greg, you have to catch my eye.
Greg Watson: I forgot my flag today. I have a couple of thoughts. The first is that I want young people to learn to the syllabus. The syllabus is an important definition of a programme and the whole point of having a syllabus is to structure and motivate someone to want to learn. I do not think that is any different for a professional exam in later life or for a school exam. I think it is a good development that whereas I never did see my A-level syllabus and wondered till the bitter end what exactly it was I was heading towards, young people today have a pretty good idea from the start. That has been healthy. Secondly, I also want young people to demonstrate what they can actually do, in the exam. By being a bit more open, a bit more transparent, and providing a few more clues, we enable people to feel well prepared, and what we are actually is assessing what they can do not how successfully they have guessed what they are about to do, or their ability to cope with the surprise of what they have been faced with. I think that that has also been a positive development. But I would set against that the fact that there is a challenge and we employ expert people in the field of assessment to make sure that we keep the level of challenge right, and that it does not become too formulaic and too predictable, which obviously would mean beginning to lose some of that effect.

Q182 Mr. Heppell: I want to add a further thing about the idea that there is a sort of competition to pick the easier course. We actually had someone in this Select Committee telling us that that was common practice in schools; that schools would look at the syllabuses and pick the easiest. What about from the other side? What about universities that blacklist qualifications because they say they have been made too easy?
Chairman: That is common practice, is not it?
Mr. Heppell: Is it common practice?
Dr. Bird: I think some of the universities have got a list of subjects that they would encourage people not to do more than one of, in the sense that if you were doing a number of these newer subjects—media studies and something else, maybe—they feel that that might unbalance your curriculum. I do not think they have actually blacklisted A-levels. They have discouraged people from doing too many of them.

Q183 Chairman: It is a bit of a mafia, though, isn't it—all the senior tutors in charge of admissions at Oxford and Cambridge get together on a regular basis. They do not have to have it written down on a list.

Greg Watson: I think everyone who—
Mr. Stuart: Greg is on the high table.
Greg Watson: Bearing in mind that there are over 100 universities in this country and 50,000 different degree courses, I am not surprised that different universities and different departments in universities come to a particular view on what is the most appropriate basis for being prepared for their courses. It would be pretty extraordinary to imagine that a maths degree would be available to someone who had not done maths A-level. That is perfectly proper, and perhaps that is a question for the universities to answer, rather than us. Our job is to run A-levels to a common standard, right across the subject range, and to make sure that the A-level is worth the same, as a qualification, regardless of which subject you do it in. What use universities subsequently make of it is down to them trying to target the right kind of young people to get on to the right kind of courses.

Q184 Mr. Heppell: I was thinking about common qualifications. I understand that some universities differentiate the courses and say, "We will take the OCR one as one we accept but we won't accept another." Does that happen? I thought it did.
Greg Watson: Not to a great degree that I am aware of.
Jerry Jarvis: If it does, it must be very limited. There are areas of the country where there are affinities between certain universities and certain exam boards. There are traditions, and there are preferences in the selection process, but I cannot believe that it is widespread. It has not come to us.

Q185 Chairman: What does Murray think? One thing that comes through from reading all this stuff, as we have got into the inquiry, is that even if a parent does not remember their A-levels and GCSEs very well, they remember their degrees. I was leafing through my examination results from when I was a small person right through to my postgraduate time. I looked at the year in which I got my degree from the London School of Economics, and eight people got firsts throughout the whole LSE, in all subjects. That does not happen now; the figure would be one third. It is not your area to set degree papers, but surely you must be worried about that. Does it not raise a big question about whether a first is good as it used to be, if a third of people get firsts compared with a small number not very many years ago?
Murray Butcher: By definition, yes. You are either talking about an elite, or you are not. I spent many years as an employer of degree-holding candidates, and the university from which the degree is gained and the subject always creates a choice for employers. Gosh, we are moving off the subject.

Q186 Chairman: No, I think it is very relevant, because what part of testing and assessment is there to try to dispel this view—if it is wrong among parents—that standards have got easier? I pitched to you the idea that in degrees that it is the case, with so many people getting firsts compared with lower and upper seconds.
Greg Watson: I honestly do not think I could comment, not being involved in university-level assessment.

Q187 Chairman: That is why I was going to ask a second question. What I am getting from you is that you do not really know because you do not have that much of a link with universities. Surely, part of your job, and one criticism of the examining boards, is that you do not any longer have enough of a relationship with the teachers who teach the subject, or with the people who take the products after they have taken your examination—that you have become rather isolated, both from the teachers below and from the university teachers above. That surely cannot be good if that is what people are saying about you.
Dr. Bird: Is that the evidence you have had presented? I would certainly reject that on behalf of AQA. We have active teachers and head teachers on our governing council all the way through to people involved in our subject committees. They are active teachers who act as senior examiners, preparing and developing our material. From the classroom perspective, we have got that active involvement with the chalk face—the whiteboard face as it now more correctly is—and HE is represented from admissions and senior academic perspectives on our governing council.

Q188 Chairman: So most of your examiners will be teachers, will they?
Dr. Bird: Yes.
Jerry Jarvis: All of them.

Q189 Chairman: All of them?
Jerry Jarvis: Decisions in all our awarding bodies are taken essentially by practising senior teachers—from the grading decisions to the writing of examination papers and so on.
Greg Watson: I am actually quite concerned about what you say, though, because we could not possibly compete if we did not have an intimate relationship with people who are ultimately our customers—the takers of our qualifications.

Q190 Chairman: Any hints that you would knock that to one side?
Jerry Jarvis: Certainly, in the development and creation of curriculum materials and in supporting a teacher in delivery, I cannot subscribe to that view. However, remember that we are required to demonstrate rigour, and part of that demonstration necessitates distance. I could well understand that when it comes to a disagreement over an examination outcome, we have a role to play that is authoritarian, and we must maintain some rigour and some distance from the process of learning.
Greg Watson: As awarding bodies, we occupy a unique position. We sit in the middle of lots of different stakeholders in the qualification system. On the one hand, we have schools and colleges, teachers, young people and parents. On another side, we have subject groups, subject associations, leading thinkers in a subject and groups in

universities that do research on particular subjects such as the Salters group at York. On the other hand, we have the Government of the day and their political drives of various sorts, and, of course, we as a charity also have our own education mission. We have a lot of experience—150 years—and a lot of research to tap into. They give us good clues about the right direction of travel. Inevitably, sitting in that situation, we do not please everybody all the time, but we do play an important role in drawing on all the different views and trying to square the circle in a way that I do not believe anybody else in the system does.

Chairman: Because I was following the run of your answers, we have been stealing Lynda's questions. We are over to Lynda now.

Q191 Lynda Waltho: I was actually going to extend the question, although I wanted to deal in particular with differentiation. What input do you get from employers? There is a general feeling, certainly within the trade press—specific evidence has not been given to this Committee—that there is not sufficient input from academics and employers, particularly in respect of the closed questions that make up a large part of exams. They may be deskilling our school leavers, and not expanding their analytical skills. What input do you get from employers, who are receivers as well?

Greg Watson: A couple of comments. First, there has been direct input from employers in respect of many of our longest standing and most successful qualifications. It is still the case that the single most heavily used qualification in the history of this country is an adult IT qualification called CLAiT, which was developed almost 20 years ago in direct response to trends in the employment market and signals that were coming from employers and further education colleges speaking on behalf of employers. More recently, very successful new vocational programmes called OCR Nationals have been rolled out to schools. We built those—again, sitting in that unique position that I mentioned—listening, on the one hand, to schools and colleges and what they hoped they could offer to young people, and, on the other hand, to employers and their views on what was relevant to modern employment. I think that we do play that role. That said, I spoke in my opening statement about the impact of the past 10 years and the growing use of qualifications as a public policy tool—perhaps more change has been driven directly by public policy. In my position sitting in the middle of everybody, I would say that in the past 10 years we have probably been drawn closer to satisfying the Government's direction of travel, and the trade-off has been that we have given less time than we would have done 10 years ago to consulting directly with universities and employers. There are only so many hours in the day, and only so much time available to develop new qualifications. I hope that, with the arrival of an independent regulator, there will be the opportunity to rebalance that slightly, because there are times when I would like to be closer to employers and universities in developing some of our qualifications.

Q192 Lynda Waltho: That is the general view?

Jerry Jarvis: I would endorse to a large extent what Greg said. Half of Edexcel's provision is in the vocational workspace with BTECs. I believe half a million students are doing them. Those qualifications would not work if they had not been developed in conjunction with employers. They are vocationally oriented qualifications for the marketplace. The degree to which employers play a part in the setting of criteria for what is taught at GCE and GCSE is open to question. Examination organisations such as ours are required to create qualifications that conform very closely to criteria set down by the regulator. There is always a case for better collaboration with all stakeholders, including parents, in the putting together of criteria. We are all constrained by, and must work within, the criteria that are set down but, certainly, when you move away from GCE and GCSE, the degree of employer collaboration is massive.

Murray Butcher: My answers would be from a vocational context. Employers can contribute in two ways. First, they can contribute through their role on sector skills councils, in which the creation of national occupational standards takes place. They lead to the creation of national vocational qualifications. Secondly, they can contribute through the provision of supervision of that activity. The key person within vocational qualifications is the external verifier. He or she works to a particular awarding body and visits locations to ensure that practice accords to standards. They may be practitioners from further education, or people drawn straight from employment. There is considerable opportunity for employers to play a role, but I concede that it is quite difficult on occasions to gain such contact with employers because, naturally, they see their principal activity as earning their particular income from their role—finding time to engage in education can be quite difficult for them.

Q193 Lynda Waltho: That finishes that part, but I would like to extend the question and to look at the differentiation of students for selection purposes. Certainly, top universities argue that too many applicants attain the top grades, which often leads to more testing. The Committee has heard that high-stakes testing and the need for schools to show well in league tables has resulted in teaching to test. What are your views on that? Does the league table culture in schools distract teachers from the task of preparing school leavers with the deep knowledge and the independent analytical study skills that are necessary for higher education and higher training?

Jerry Jarvis: Let me take the first question on whether testing has an effect on the behaviour in the classroom. There is huge pressure on schools and teachers to increase performance. When we have one of our regular focus groups, I sit with a panel of teachers and ask them to introduce themselves. Almost without exception, they tell me their name, what school they are from, the subject that they teach, and the pass rate in their subject. Unfortunately, there are huge issues at stake in most

schools, and teachers are human. Having said that, the huge overwhelming majority of teachers aim to deliver on education—that aim comes across strongly in what they do. However, there is no question that there is pressure. We are talking about teaching to test. If we can write examination materials that cover the whole syllabus in a way that means that people cannot predict how something will be questioned or what the questions will be about, it would be necessary for schools to teach the whole syllabus. When I came here today, you very kindly issued me with some instructions and notes about what was going to happen today; you made me feel comfortable, and you gave me a brief outline of the lines of questioning that I would be given. You were either leading me and helping me to respond, or you were just trying to set down the rules of engagement. Providing we stick to the notion that we are talking about the rules of engagement, surely we are right, but the pressure is massive.

Dr. Bird: I think the pressure is massive. My concern as a former sixth-form governor is that it is too much about output measures and not about added value. There is so much in output measures that is to do with the inputs that the students bring with them and less about what the school adds through its teaching and learning processes. So there is a place for more sophisticated league tables if they will be used to assist schools in improving. As Jerry said, we aim to provide a broad curriculum. We aim to test all of the student's skills. We are a strong supporter of the introduction of A* at A-level. We certainly believe that there is enough evidence in the current distributions of marks that that could be done from the current exam papers, if the Government were minded to do so. We feel that it is a cause for celebration that more than 20,000 people are getting three As. It is a very small proportion of the total candidate group and it is a very small number of university departments that have a problem with that. We appreciate that it is a problem, but those departments that select have known for many years that it is not just three A-levels that you need to be a good vet or a good lawyer—you need a range of other things—and they go out of their way to establish the other skills that good sixth-form curriculums provide, such as extension work, project work, community work and work experience. Those things can also be certificated. We have a qualification that we are trying to get approved that would support the bundling up of those activities.

Q194 Lynda Waltho: So it would be a combination of those extended activities plus A*. What do your colleagues think about the A*? I would be quite interested to hear that.

Greg Watson: The A* is the right development at the right time. Andrew has mentioned that the number of candidates getting three As at A-level is small; we are talking about 3% of the 18-year-old age group. So it is still a pretty tall order to end up with three As. Nevertheless, there are now enough young people in that category that certain universities are saying that that achievement is not enough to differentiate. As I said earlier, I think that we have made a bit of a trade-off in the evolution of A-level, between broadening it as a general qualification and maintaining it as the most stretching assessment for getting into university. I think that reintroducing an element of stretching in each subject and marking out an A* grade as a higher bar for the best candidates to get over is a perfectly sensible thing to do, and I know that a lot of universities will welcome that development. I think that the big pressure in this area is the GCSE A* to C measure. When Jerry is talking about the teachers round the table, that is the thing that I know drives a lot of institutions. That is probably the greatest risk of creating any distorting incentive and of distorting things in two ways potentially. One is that I think that if I were in school today, I would rather be a D-grade candidate who was very close to the C grade boundary than a D-grade candidate who was at risk of ending up in the Es. There is always the danger that the last half an hour of the teacher's attention will inevitably be drawn towards the candidate who has a good chance of getting over the C grade hurdle. The second issue is that, as the curriculum evolves and we have new qualifications coming along, the pressure to find new qualifications that can be treated as equivalent to a C grade at GCSE can itself become a distorting incentive in the way that those new qualifications are developed; you can see that from much of the public debate about Diplomas and how they are going to be treated. Attention could become concentrated on that issue, rather than on whether it is a good curriculum innovation and will be a good new offering for young people.

Lynda Waltho: That is fine. I think that Dawn wanted to come in on Diplomas.

Q195 Ms Butler: I just wanted quickly to touch on Diplomas, because I know that Graham is going to talk a little about them; he thinks that I am stealing his thunder. Do you think that the new Diplomas will help to bridge the gap between school and higher education? We know that young people who come from more affluent families already have this type of mapping; going to school, then going on to higher education and going to university. Will Diplomas help to bridge that gap?

Jerry Jarvis: One thing that the Diploma may do is broaden the experience and learning of students who enter university. They might have taken a much broader, different curriculum, and perhaps one that was closer to practical learning than previous syllabuses were. There is an opportunity there, although a huge number of students enter university through BTEC nationals and, by proportion, do very well in taking degrees. I am interested in your identification of the gap.

Q196 Ms Butler: As I was saying, we are trying to inject aspiration into the learning agenda. We want to give those who might not have thought about going on to higher education the opportunity to do so, as the BTEC does. The question is partly how we can ensure that employers take Diplomas seriously. Also, how can we ensure that a Diploma is equivalent to three and a half A-levels?

Jerry Jarvis: We have to work very hard to have that Diploma earn its spurs and get it the reputation that it needs. It must cover a great deal of ground, and it must be valued by everyone, not just those who would not necessarily have considered themselves able to go into higher education. It has the potential to offer a very different form of learning on the way to either higher education or employment. The Diploma is a big ask, and no subject is dearer to us right now than to make it work, as the first teaching starts in September.

Chairman: We are deep into Diplomas now, so I shall share the Diploma questions around. Fiona first.

Q197 Fiona Mactaggart: Greg, in your evidence you quoted a delegate at a conference that you organised, who said that the Diplomas were examination officers' worst nightmare—a telling phrase, not usual for evidence to a House of Commons Select Committee. You clearly feel strongly about this. What is nightmarish, and what would you have done differently?

Greg Watson: Just by way of explanation to members of the Committee—I am conscious that our sector, like many, is full of all sorts of strange language—exams officers are the administrative hub in a school or college, responsible for making exam entries, ensuring that results are dished out in time and so on. I mentioned Diplomas because they are the most complicated qualification that I have ever seen. A typical learner is likely to take part of their Diploma in a school and part in a college, and they will do different elements of the Diploma with different awarding bodies. The exams officer will be faced with a brand new IT system built at great speed, and just about in time, with which to administer all this, and any given exams officer will probably have to work with an exams officer in at least one other school or college to ensure that they use the same reference number for a candidate and keep track of what a candidate has done so far. They must also know exactly what the candidate still needs to do to complete the qualification. Running the exams office in a school or college is a tough job anyway. It is often a temporary or part-time job, and has a churn rate of about 50% of individuals in any given year. In many places the exams officer struggles to have authority with the senior management team in the school or college to get things changed. We need to recognise the complexity of the qualification in its own right, the practical logistics of how it will be taught and the fact that it needs to be sorted out in pretty short order for the first learners, who will be going through the programmes in September 2008.

Q198 Fiona Mactaggart: That comment is about the administration of it; let us consider the assessment of it. Murray, most of your assessment experience is in the thing that is worrying us in some ways about the Diplomas—whether the person can do the extended portfolio. Do we have people who can carry out that quality of assessment? Will it work?

Murray Butcher: At this stage, a great number of questions remain on the form of assessment that will take place within the various themes of activities, from the principal learning and the additional specialist learning to the extended project and the functional skills. We are likely to find a range of assessment practices that will cover all of those. Some will be internal, some external, some will be moderated and some will be verified by the awarding body. Even though we are planning to release the first teaching in September, there is still quite a long way to go, and quite a lot of discussion to have with the regulator to agree on the types of assessment that will take place. I believe that quite a bit of responsibility will fall on the schools and the delivery consortiums, and funding is already in place to support that. It will be fairly heavily pressurised in order to ensure that the teaching staff at schools and colleges and even the local employer have the necessary skills. It is a very big question.

Chairman: May I ask everyone to be quite short? My colleagues are tending to ask all of you questions on each section.

Dr. Bird: Very quickly, I agree with all that Murray said. I would just add that none of those means of assessment are new. We have experience of using them in centres now. We have lots of experience in training teachers to be assessors; we have done it for many years. It is a big ask and a lot to do. That is why we are encouraged by the Government's process of controlling volume in the first few years with Gateway centres and so on and by the money that they are providing to help with that verification and assessment development process.

Q199 Fiona Mactaggart: I was interested in the bit of your evidence in which you said that coursework had changed over time and that assessment used to be more embedded in courses. You also said that the development of extended portfolios had created greater opportunities for plagiarism, and that some of the risks that people have suggested has occurred with coursework. I hope that I have not misread you. I wonder how you assess the extended project element of the Diploma and whether such an element holds the same risk.

Dr. Bird: The extended project is a piece of coursework that is chosen by the student and arises out of their other studies. It is not a subject set by us, but it is approved by us. It then requires a student to do a piece of work and present some material. Clearly, that involves them doing research and using the Internet and so on. By its very nature, it is a unique piece of work that is based on a student's interests and the A-levels that they are doing alongside the project. The chance of copying great chunks of it, therefore, is massively reduced. That was our point. Old-fashioned coursework embedded in the curriculum arising out of a student's experience requires higher quality assessment skills. However, because it is different for every student, such work reduces the chances of plagiarism being a serious problem. With coursework in general qualifications, we have reduced its variety, standardised it and made it a task. That task,

therefore, is difficult to differentiate, which means that you are looking at lots of very similar things. The chances of people sharing that and copying are therefore increased because it is the same piece of work. The thing that we find disappointing about coursework in general qualifications is that just as a number of us are launching the portfolio products that allow students electronically to publish evidence of their work and have it assessed remotely, which means you can provide photographs, audio clips, short videos and texts, the system is closing coursework down and making it a fairly repetitive and limited task. The technology enables things to be checked, which, we think, would allow coursework to flourish. We hope that that will be the case in the Diploma environment.
Chairman: Very interesting.

Q200 Fiona Mactaggart: I have heard that administration is hard, and that we need particular skills to assess the can-do bits of the coursework of the Diplomas. Do you think they are going to work? A yes or no answer will be sufficient.
Dr. Bird: Yes.
Murray Butcher: Yes, we are working very hard to ensure that.
Jerry Jarvis: I would rather come back with the statement that they have to work. We are making a huge investment, and we have made a very bold decision. We all feel exactly the same way. We all suffer from trepidation in a whole series of different areas, but we have to make this Diploma work.
Greg Watson: I will say yes, with one important proviso. I return to an earlier question about whether employers and universities will really feel that the Diplomas are worth while. The answer is that they will see the proof with their own eyes, because they will get to know young people who have been through these programmes. Whether the Diplomas will succeed in the early years will be more about how they are taught, and then about the kind of young people that they produce. Regardless of all the technicalities of assessment—and there will be issues that we have to iron out over time—if the Diplomas are really to succeed in the first two or three years, it will be all about the teaching. That means that support for teaching has to be exemplary and that we should be sensible about the number of students that we want to see on those programmes early on. As Andrew said, it has been helpful that the Government have put some criteria in place. If the numbers do not grow too fast, and we devote an extraordinary effort—not just the Government, but people like us—to supporting teachers in the first few years of teaching, Diplomas will succeed.
Chairman: Murray, may I ask you to be one of the first respondents to Graham?

Q201 Mr. Stuart: Fiona has given you a pretty good work through. We ended there on a positive note that, "We are going to make it work because we've got no choice." City & Guilds has the greatest experience of vocational training; may I put it to you that there has been insufficient time for component awarding bodies to research and develop new assessment approaches? Are Diplomas being introduced to a political, rather than educational, timetable?
Murray Butcher: Timetable, perhaps. In terms of trying to draw—

Q202 Mr. Stuart: Is that a yes or a no? Would any one of you, or anyone whom you know in the educational establishment, have gone for this timetable if it had not been dictated by the Government?
Murray Butcher: I think I would need to collaborate with my colleagues on either side to reflect on how some of the Curriculum 2000 qualifications were introduced. That was also an extremely short time scale, but it was achieved. It is possible that we can achieve this, but some very concentrated thought is needed. It is, in historical terms, more of a political timetable than a curriculum development one—that is true.

Q203 Mr. Stuart: Yes; it comes in in a few months, and you say that there is a lack of clarity over the purpose of the qualification and its underpinning curriculum.
Murray Butcher: Yes—

Q204 Mr. Stuart: That is pretty damning is it not?
Murray Butcher: There are considerable concerns—if I can explain. As a vocational awarding body, we see this as an opportunity to introduce a hint of vocationalism into the school and post-school curriculum. We carry considerable anxieties that, as we go through that process, the pressure on the award may limit the amount of vocational experience available and that the award may drift back to the standard, general qualifications—much as we believe GNVQs did. My anxiety is that if it is successful, this qualification will also be a challenge for higher education, because I believe that it will draw in a slightly wider cohort than has been experienced up to now. There is a lot of pressure on HE to increase its intake, and it must be borne in mind that the nature of the intake, as well as the skills and approach of young people going into higher education, may change. There will need to be a response to that.

Q205 Mr. Stuart: The advanced Diploma is supposed to be worth 420 UCAS points, whereas three As at A-level are worth 360 points. Is that realistic? Will it carry public confidence?
Murray Butcher: I do not see why it should not. We are devising a new qualification that is trying to attract young people with a range of skills. I would not wish to compare the skills of this young person to that young person. If they are successful in the award, they deserve the three or three and a half A-levels that they achieve.

Q206 Mr. Stuart: People have had theories over the years and they have always come up with conceptually excellent ideas that, like GNVQs, end up not working. The two biggest challenges, according to *The Guardian*, are university

endorsement and parental endorsement. Is not the real challenge in getting the new qualification going that you have to get enough good pupils and good teaching to get it off the ground? The tendency is that a qualification comes in and it is the poorest cohorts at the poorest establishment who are often the first adopters. Is that not a danger? How can we prevent it from happening, so that the Diploma can flourish?
Jerry Jarvis: You are absolutely right. I made the point earlier that the reputation of this qualification will be critical to its success. The programme of introduction is too fast. There is too much involved in this programme, and too many other parallel changes are going on at the same time. We would not have done it this way, but we have made a commitment. We have had our arguments with our regulator, with the Department, with each other and inside our own organisations, and we have made a commitment, so there is a point at which we have to say, "We wouldn't have started from here, but we have to make this work." So much is at stake. The ultimate success of the qualification—

Q207 Mr. Stuart: What if we had had you here when we were talking about GNVQs? We can look at every previous initiative, because we have always recognised we do not have the right system for vocational education. If we had had your equivalents in the past year, you would have all said the same thing, wouldn't you—about the need to make it work? And they have not worked. I am trying to tease out just how serious the problems are with the lack of clarity over the purpose, the timetable, and the risk of undermining GCSEs and the AQA coming in.
Greg Watson: I am not sure I necessarily agree with the statement that GNVQs did not work in the end. Through a process of iteration over about 10 years, GNVQs became confidently used in some schools and colleges, but it took 10 years. When the Education and Skills Committee looked into Diplomas—we offered evidence about the speed—there was, I think, recognition in the Committee's report that going at this speed does not help and putting pressure on numbers early on does not help. I mentioned OCR Nationals earlier. When we were developing those, we probably took a year longer than we took over Diplomas. That gave us a year to triangulate more effectively between what employers wanted and what schools and colleges thought they wanted, and to look hard at assessment structures and ensure that grading arrangements and the like were going to work well. I think we would have all wished for a little more time. On a positive note, there has been an excellent dialogue between ourselves, the Department and the QCA about the first phase of Diplomas and how that process works. I think everybody has recognised that it was very loaded towards the employer view and we did not take on board enough about teaching and pedagogy and did not, early on, spend enough time on the business of assessment and standards and ensuring that the outcome of the Diplomas would command confidence. There is a commitment all round, among all the people I have just talked about, to go about the second phase in a rather different way. That in itself shows that we are perhaps recycling the learning faster than we might have done with the GNVQ.
Chairman: There is a bit of pressure on time. We have to move on to coursework.

Q208 Annette Brooke: You have already hinted that there are quite a lot of advantages in using coursework in the examinations system. Given the rather dramatic moves, particularly with the GCSE, is this a question of throwing the baby out with the bathwater?
Dr. Bird: In my early remarks about coursework and the extended project, I was, I suppose, broadly hinting in that direction. Very few people in their daily lives do not discuss with colleagues an issue they are trying to tackle, or use the Internet or any of those wider research skills. Clearly, properly constructed coursework questions allow people to demonstrate, in preparing a document for moderation, those wider study skills and extended communication skills, which it is very hard to assess in a quite short, formal examination.

Q209 Annette Brooke: I shall ask another question and perhaps somebody else can pick it up. Is it simply that you guys have not been smart enough in setting the coursework?
Chairman: Who is that question to?
Annette Brooke: I am looking for volunteers. I though that it would be a bit unkind to pick on somebody.
Chairman: Jerry can go first, followed by Greg.
Jerry Jarvis: I think that you have two volunteers here. It is probably a case of throwing the baby out with the bathwater. It is not that difficult to detect plagiarism. After all, the same mechanism that allowed the student to find the material to plagiarise makes it just as easy for the moderator to find it. We believe that coursework makes a hugely valuable contribution to the way in which we conduct assessments. It is a far richer method than an end-of-term examination. Last year, Edexcel did a huge amount of work on detecting cheating throughout the UK. We examined all sorts of issues, but detecting cheating in coursework is easy, actually.
Greg Watson: Yes, I am disappointed with our direction of travel. Hopefully, some of what we learn from the extended project within the Diploma will flow back into coursework. Ironically, the extended project could prove to be one of the most exciting parts of the Diploma development. It could be a genuinely personal and independent project that would build evaluation, research, thinking and self-organisation skills, which are exactly the sort of skills that coursework—done well—has the power to develop. Coursework is not right for every subject, and in some ways we have suffered from having veered from one side of the road to the other. At one time, we decided that coursework was a good thing and, through over-mighty regulation, we were compelled to put coursework into subjects such as maths, where it does not sit readily. There has been a sudden change of tack and now we are compelled

to take coursework out, even when, in subjects such as geography or history, it can breed fundamental skills that, incidentally, carry on very well into higher education, too.

Murray Butcher: I was just reflecting on some good, simple assessment reasons for maintaining coursework. We are trying to get as broad a picture as we can of individuals' abilities. If you take out coursework, you are focusing on just one or two other forms of assessment, which gives you a biased picture of the individual. I share the view that by withdrawing coursework, we would be going in the wrong direction.

Q210 Annette Brooke: On gender balance, I have the impression that girls started performing rather well in some GCSE coursework. Is that statistically true? I suppose that I should look to the awarding bodies for that information, but I thought that you might have the gender breakdown of results with and without coursework.

Jerry Jarvis: We could provide it but, anecdotally, it is generally the case that girls do better in coursework, as well as the other subjects, although that varies enormously in coursework subjects. However, yes, they are better at coursework.

Q211 Annette Brooke: I find that interesting. I wonder whether getting rid of coursework would solve one of the Government's problems by potentially closing the gender gap—I accept that that is rather cynical. Something that concerns us greatly—this might be in line with our questions and other lines of inquiry on social mobility—is that young people from different backgrounds and with differing levels of family support clearly will have different opportunities within their coursework. For instance, I recall that the Prime Minister's wife could get help with certain coursework, and I used to take my children to museums and so on. How do you factor in those things for a true assessment?

Chairman: You cannot, can you? Is it fair? Is it biased towards middle-class kids? Do less privileged kids find it harder to achieve?

Greg Watson: That is very hard to iron out. However, returning to something that a couple of us have said, the notion of a truly personal piece of learning could get over that problem, because it would be in the bounds of the imagination and creativity of the young person and their teacher. I can think of good examples in GCSE business studies, in which coursework is open-ended. A student goes to a local business, works up a project on one aspect of the business, and writes it up. A teacher who encourages young people to think imaginatively, or a school that maintains good links with local employers, is able to do that, regardless of any advantage or disadvantage in the home background. Part of what creates the difficulty is the tendency, through regulation, to drive coursework towards being a standardised task. If anything, those other advantages kick in to a greater degree when that is the case.

Chairman: David, do you want to come in on coursework?

Q212 Mr. Chaytor: Yes. Specifically, what are the most effective means of preventing plagiarism in the first place? I take the point about making coursework more personalised, as against standardised, but how do you prevent someone from accessing the Internet or from getting a disproportionate advantage from having well-informed parents?

Jerry Jarvis: The view that we have taken is that the best way to prevent plagiarism is the likelihood that your efforts will be detected—it is like the idea that you will probably drive slower if you suspect that there is a traffic camera around. The best way to do that is to sample students' work. As I said earlier, moderators and examiners can use the same methods to access non-original material as the students did. These days, even with material that is being shared among students, which might not be generally available, we have deployment technology and we digitally scan everything that we do, so we have the ability to have a machine literally looking for similarities that could not otherwise be detected without the deep reading of work. For me, the way to discourage plagiarism is through the probability that you will be caught.

Chairman: Andrew, do you want to come in on that?

Dr. Bird: I agree with some of that. I think the real way of inhibiting plagiarism is to make sure that you have set the right task. If we set a task that is basically a knowledge task, we are encouraging somebody to go to the Internet and find out all they know, so we need to set a task that requires them to acquire some knowledge, from wherever they get it, and then to do something with it—to process it and convert it into something that is about them and their insight into that knowledge. The task needs to be about underlying ability, rather than a recounting of what they can find in a book or on an Internet site. Tasks that require people to give empathy of analysis are the sort of coursework tasks that we ought to be setting. That, in my view, is the strongest way of avoiding plagiarism, followed by the sense that there is a speed camera around the corner.

Q213 Mr. Chaytor: City & Guilds is particularly keen on the role of coursework, and local assessment of coursework is more frequent in vocational qualifications than in more academic qualifications. What are your views on all that? Do different standards apply to courses that are defined as vocational, as opposed to those that are defined as academic?

Murray Butcher: There are similarities and differences. Within a National Vocational Qualification course, you are dealing with the fact that the curriculum contains certain criteria. When you come to assess what, in vocational qualifications, is called the portfolio, which is a collection of evidence, you are testing that evidence against certain specified criteria and looking for the supporting evidence within it. That does not mean to say that there are not occasions when you find that individual students and trainees come up with extremely similar evidence. Supported by regulation, we must carry an investigations team

within City & Guilds so that whenever questions occur that might suggest plagiarism—or not so much plagiarism as the possibility that the centres providing the documentation are not encouraging individuals to do things themselves—we need to explore that and confirm whether something untoward has taken place. So, in a sense, plagiarism is possible in NVQs, and we have to make sure that we eliminate it.

Q214 Mr. Chaytor: Broadly, what proportion of awards are rejected because plagiarism has been identified by City & Guilds and the other boards?
Murray Butcher: Within City & Guilds, it is going to be an extremely small percentage—less than 1%.
Greg Watson: For OCR, 192 candidates in 2007 had sanctions applied for plagiarism.

Q215 Mr. Chaytor: As a percentage of the total?
Greg Watson: Minute.

Q216 Mr. Chaytor: Broadly, on the use of the Internet in coursework and to support learning for external examinations, is there any evidence that the digital divide between families is leading to a widening gap in the achievement of young people? Is access to the Internet a factor that drives up standards very quickly for the most affluent families, leaving the children of poorer families behind? Is anyone doing any research on this, or do you have any thoughts about it?
Greg Watson: That is actually a very difficult question to answer from where we are sat. We assess what is put before us. Given that most of it arrives in a digital format, you could say, *prima facie*, that there is no obvious divide, but I cannot gauge—I am not sure that I even want to—who has used what means to get there. We simply mark it all to the same standard when we are presented with it. There is an interesting research subject there, but I cannot comment from the evidence that we see.
Jerry Jarvis: Following on from an earlier question about setting an examination to ensure equality of access, particularly for disadvantaged children, we do quite a lot in setting all our examination and assessment work on the more obvious things, such as language and culture. We would not want to set a paper in geography in which someone who was a non-UK resident would be disadvantaged, so we think very carefully about that. We would not require someone to do a field trip to France, so we can make sure that when we set our assessment instruments, we do not. However, I am afraid that one or other of the population will be inadvertently disadvantaged. Geography comes in as well, with remote villages and so on where access to appropriate research facilities is limited. We have to be careful but, inevitably, some students have greater advantages than others.

Q217 Mr. Chaytor: May I move on to the QCA? All four boards agree that the changes to the QCA are welcome. That is right, is it not?
Dr. Bird *indicated assent.*
Greg Watson *indicated assent.*
Jerry Jarvis *indicated assent.*
Murray Butcher *indicated assent.*

Q218 Mr. Chaytor: But what has it done wrong? Where has it gone wrong?
Chairman: And why were you not lobbying for the change before?
Greg Watson: I was.
Chairman: Oh, you were. Okay.
Dr. Bird: We were as well. The QCA has not done an awful lot wrong, but having the two activities together can generate a conflict of interest, and separating curriculum development from the regulation of assessment can only be a good thing. The regulators spent a lot of time thinking about the development of the new A-levels, but never mentioned what it would cost to assess those various models of assessment. Sometime later on, QCA quite properly wanted to explore our pricing strategy for those A-levels—after we had been in development for some time and the die had been cast. That is the very worst set of arrangements, because we are already committed to a style of assessment in association with the curriculum development side of QCA at a time when the regulatory side started talking about pricing, which is influenced by what it costs to deliver the style of assessment. Separation of the two roles should lead to a more robust dialogue between the regulatory side and the development side so that we triangulate the issues more effectively, because the consequence of development requirements on us changes what we can deliver and what it might cost to put it into the marketplace. One can see a more robust triangular relationship across the industry, which is why we favour it.

Q219 Mr. Chaytor: The example that you have given could equally be used as an argument to scrap the curriculum development agency function completely and devolve it to each of the four boards.
Dr. Bird: It could be—
Mr. Chaytor: There is no guarantee that a separate development agency will also not overlook the practical implementation costs of the financial arrangements
Dr. Bird: I agree, and I think some of us might say that that might be quite nice, from the innovation perspective. Certainly we would encourage the development side of the organisation to try and be strategic and output-driven, rather than detailed and prescriptive, in its future thinking. We do suffer from the very detailed and prescriptive approach to what A-level English looks like, which does not help innovation across this table.
Greg Watson: I think, to go right back to what I said at the start this afternoon, that managing a large and high-stakes qualification system means constant trade-offs between change in order to keep qualifications relevant, to respond to changes in the economy, and to adapt to the different needs of higher education versus stability, which is the thing that builds public confidence and familiarity. I think

that a good regulator would constantly be balancing those two drivers. I think that QCA, because of the position it has occupied very close to Government, has tended to find that its role in being a sponsor of change has far outweighed, over time, its responsibility for stability. I hope that the one difference that we will see with the new regulator is a greater weighing of the pros and cons of change at any given moment, whether it is at a system-wide level in introducing a whole new qualification, in the form of Diplomas, or at an individual qualification level, in deciding whether to have more or less coursework in GCSEs. I think that that would be a very positive development.

Q220 Mr. Chaytor: Greg, in your submission, you were critical of the QCA for intervening in too heavy-handed a way, whereas in the AQA submission there were accusations of not regulating strongly enough in respect of the current basic and key skills tests. It said that the problem is that the regulation is weak. Where is the balance? Is it that the QCA has been too interventionist, or has it been too hands-off? Each of you says different things.

Dr. Bird: I think we were saying that it was conflicted because it was the provider of the tests. It was not a matter of being too hands-off; it was the regulator of itself. That was our point. Those tests are derived—we only deliver them—by QCA itself, so it sets the standard.

Greg Watson: I would say that it has been a case of too much in some areas and too little in others—too much intervention at the detailed level, getting in the way, I think, of producing more stimulating approaches to geography, more relevant vocational offerings, but on the other hand too little attention to the long-run effects of a series of changes piling up. If you were to look at A-level over time and take modularisation, the move to six units, the balancing of coursework, the splitting of AS and A2 into two lots of 50%, compounded with various tweaks and twists applied along the way; if the regulator is looking after public confidence it should be less concerned with any one of those changes in any one subject and more with looking at the long-term effect over time on actual standards and on public perceptions of standards, and therefore, with measuring the rate of that sort of change in a more moderate way.

Murray Butcher: The opportunity for the independent regulator may mean that QCA, or whatever it becomes known as, may be able to focus on regulation and refine that process of regulation, which I would say, from City & Guilds experience, has been delivered in fits and starts. There has been urgent attention to this activity, which then suddenly dissipates; then it comes over here somewhere, and gets distracted by some of its curriculum activities. I really look to this as an opportunity to have stable regulation. The one principal issue that we still have, which Andrew mentioned in his opening statement, is the fact that we have four regulators to deal with, and that is a big bureaucratic problem.

Q221 Mr. Stuart: Following on naturally from that in terms of maintaining standards, should it be an important part of the new regulator's job to look at comparisons internationally? Sir Peter Lampl said, of standards, that "the key question is are they improving fast enough to maintain or improve our position relative to other European countries, America as well as China and India?" Do you agree with that?

Jerry Jarvis *indicated assent.*

Greg Watson: Any sensible regulator in this field, with a primary focus on standards, will want to base its decisions on evidence. That seems a strangely obvious thing to point out, but it is a point worth making. There is good evidence abroad. As I mentioned, we are part of a group operating in 150 countries. It is partly through the process of synthesising research from experience in different countries that we can draw some of the conclusions that we can about the impact of change on standards and so on. I would have thought that any sensible regulator would want to build relationships around the world with all sorts of research communities.

Q222 Mr. Stuart: In the context of today's evidence, what do you make of this country's tumbling down the international league tables that came out recently?

Chairman: Do not roam too widely, but comments would be appreciated. Are we tumbling? Is that your fault? Is it just a rotten examination board system causing us to tumble down the international league tables?

Jerry Jarvis: First, that is a very difficult question to answer here. We should also be very careful about what we compare. When Leach looked at the various standards in different countries, one of the conclusions that we started to come to was that we probably do as much training, but that we do not certificate it in the same way, so we do not count it. Counting eggs and counting eggs is an important issue. We certainly share a massive international business with OCR, and this education system still has huge respect abroad. We cannot be complacent but it can be very dangerous to build any sort of policy on an international comparison because of the sheer difference of literally comparing the same thing.

Chairman: We must move on to the last section.

Q223 Mr. Stuart: May I just deal with the key stage and testing at different times? Edexcel's evidence was that that will increase the testing burden and that the Key Stage test should be at the end of the key stage, if retained. Do you think that if we are to have tests when the pupil is ready rather than at the end of the stage, we need to scrap Key Stage tests as such?

Jerry Jarvis: Forgive me, Barry, I need to make a slightly longer response to this question, because it is something I feel strongly about. We constantly accuse ourselves of over-testing, but summative, quantitative testing needs to be done to ensure that the investment that we put in, and the teaching and learning that is happening, happens. Let me return to what Ken Boston said—three things: personalised

learning, continuous—essentially, internal—assessment and training of teachers. The reason why I care so much about it is that after about £80 million of investment, we can do it now, and if we really were interested in shifting that stubborn set of five good A to C grades, we could do it now. It is here, and Ken Boston was absolutely right. So, do we over-test? Yes, in certain areas, but we do not spend enough time on personal, continuous improvement, which is the key to improvement.

Chairman: Okay. We must move on, and I want to cover the last section. We have the duo of Fiona and David covering it, with Dawn throwing one matter in if she wants.

Q224 Fiona Mactaggart: We have had quite a bit of evidence about the narrowing of the curriculum and teaching to the test, as you would expect, and in a way, some of your evidence is specifically about that. I am very interested, for example, in your analysis, Andrew, of why aptitude tests are not an appropriate alternative because you can train young people to succeed in them—a fact I know very well, representing a town that still uses the 11-plus. Children of 11 perform less well on Key Stage 2 assessments, I think, because they are training for their 11-plus aptitude tests. I think that I am hearing from you that in some ways it is a good thing if people teach to the test, because then we get better results. However, all of us as educators think it is a bad thing, because the domain that young people learn is narrower.

Dr. Bird: I think that we all said teach to the curriculum, rather than teach to the test. We want people to cover the whole curriculum and know, as Jerry said, the experience that they will have when they are examined, so that it is not a shock and they can show their best when they are being tested. That is about understanding the style of the examination that they will be required to sit. Our job is to ensure that the tests maintain sufficient variety and coverage of the curriculum over time, so that it is not easy to teach to the test in the narrow sense in which I think you mean it—saying "I am going to predict the six questions that will come up next summer, and teach you seven topics. I guarantee that six will come up." You are leading us towards a very narrow, output-driven thought. We certainly want to give people every encouragement to cover the actual curriculum, which is why we make specifications available, as Greg said. When I was doing A-levels, I had no idea what the curriculum was, which cannot have been a good thing.

Greg Watson: If I may add to that, we have talked quite a lot about innovation, and it is a real benefit to have more than one awarding body. The competition of ideas has been a powerful driver to keep syllabuses interesting and make subjects stimulating. In the era before micro-regulation, which really started in 1998, there were outstanding partnerships in a whole range of subjects—science, geography, maths—between individual awarding bodies and very creative university departments in places such as York and Cambridge. That was true also between groups of forward-thinking teachers—I can think of the Suffolk science movement. That interplay of ideas put a range of options out there for young people and their teachers. A school would pick its syllabus to suit what kind of school it was and the young people that it had. We are beginning to get some of that back. There have been signs that we have moved back in the right direction, and if we get the form of regulation right and it moves from the micro level to the level of looking after the system, there is scope for the interplay of different ideas among organisations such as ourselves. All the evidence from the work that we have done when we have had a bit more freedom to operate, and been less under QCA diktat, is that we have been able to reignite interest in some subjects. We are doing it again with A-level history, in which we have gone back to an alternative, more research-based approach, sitting alongside what we might think of as a more traditional style. Everything that we hear from the chalk face is that some people learn the skills of history much better by going out and touching it with their hands than by simply learning a load of facts from a textbook.

Jerry Jarvis: Teaching to the test is a perennial issue for us. When I talk to the Specialist Schools and Academies Trust and say, "Help me to avoid being cast in the role of someone who is encouraging teaching to the test," it slaps me on the back and says, "Get on with it. That is exactly what we are supposed to do. We are supposed to set a standard, teach the syllabus and assess whether students have met that standard." I do not think that, with the enormous stakes that are placed on league tables, you can avoid the accusation that teachers are spending time on examination preparation. It is inevitable, and it was always there. A long, long time ago, when I was doing qualifications, I remember practising on past papers and wishing I had a father who knew more about maths than my friends' fathers did. What has changed is the pressure to succeed. It really defines careers inside schools and colleges, and it is something that we have to guard against. The thing that drives me along is that if we can use technology and all the techniques that we have to make exam preparation relatively simpler, we can spend much more time in enriching education. Why are people educated? Is it for entry to university? Is it in preparation for work? Or is it for the sheer joy of learning? I seem to remember that there was a lot of that in the past.

Q225 Fiona Mactaggart: Absolutely, Jerry, yet your organisation is producing ResultsPlus, which, as I understand it, will be able to show me, as a teacher, that, fine, I might be wonderful at teaching Jane Austen but when it comes to Shakespeare my students are much more wobbly than Barry's students are, and that I have to shape up on that bit of the curriculum. Do you think that there is a risk that ResultsPlus will actually encourage that sort of thing?

Jerry Jarvis: What I have had back from a large number of teachers is that we are moving towards true personalised learning. Because such information is available, they can spend less time on

revision, which becomes far more to the point. Teachers are able to go at the speed of each different student, and they have more time to indulge in the richer aspects of delivering the syllabus. In fact, in our experience, the process actually works the other way around.

Q226 Fiona Mactaggart: You make it sound as if it is the child who will be the level of reflection that a teacher will make. Is it not more likely that it will be the teacher who will be the level of reflection in this sort of a system?
Jerry Jarvis: Having deployed the technology, it is clear that it operates at many different levels. As a tool for the senior management team inside a school or college, it pinpoints exactly how well the school or college is delivering its curriculum across all the subjects. For an individual teacher in the classroom, we are seeing a jump in the performance of their students. Teachers are literally taking the information and using it to their advantage.

Q227 Chairman: But you do not all agree with that, do you? Some of you are not using the technology.
Greg Watson: We are looking at similar ideas at the moment.

Q228 Chairman: So you approve?
Greg Watson: This is nothing new, in fact. Any teacher who wants to do the best for the young people in their classroom will reflect on how they are doing, and will compare how they did this year with last year. For some time, teachers have been able to get hold of exam answers and have a look at them to see how well they did. For many years, many teachers have come to professional development events—INSET events—that we lay on. Some involve a general run across the syllabus, but some are targeted at particular areas. If we get feedback from teachers that they are struggling with the coursework element, we always lay on support for them to come and talk about coursework and perhaps learn from other teachers. I think that the desire to understand, to diagnose, to target a bit of effort to develop professionally has been with teachers for a long time. We have started to explore whether technology and some of the data that we have could add to that.

Q229 Chairman: There are no dangers?
Dr. Bird: There is the issue about releasing detailed results directly to students. We think that that can be mediated through the relationship that students have with the centre and with their teachers rather than giving results to them directly, if that is what you mean by richer information being available. We certainly think that teachers can be assisted by having richer feedback about how the cohort of students is performing so that they can modify their teaching methods. It is obviously better if they get that from formative assessments that they carry out in the classroom rather than from summative results for the cohort that has just gone, as those students already have their qualifications, whether good or bad. Many people's support materials for classroom teachers contain formative tests in the form of homework, which essentially have the teacher ask, "Did I get that algebra across correctly or not?" That must be a helpful process, to assist teachers and curriculum heads in understanding how different teachers are performing in different parts of the curriculum, so that they can play to their strengths and fill in where there might be weaknesses, because people are not universally good at doing everything in the curriculum.

Q230 Fiona Mactaggart: When you design questions for examinations, which of course is a mystery to most people—we can only imagine the types of questions—you test all sorts of questions and design them to find out if they show specific knowledge or a range of skills, and such like. When you do that, do you think about the skills or knowledge that reveals something beyond the ordinary? We talked earlier about creating an A* grade, for example. Clearly that must have implications for the design of questions, to create the space for a young person to show that ability. None of your evidence tells us how that is done.
Chairman: One of you answer that question, or we will be here all night.
Greg Watson: It is a very careful and very rigorous process; a lot of what you suppose to be true is true. In writing a syllabus, we are not simply writing down, "Oh, read this book and you will know all about the Second World War." We will actually unpack that process to say, "By the time someone has successfully completed a GCSE course, they should know the following things and they should be able to do the following things with that knowledge." We will then describe a series of levels, which we call grades, that will differentiate the level of knowledge that you would be likely to have and what skills you would be able to apply to that knowledge. Interestingly, in the context of A* we have been doing a research project in advance of introducing A*, because we have been through the process of redescribing the syllabus. We have now described to ourselves, "What does an A* grade historian have by way of knowledge and skill that you would not have expected of an A grade student?" We are now trying out questions at the moment for exactly that reason; we want to discover what types of questions unleash that potential and what types of questions are sufficiently open-ended to provide that extra stretch and also, importantly, what types of questions successfully differentiate the most able students from the rest and do not lead them all to being clustered around the middle of the mark range. So that is the type of work that we are doing. We have a great army of research people sitting behind everything that we do and that is what they do all day and every day.

Q231 Chairman: So you want to take over the world, do you not? You do not want the Qualifications and Curriculum Authority or a national curriculum. You want to set the whole thing really, because coming through is the idea that you can do it all. Your exams drive the curriculum.

Greg Watson: I think that we have very important expertise. We are capable of adding more value in terms of innovation and enriching and discovering new approaches to subjects.

Q232 Chairman: Is it not dangerous to let you take over the world of education?
Greg Watson: There is something missing in my equation, which is why we need a regulator, and it is that the public's stake and society's stake in the qualifications system is much bigger now than it has ever been. Because of that, the public want to know that young people's education is safe in our hands and that we are doing what we should be doing absolutely properly. It is quite right that somebody who is not sitting along this line of witnesses here should have an independent view and should be tracking effectiveness.

Q233 Chairman: So you want a weak regulator to defend you from public criticism?
Greg Watson: I want a strategic regulator who can really stand up on the basis of hard evidence and support us when we say that we are looking after standards and everybody is getting a fair deal. There should be an independent body that is not a vested interest, which we might be seen as being, and that can say exactly the same thing and say it with confidence, on the basis of hard research.

Q234 Ms Butler: I just wanted to go back to an earlier question. You talked about Key Stage tests and when those tests are taken. As you know, there is a push for stage not age; the stage when the person is ready to take the tests. I wonder whether the panel could just say what they feel would be any unintended consequences of that.
Jerry Jarvis: Did you say "unintended"?

Q235 Ms Butler: Unintended, yes. Will it be an extra burden on the teachers? Is it reliant on the teachers' understanding of that child's abilities, or is the focus all on that teacher?
Dr. Bird: I take it that you are talking about the confirmatory single level test.
Ms Butler: Yes.
Jerry Jarvis: I think that the point that you are making relates to the type of "test when ready" issue and whether that would have unintended consequences. One positive unintended consequence is that it would perhaps take the pressure off that terrible examination point in the cycle. To me, it feels right to test when ready. That would require a much higher degree of administration and control and the provision, perhaps, of technology to assist the teacher in the classroom. It would be bound to have a number of knock-on consequences. My view is that examining when ready would have far more positive than negative outcomes, because it would tend to be about successful delivery. It would be a huge advantage for the student and the teacher.

Q236 Ms Butler: Are there any other views on that from the other members of the panel?
Dr. Bird: I agree with Jerry. There is a danger that teachers who lack confidence would fall into a test, re-test, re-test mentality with students who are bouncing along just below a level. That might be a problem but, generally, testing when ready would have many more positive than negative outcomes provided that it is allowed to replace mass testing. There is a danger that we would have single-level tests for one purpose and some sort of national audit test run at the end of the key stages. In many ways, that would be the worst of all worlds.

Q237 Ms Butler: What do you think would be the best way to ensure that there are more positive than negative outcomes? For those children who are bouncing along and who may be ready but whom the teacher does not think are ready, how can we can build in a caveat to ensure that there are more positives than negatives?
Greg Watson: There is an important principle, which runs across all assessment, that says that we should be absolutely clear about the purpose of assessment up front. Ideally, there would be one purpose of assessment rather than trying to do many things at once. We should be clear that we are assessing in order to inform and support learning and to help teachers and to guide their efforts. That is a very different purpose from measuring to compare schools against one another to create league tables and support parental choice. Our field does not involve Key Stage testing—we are involved in GCSEs and A-levels—but being clear on what we are testing and why is a really important discipline. It is important to get that right in this context so that we do not have the dangers that have been highlighted about a conflict between an assessment to support learning and an assessment to rate a school's performance overall.
Jerry Jarvis: Here I am back on my subject. The infrastructure prevents testing when ready. The process is geared to delivering information at a constant rate to all students in the cohort, who will then be examined at the end. We have to do much more than simply say that we will test when ready. The infrastructure—personalised learning—must be in place for such an approach to work. If that happens, it would be worth doing.
Chairman: We are getting really tight on time.

Q238 Mr. Chaytor: I have a short question to put to each of the four members of the panel. Does the use of league tables help or hinder your organisation's work?
Murray Butcher: I probably have the easiest answer of the four panel members. Given that the majority of our work is in vocational qualifications, league tables do not greatly affect our activity.

Q239 Chairman: Nevertheless, what do you think of them? You are a professional educator.
Murray Butcher: My anxiety has already been expressed. It depends on the purpose of the assessment and on what are trying to draw from it. I recall that Ken Boston, the chairman of the Qualifications and Curriculum Authority,

mentioned just before Christmas that one of his colleagues identified 22 different purposes for assessment. I suspect that 22 is almost a random number and that you could raise it somewhat. I suggest that league tables have far too much pressure put on them and that they are probably not a sufficiently refined instrument to give the exactitude that they purport to give.

Dr. Bird: I do not think that tables impact on our work particularly strongly. Obviously, our outputs are part of what formulates them and we are aware of pressure around the boundary between D and C grades in some subjects. The inquiries about results and the seeking of scripts after the exam to check on that boundary has a relatively marginal impact as far as operational activities are concerned. I have no other comment really.

Jerry Jarvis: There is no question that the pressure creates some disadvantage to us—particularly in respect of calls for more research and more opportunities to take—but there is a huge pressure on schools to maximise the sorts of qualifications that are taken. Therefore, we see schools and colleges chasing qualifications in an unhealthy way. However, that same competition demands that we drive up service and standards in a way that we otherwise would not. There is no question that league tables affect what goes on inside a school. However, one that benefits society, if not necessarily Edexcel or my colleagues, is the pressure to continue to compete for those qualifications because they are taken so seriously.

Greg Watson: Undoubtedly, pressure creates unhelpful tensions in the qualifications system. If our fundamental job is to assess learners and provide assessment that adds to learning and that helps to structure and motivate learning; the other dimension of what we do is used to determine teachers' pay rises and that gets in the way of that relationship. That explains the distance that some teachers feel exists between them and us, and is something that I want to break down. I hope that the new independent regulator—and we have mentioned him several times—will be very aware of that in balancing change and stability in the system. I wonder whether the regulator might run the league tables or develop a carefully defined relationship with those who determine such tables so that the tables do not get in the way of what the qualifications are fundamentally there to do, which is to give young people a portable measure of what they have achieved to take on to the next stage in life.

Q240 Mr. Stuart: Just to go back to Edexcel ResultsPlus, will you look at common fields and common analysis? Referring to the earlier question about the unintended consequences, could it be that every teacher in the future will be able to have marks for their entire career? Would you be able to categorise them as a four-plus teacher, a three-and-a-half plus teacher and a three-plus teacher? Now that the Prime Minister has said that he will not put up with schools that are below a certain level and they will have to be closed, will we move to a stage in which politicians in here will say that any teacher below a certain level will be fired within two years or brought up to scratch?

Jerry Jarvis: Right now, those teachers are liable to be sacked anyway for their performance as defined by the league tables. Teachers, management teams and students can avoid that if they use ResultsPlus to continuously assess their performance and improve it—so it would have the opposite effect.

Mr. Stuart: And to continuously avoid splitting infinitives.

Q241 Fiona Mactaggart: Do exams have to be in the summer when hay fever and hot weather cause problems to some candidates?

Jerry Jarvis: That is a much broader question because of university entrance and the competition for places. As someone said to me when I joined Edexcel, the reason why we have university entrance at that time of the year is so that children can bring in the harvest. We seriously should look at the overlap that exists between the teaching curriculum and the examination process that leads to higher education. Right now, I do not think that any of us can see a viable methodology for moving that examination time.

Q242 Chairman: Even trying to get Oxford and Cambridge to have the same application process as other universities has been very difficult. You looked rather uncomfortable—Greg particularly, and also you Jerry—when I said that perhaps the opinion is that you are getting out of touch with what universities and teachers are saying. You seemed more hurt at that than at any other question. May I take you back to Kevin Stannard, the director of international curriculum development at Cambridge University? He said: "Through the late 1980s and the 1990s, the gap between academics and schools got wider and wider in terms of academics disengaging from exam boards. I think what we are living with now are the implications of that". Why would he say that, Greg, if it was all nonsense?

Greg Watson: My discomfort is that getting qualifications right is the business of trade-offs. There are so many different stakes in the qualification system. For it to work well, and for it have everyone's confidence, everybody has to have their say; and those different and sometimes competing demands have to be balanced—and balanced well. I recognise that universities have a bit less of a stake as a result of the last 10 years; employers will feel in some areas that they have a bit less of stake in some parts of the system, although it has grown in others. We are sitting in the middle, and inevitably juggling all those. We are certainly making a lot of effort to build more relationships back with higher education; they have weakened a bit, as we have had more of our focus drawn towards the Government of the day, and a not particularly independent regulator.

Q243 Chairman: Is there anything you want to say to the Committee before we wind up—anything that you think you should inform the Committee about what you have not yet said because our questions were inadequate?
Jerry Jarvis: If I had not got across my point about the importance of the things that Ken Boston said and about all the money that I have spent, I would be very disappointed.
Chairman: I think that we got that.
Jerry Jarvis: But we have a serious opportunity to shift that stubborn A to C statistic that has dogged us for many years.
Murray Butcher: A point came from Fiona about releasing information and the problems it might cause. I think that we should look at it from the reverse direction. It is sensible for awarding bodies like ourselves to find ways to be more and more open in our processes. If part of that is giving feedback, as mentioned, with all the data that we collect from various examining activities, there is a growing duty and responsibility, at least on the four of us, to be more creative and more open with that information. It will only make the system more robust.

Q244 Chairman: I feel a bit guilty about you, Murray, because we have not asked you as many questions as others. Surely we should have asked you about this vast expansion of apprenticeships. Are they going to be linked to qualifications in a serious way—or should they be?
Murray Butcher: I regard apprenticeships as they stand as serious qualifications.

Q245 Chairman: I am sorry, but they are not naturally linked to qualifications, are they?
Murray Butcher: They are an amalgam of two or three different bits.

Q246 Chairman: A lot of them are still time-served, with hardly any paper qualification. Evidence given to the Committee suggests that that is the case.
Murray Butcher: I would be extremely surprised, because the QCA and the Learning and Skills Council have identified apprenticeships as involving national vocational qualifications, possibly what is called a technical certificate, which underpins knowledge, and functional and basic skills. That is the apprenticeship.

Q247 Chairman: So you are totally happy with what is happening with apprenticeships?
Murray Butcher: I would not say that I was totally happy. They probably need more support in the workplace, and still more understanding by employers as to what they offer. I know that the achievement rates remain relatively low—at about 40%.—partly because young people who begin apprenticeships achieve employment and leave the programme. So there are obviously some problems about ensuring continuance into employment.
There are some structural issues that need to be resolved, but as a product they provide a good background and a good basis for employment. What is planned at the moment is increasing flexibility in the structure of the apprenticeship. That is what QCA and others are currently looking at.
Greg Watson: Just a final thought on standards, which is right at the heart of the review. In general in this country, we spend too much time having a debate about standards at the low level of an individual question paper, the marking of this year's examination and small changes here or there in the percentage of candidates getting a given grade. We are having that debate at the wrong level. The potential for standards to move and for public confidence to be shaken is greatest when there is wholesale, system-wide change or major structural changes to long-established qualifications. The acid test for looking at the move to an independent regulator is whether we will have a body that is sufficiently able to look at the macro-level changes and the effect that they may have on standards and public confidence and worry much less about the detail of which individual qualification is which. As I hope you have heard, that is something on which we have tremendous expertise, tremendous power to innovate and a tremendous ability to add to learning by making exams sympathetic to the business of learning, rather than something that sits outside it.

Q248 Chairman: But, Greg, you are still selling a product. We on this Committee and our predecessor Committee have argued for a long time that you do not need a new A* plus and that you can just let the universities have the full marks at A-level. Would that not be just as good as an A*? Why do you want an A*?
Greg Watson: We do not just want the A* grade, but the new style of syllabus that we have developed to support it. We have been back to every single A-level subject and—to answer your question, Fiona, about where these questions come from—defined in new terms what you have to be able to do to get the highest grade in an A-level. Now, we are setting questions to match that.

Q249 Chairman: Will that inevitably lead to even more children from elite independent schools dominating the leading research universities in our country?
Greg Watson: We will just mark what we are faced with and give the A*s to the A*-standard candidates.
Dr. Bird: There is a forecast in our evidence to that effect. I cannot remember which paragraph it is in, but we have provided a forecast of that in our evidence.

Q250 Chairman: Jerry, do you think that, too? Do you think that will be the inevitable conclusion? Will an A* mean that even more children from the independent sector and the best of the state school system will dominate?
Jerry Jarvis: That is certainly a tendency and something that we really have to work on. However, it is striking that if you look at individual schools' performance, you can see that two schools facing each other in the same street and drawn from the

same community can have dramatically different outcomes, and we need to understand much more why that is. The fact that schools that have dramatic advantages will do better is almost an inevitability. The private sector already has the vast majority of passes at A-level, so you would expect it to excel at A*, but a number of other schools also excel and can also differentiate themselves hugely. In the evidence that the awarding bodies gave during last summer's results, we faced the press together and pointed out that there were differences in different school types and that the overall achievement rates at some school types were doing down, not up. However, there is, of course, a general tendency that the better-off and the more advantaged will gain higher grades.

Chairman: Thank you very much for that evidence. It has been a long session, but it has been a good one from our point of view. Sorry if we pushed you too hard at any time or if we have been hard on you, although I do not think that we were. You have been excellent witnesses. Thank you very much for your contributions.

Monday 28 January 2008

Members present:

Mr. Barry Sheerman, in the Chair

Annette Brooke
Ms Dawn Butler
Mr John Heppell
Mrs Sharon Hodgson
Fiona Mactaggart
Lynda Waltho
Stephen Williams

Memorandum submitted by the Confederation of British Industry (CBI)

SUMMARY

1. As the UK's leading business organisation, the CBI speaks for some 240,000 businesses that together employ around a third of the private sector workforce, covering the full spectrum of business interests both by size and sector. The CBI understands the Committee's wish to examine the current the testing system, and the role of the testing and assessment regime in ensuring accountability and raising standards.

2. CBI members are committed to investing in the skills of their employees. In a global economy characterised by rapid change, young people need to have transferable skills to ensure their continued employability. The UK must have an education system which produces young people with the skills employers need, so that we have a world-class workforce that can compete with Europe, the US and Japan and the growing challenge from China and India.

3. Employers value test and examination results as a way to monitor the performance of our education system in ensuring young people attain the skills they need for success in life and in work. In this paper, we focus on the public examinations taken between ages 16–19—GCSEs and A-levels. Employers use an individual's GCSE and A-level examination results as a good indication of a young person's abilities, particularly during the early stages of a person's career. However, employers are also interested in the tests taken in schools by children at Key Stages 1–3 as they provide an indication of whether educational standards are improving. Such tests must provide an objective and reliable measure of the standards achieved by pupils at crucial stages in their development.

4. At present the CBI has two key concerns. First, too many young people leave school without the necessary literacy and numeracy skills to succeed in work and, secondly, insufficient numbers of students leave school and go on to study science, engineering and maths at university.

5. Problems with basic skills continue to manifest themselves in the workplace—20% of the current workforce lack either the literacy or numeracy skills expected of an 11 year old. It is estimated that lack of basic skills costs the economy £10 billion per year. Employers invest £33 billion in training their staff every year and recognise their responsibility in training their employees with job-specific skills. But it is the Government's responsibility to ensure young people leave the education system with the basic skills.

6. Business demand for Science, Technology, Engineering and Maths (STEM) skills is high, but supply is not keeping pace. Employers are concerned that the number of graduates in key disciplines such as physical sciences or engineering is falling. A third of businesses think raising the number of STEM graduates should be a top priority for Government as jobs in these sectors are set to expand. Whilst the number of science degrees awarded at university has increased, the large rises in biological and computer sciences hide a decline in physics and chemistry. Falling numbers of STEM graduates can be traced back to the shortcomings of school science: fewer students studying triple science; too few schools with specialist chemistry and physics teachers; too little time spent doing experiments; and patchy lab provision.

7. The focus on English, Mathematics and Science in the curriculum and assessment regime is appropriate, and reflects how important these subjects are to the future prospects of young people. Employers' confidence in academic qualifications is determined by the number of young people leaving education with the skills that business needs. The CBI has been examining trends on science qualifications because of concern about declining number of students taking STEM degrees.

8. Finally, employers understand and recognise GCSEs and A-levels and believe them to be world class qualifications. But they remain deeply concerned about the number of school leavers who have inadequate standards of literacy, numeracy and general employability skills. The new vocational qualifications—the diplomas—will help to create the skilled and competent employees business needs. Work-based qualifications must also be improved so that they reflect the skills and competences employers need.

EMPLOYERS ARE CONCERNED ABOUT THE EMPLOYABILITY AND BASIC SKILLS OF YOUNG PEOPLE

9. Employers are concerned that too many young people leaving school at 16 do not have basic skills. Poor literacy and numeracy skills damage people's quality of life and their employment prospects. Those with poor basic skills are more likely to suffer higher unemployment rates and low earnings, with poorer chances of career progression and social exclusion. They are also less likely to be fully effective in the workplace, damaging the competitiveness of UK firms. According to the latest CBI/Pertemps Employment Trends Survey nearly nine out of 10 employers (86%) think that ensuring young people leave school with basic literacy and numeracy should be the Government's top education priority.

10. The CBI/Pertemps Employment Trends Survey 2007 also found over half of employers were concerned with the literacy levels (52%) and numeracy levels (50%) of school leavers. Employers often find that they need to remedy deficits in the basic skills of their employees. The survey also found that 15% of employers had to give school leavers remedial training in numeracy, and 13% in literacy.

Employers' dissatisfaction with school leavers' key skills (%)

Self-management	67
Basic literacy	52
Employability skills	51
Basic numeracy	50
Positive attitude to work	44
Use of IT	8

11. Employers expect young people to enter the world of work with the basic numeracy and literacy skills required—in practice this is the equivalent of a grade C or above in Maths and English GCSE. Pass rates have increased—the percentage of pupils achieving a GCSE A*-C has increased from 57% in 1997 to 62% in 2007 in English and from 47% to 55% in Maths. But there remains a significant proportion of pupils still underachieving at GCSE level. In 2007, only 46% achieved five or more A*-C including English and Maths and a fifth left school with no qualification graded A*-C. This is a significant proportion of pupils who are still underachieving at GCSE and this is a key concern.

12. Given employers' concern that young people are leaving school without basic competency in literacy and numeracy, the CBI undertook to explore the ways literacy and maths skills are used in the workplace, and the shortfall in skills that employers experience. This work, sponsored by the DfES, was designed to identify the key functional skills needed by people at work—and the study was informed by survey and case study work. The final CBI report, *Working on the Three Rs*, was published in 2006 and defined the functional literacy and numeracy skills necessary to be competent in the world of work.

13. On literacy, the CBI report showed that reading basic text is a vital skill for the workplace and that writing a short report, with legible handwriting, is also essential. It is important to consider reading and writing separately as they are different skills—and they should be assessed separately too. To be functionally literate, an individual must be able to:

- read and understand basic texts;
- construct properly spelt, grammatically correct writing that is suitable for the audience;
- write with legible handwriting;
- understand oral communications and react appropriately; and
- be sufficiently articulate to communicate orally.

14. A good grasp of basic numeracy is also a vital tool for work, and is used in a wide variety of contexts—from checking change in a supermarket to understanding performance targets. The ability to interpret and respond to quantitative data is also an essential skill for modern working life—there are charts, graphs and tables in most workplaces. It is important that employees understand these in order to contribute to problem solving and quality improvement and help create high performance organisations. To be functionally numerate, an individual must have confidence with:

- multiplication tables and mental arithmetic;
- percentages, ratios, fractions, decimals, ratio;
- different measures and conversion between them;
- spotting errors and rogue figures; and
- odds and probabilities.

15. The CBI was delighted that the Government commissioned the development of functional skills modules designed to test the practical application of numeracy and literacy—building the skills people will need in everyday work and life situations. These functional skills units will be incorporated into: English and Maths GCSEs; the new diplomas and apprenticeships; in addition to being available as 'stand alone' qualifications. The functional units—which will be taught from September 2010—will be offered at the standard of Level 2 (GCSE A*-C), Level 1 (GCSE D-G), and entry level standards.

16. In 2005, the CBI set a minimum target of 90% of young people achieve functional skills modules at Level 1 and 80% to achieve a Level 2 by 2010. We were therefore pleased to see that last year 90% achieved a Level 1 in functional literacy and numeracy. However, employers do expect young people to have the skills commensurate with a C or above at GCSE level (Level 2) and it is therefore disappointing that we remain so far from achieving this target. While functional skills modules within GCSEs offer a welcome strengthening to the system, they cannot replace the goal of having more young people achieving a C or above.

17. Employers also expect young people to have employability skills. The CBI's *Time Well Spent* report published in 2007 identified the key transferable employability skills: self management, team working, problem solving, communication, application of literacy, business awareness, customer care, application of numeracy and Information and Communication Technology (ICT). In 2007, 50% of employers were dissatisfied with the employability skills of school leavers—67% were dissatisfied with the self-management skills of school leavers whilst 92% were satisfied with their IT skills. The Government intends to introduce Personal, Learning and Thinking Skills modules to diplomas and apprenticeships. While these are a step forward, these skills do not reflect employers' definition of employability skills. The framework comprises of six groups of skills: independent enquirers, creative thinkers, reflective learners, team workers, self-managers, effective participators. The CBI would be happy to work with the Government to resolve these concerns.

Examination Results show the Declining Number of Students taking Science and Maths

18. The CBI shares the Government's ambition to become the world's leader in STEM research and development. The UK must continue to attract—and attract more of—the brightest and most creative minds to these sectors. The CBI has proposed a target of 25% of young people studying STEM subjects at university. This target is essential if the UK is to fill the graduate level jobs that are predicted in these sectors by 2014.

19. However, a study of A-level entries reveals some worrying long term trends. The number of A-level exam entries increased by 14% between 1984 and 2006 but science subjects have not followed this trend:

- *Physics*—the absolute number has fallen by 57% (31,065 fewer pupils) and as a proportion of all A-levels from 9% to 3% (a 66% fall). Only 21% are girls. Physics was the most popular science A-level in 1984 but is now the least popular.
- *Chemistry*—the absolute number has fallen by 28% (13,534 fewer pupils) and as a proportion of all A-levels from 8% to 5% (a 37% fall). Virtually equal numbers of boys and girls take Chemistry—49% are girls. Chemistry was the second most popular science A-level in 1984 and retains its middle ranking.
- *Biology*—the absolute number has stayed broadly unchanged—a 3% rise (1,453 more pupils). As a proportion of all A-levels has similarly remained constant at a steady 7%. Biology was the least popular science A-level in 1984 but is now the most popular.
- *Maths*—the absolute number fell by 25% between 1999 and 2005 and as a proportion of all A-levels from 9% to 7% (a 29% fall). But last year saw an encouraging 8% rise in the number of young people taking Maths A Level (to 49,805 students). Only 38% are girls.

20. *The fact that too few students are taking science A-levels is having an impact on the number of students obtaining first degrees in science subjects. While the number taking "science" has risen by half (49%) since 1994, much of the increase is due to the number taking biological sciences and ICT. The underlying figures show very concerning trends for those sciences business needs—with a long term decline in the number taking physical sciences (physics and chemistry) and engineering and technology. After a dip in the numbers taking these subjects, we are now possibly seeing a recovery, with an increase in the number of applicants to university courses but the absolute numbers remain very much lower than in 1984.*

GCSEs and A-levels are Valued and Understood by Employers, and Vocational Qualifications must Deliver Business Relevant Content in ways that Motivate Young People

21. Employers use GCSEs and A-levels as a key method of benchmarking potential employees. For example, companies set minimum requirements for entry to jobs such as five GCSEs A*–C including English and Maths. At higher levels, employers will look at A-level grades when recruiting graduate candidates. Employers understand GCSE and A-level qualifications and have no problem differentiating between candidates with good grades. This is because employers always interview candidates and do not offer jobs without interview as many universities now do.

22. The CBI welcomed the Government's decision to incorporate the planned A-level review into a wider 14–19 review of educational qualifications in 2013. We urged the Government not to prejudge the 14–19 review, as although employers involved in the development of the new diplomas are enthusiastic, there remain concerns about the new qualifications. CBI members believe that it is too early to talk about withdrawing A-levels and GSCEs. The new diploma qualifications must:

- ensure more young people attain vital literacy and numeracy skills;

— stretch our brightest children—rather than became a dumping ground for the less academically able; and

— be attractive to young people currently disenchanted with the education system.

23. It is vital that young people have a wide range of attractive routes to choose from. Therefore, together with a high quality 16–19 academic route, there must be similarly engaging vocational and work-based choices available to young people—particularly the less academically inclined.

Quick Progress must be made in Improving Vocational Qualifications and Accrediting Employers' Training

24. The CBI welcomed the Government's plans for the creation of a new independent regulator for the exams and qualifications system. A key function of the new regulatory body will be the accreditation of qualifications. Strengthening employer confidence in the vocational qualifications system will require developing qualifications with business relevant content.

25. It is essential that when the education/training participation age is raised, employers have access to work related qualifications that make them willing and able to provide training to young people that leads to recognised qualifications. The Government has set challenging apprenticeship and qualifications targets for both 2011 and 2020. If these are to be met it is essential that employers and young people see the point in getting involved in work based qualifications. Employers will only see value in employing young people and arranging training towards qualifications if an employee is developing "economically valuable skills"—ie those skills that will lead to improvements in productivity and business performance.

26. The primary objective of training from an employer's perspective is to raise business performance, by having a competent workforce. Too many CBI members report that the recognised qualifications available are often out of date and irrelevant to their business—and do not reflect their specific skills needs. Findings by IRS, for example, indicate that NVQs, SVQs, and National Occupational Standards were used by only 32% of employers.

27. The Government has tasked the Qualifications and Curriculum Authority (QCA) with developing an accessible system for recognising high quality employer training within the qualifications system. Employers often provide staff with excellent training which is tailored to meet their individual business needs—and accrediting this training will build employer engagement with the vocational qualifications system.

28. The CBI has welcomed the recent QCA pilots to accredit employers' in-house training towards qualifications. The Sector Skills Councils also have an important role in ensuring their sectors' qualifications are fit-for-purpose. The CBI believes two models should be available: an employer becomes an awarding body to develop and award qualifications itself or an employer partners with an existing awarding body. The fundamental principle underlying a system for accrediting vocational training must be that qualifications fit employer training needs, and not vice-versa. Maintaining high quality standards will be essential, but this must be coupled with a flexible approach from the new regulator. For example, employer training should not necessarily have to conform to national occupational standards if robust industry or internal employer benchmarks already exist.

29. The QCA was tasked with showing demonstrable results by Christmas 2007 and we welcome the progress that is being made with employers achieving awarding body status and getting training accredited. The CBI is pleased to be working with the QCA, awarding bodies and employers to ensure that fit-for-purpose vocational qualifications are designed and delivered. We are pleased that John Denham, Secretary of State at DIUS has made progress in this area a priority.

January 2008

Witnesses: **Professor Steve Smith,** Vice-Chancellor, University of Exeter, **Professor Madeleine Atkins,** Vice-Chancellor, University of Coventry, and **Susan Anderson,** Director, Human Resources Policy, and **Richard Wainer,** Principal Policy Adviser, Education and Skills, Confederation of British Industry (CBI), gave evidence.

Q251 Chairman: I welcome the distinguished group of witnesses that includes Professor Madeleine Atkins, Professor Steve Smith, Richard Wainer and Susan Anderson. Thank you for coming, and I apologise for the delay. As you know, for 20 days we will be discussing the European treaty, and that was the start of the voting—it was on a programme motion, I think. I hope that now we will not have any more interruptions for a considerable time. You will be aware that this is our first major inquiry as the new Committee for Children, Schools and Families, and we are very keen to get to the bottom of the questions around testing and assessment. We are particularly keen to speak to the end users, such as the universities and employers. We always give witnesses a chance to say something to get us started, or they can go straight into answering the questions. You know the topic: is our testing and assessment system fit for purpose? When you wander around the world answering questions about UK education, a lot of people say, "We would like to know more about how our students perform, but not like the United Kingdom, which tests at seven, 11, 14, 15, 16, 17 and 18." They go through the catalogue of testing

that they are sure that we have. Are we where we want to be in testing and assessment? You can answer that question, or you can introduce yourself and say what you want to say.

Professor Atkins: I am Madeleine Atkins, Vice-Chancellor of Coventry University. I would like to make two comments on your introductory questions. First, I question whether we have got the balance right between a deep synoptic understanding of subjects at AS and A2-level, as opposed to having a broad range of quite superficial knowledge. As we move our curriculum towards problem-based and activity-led learning, which is very much in line with what the employers say that they want, we are finding that a gulf is opening up in the way in which students are prepared to learn as they come into university. My second comment is that as we move further towards the knowledge-based society, we find that many young people coming from school and college underestimate the amount of mathematics and numeracy that is required in higher education vocational programmes. I am delighted to see that Diplomas will have numeracy as a requirement in the new 14–19 Diplomas. Nevertheless, I wonder whether we have got the balance right.

Chairman: Thank you.

Professor Smith: Very briefly, I am here as Chair of the 1994 group of research-led universities. As Vice-Chancellor of Exeter University, a member of the Prime Minister's National Council for Educational Excellence, a member of the Universities and Colleges Admissions Service board and a member of the UK Post Qualification Admissions Delivery Partnership, I am particularly interested in answering questions and discussing issues of fair access and wider participation and the way in which the current assessment regime at A-level supports those aims. Given the nature of our intake. I am also very happy to talk about Exeter's experience with A grades and the prospect of grade A* at A-level, which opens up a series of issues, and also the accuracy of A-level predictions. I am very happy to talk about all those issues as well as the measures that we have to put in place as a university in order to cover the knowledge gaps that we identify in the existing A-level curriculum.

Chairman: We invited you as individual vice-chancellors, Professor Smith. Universities UK did not feel that it could add any value by appearing before the Committee on this subject, as it could not get any agreement among its members, which rather surprised us so we thought we would go for individual vice-chancellors instead. I was surprised at Universities UK's reaction, but never mind.

Susan Anderson: I will make a comment for Richard Wainer and myself. Our key message is that employers recognise and understand GCSEs and A-levels as high-quality qualifications. They see value in the new Diplomas, but they understand GCSEs and A-levels. We have concerns about literacy, numeracy and employability skills, and also about the number of students studying science and maths A-levels, because we are not getting enough young people choosing to study the science subjects at university that employers want. Those are the key points that I would like to emphasise in my opening remarks.

Q252 Chairman: I shall go back to the point that I was making earlier and want to question the two vice-chancellors. This morning, I visited Southfields Community College, which is a fascinating place, to look at what it does. It is an extraordinarily interesting, innovative college. I popped into some of the sixth-form courses and, as I usually am, was absolutely amazed at how hard the students work. We hear vice-chancellors and other organisations saying that they get people who they do not think have the depth or breadth of knowledge that they should have, which surprises me because such visits to schools point me in another direction. I went into an English class where people immediately wanted to know whether I knew Simon Armitage, because I am from Huddersfield, and what I thought of Shakespeare as a poet rather than as a playwright. They were fascinating and stimulating young people. A recent report from the Higher Education Policy Institute (HEPI) suggested that students, having worked through sixth form frenetically to achieve good results to get into university, are actually not worked very hard when you get hold of them, and that we have the most easy-going university regime in the developed world. If it is true that they are not that good, why do you not work them harder? Professor Smith, will you answer that?

Professor Smith: I will happily answer that. I actually think that university students work very hard, and the evidence supports that. There are, of course, questions about the evidence, but the basic evidence, alongside the number of 2.1 and 1st grades, is the proportion of people completing courses, the proportion of people satisfied with the experience and, crucially, the proportion of people moving into graduate-level jobs. The HEPI study is a good study, but please note that it talks about contact hours, full stop. I think that it is absolutely crucial to distinguish between being in a lecture theatre in some of the continental countries with 500 people present for an hour and doing seminar and tutorial work for an hour. I think that the HEPI study has limitations because of its methodology, but students at universities in the UK work hard and the results objectively show that.

Q253 Chairman: What about the first point? How do they arrive with you? It is a gross term to use these days, but are they oven ready when they get to you in Exeter at 18?

Professor Smith: Oven ready might not be quite the right term. Let me put it this way: we need a certain degree of subtlety. I have read all the transcripts for the Committee's previous sittings and think that the key point is that it is neither one thing nor the other. By that, I mean that the students come with skills that are different to those of people who went to university 20 years ago or 35 years ago, when I went to university. They are different sets of skills.

In preparation for today, we asked all the admissions tutors at Exeter what they find. You get two basic sets of comments: first, with regard to the right sets of study skills, they are actually rather well prepared, with the exception of independent critical thinking, which is why the extended project in the Diploma looks very exciting; secondly, there are differences in the subject knowledge that they arrive with, especially in some of the sciences. You might wish to push on that a bit more. In Exeter, for example, the level of maths that the students come with is a major issue. We put on additional maths in the first year from the business school right through to biosciences, physics, engineering and computer science because we do not think that everyone comes at the right level. Equally, in our English school they actually put on additional study skills modules for people in the first year because gaps were identified there as well.

Professor Atkins: We would say the same sort of thing. With regard to mathematics, our colleagues in science, technology, engineering and mathematics—STEM—would say that the range of mathematics now studied in the sixth form is much wider than it used to be, but that there is less depth, particularly around subject topics such as calculus. For example, for students reading engineering subjects, we have to put on supplementary work so that students—even those with a grade A or B at mathematics—can cope with things like fluid dynamics, heat transfer and engine cycle calculations. That area seems to trip up our inbound students. I absolutely endorse what Steve has said; there is a requirement for mathematics and numeracy right across the vocational field. Particularly in nursing and professions allied to medicine, we find that students who dropped mathematics at 16, having got their grade A to C at GCSE, forget it. It is not like riding a bicycle. Also, if they have learned maths as a selection of discrete random techniques, they arrive to learn about drug medication, for example, and cannot quite remember which bit to put above which bit when calculating the percentage. That is quite worrying. The amount of additional confidence-building that we have to do with numeracy, as well as the actual mathematics input, is quite considerable. As with Exeter, we fund a large maths support centre that tests 800 students coming into Coventry University on the induction week and continues to offer drop-in sessions. We see about 2,000 students a term on a drop-in basis. This is a major issue. It is not a problem just in STEM subjects, but in business, nursing and other areas. I also agree with Professor Smith about academic writing. We find that there are problems in the ability of students to do two things. First, we are having to focus on bringing independent critical analysis to web-based materials or Internet-based sources. Students are very able to source materials on the Internet from many different places, but are not quite so able to bring critical appraisal to those sources. For example, they get information from "Wiki" and it arrives in the essay without any greater consideration. Secondly, students are often unable to structure a report at length, rather than produce a short piece of writing. They do not always understand that writing is a recursive process that needs to be worked at. Those are some of the areas where we have to put in additional time, help and resources to aid students' academic writing.

Q254 Chairman: Thank you. That evidence will be familiar to you, Susan. It is the sort of thing that the CBI has been saying for some time.

Susan Anderson: Those issues certainly resonate with us. When we talk to employers about graduate skills, more often, they are concerned about the quantity of graduates, particularly in the STEM areas of physical science, maths and engineering. They are in very high demand among employers, not just in the specialised manufacturing or engineering sectors, but in the financial services. Employers are concerned more about the quantity than the quality of such graduates. Sometimes, an employer from a pharmaceutical firm, for example, will say that somebody has come to them with bioscience or another relevant degree, but has not done the particular bits that they want. It would not take much to fix that. Often, major employers are working with universities—whether in the IT or the pharmaceutical sector—to see how they can get courses in those disciplines that reflect business needs. That is an area where we in business can work more closely with universities and that is happening. On literacy and numeracy, we are more concerned about school leavers at 16 and 18 than about graduates. I emphasise that when we ask employers for their views on school leavers and graduates it is literacy and numeracy that they are concerned about. We get very few complaints about IT skills. That is rather different from the work force, where there are some issues with IT. Only 2% of employers say that they have any problems with graduates' general IT skills and about 8% say that they have problems with those of 16 or 18-year-olds. For whatever reason, in some areas there are very few problems; that is probably because those areas are reinforced by home use. It is important to us, and we have worked closely with the Government, to establish what we mean by basic literacy and numeracy. We are happy to expand on that, as we have in our paper. Similarly, on the employability skills that employers are looking for, we do not label them in quite the same way as universities do. But we know that young people are getting those skills as we have defined them—self-management and team-working—from their school and university experiences. What they are not always very good at is calling them by the right labels, or being able to talk about and demonstrate how they have worked as part of an effective team on a university project, or become an effective self-manager. Sometimes the problem is with labelling and terminology, rather than the fact that particularly university graduates do not have those skills at all.

Q255 Chairman: So, does your organisation aspire for well rounded graduates with a broader competence?

Susan Anderson: Our bigger problem is lack of STEM graduates. If we have an issue to address, we are saying that those areas come pretty much top of the list. Employability skills are also important, but we think that universities, at least the good ones, really get those skills, and understand that employers are looking for people who can apply their literacy and numeracy skills in the workplace. For the university graduates, it is often a question of talking the language that employers talk and understand, and being realistic. For example, if you say that you are going to go into a business environment and be a leader, you will not be a leader on day one. Sometimes, graduates need a bit of realism. Primarily, when we talk about quality, that refers to employability skills, but the key concern is quantity of STEM graduates.

Richard Wainer: I would like to add to that. While the quantity of STEM graduates is probably the primary concern in HE, you are right about employers wanting well-rounded people coming out of university. Our surveys show that about 70% of graduate jobs require a specific degree discipline, because employers are looking for a well-rounded person with good literacy and numeracy, and good employability skills.

Q256 Chairman: Someone with a Diploma?

Richard Wainer: Quite possibly. Our members have looked at the Diplomas and, in principle, they can see them working. But there are many issues to work on between now and September, and going through to 2013, to ensure that they really are a high quality route for young people, both into university and into employment.

Chairman: I am not saying this to take a pot shot at you in the CBI, but some of us who were around at the time thought that the original Tomlinson reforms were rather stymied by the CBI attitude. But, we will come back to that and drill down on it. What I want to get out of this sitting is whether, if things are not quite as you want them now, is it because of the supply chain, what is happening down there, too much teaching for tests? I see the Permanent Secretary has just come in and is sitting behind you. Is it something that the Government have been doing over the past 10 or 20 years—not limited to one Administration? Is something in the schools not quite right for giving the right kind of product? Do not answer that now, but, by the end of the sitting, that is what my colleagues hope to be able to drill down and try to discover.

Q257 Stephen Williams: I would like to start with some factual questions to the two vice-chancellors. What is the social composition of undergraduates at Exeter and Coventry? How many, as a proportion, come from private schools, and how many come from the lowest socio-economic group?

Professor Smith: At Exeter, 74% of undergraduates come from state schools, and, I would need to be absolutely sure, but I think that 17.2% come from the lowest socio-economic groups.

Professor Atkins: At Coventry 97% of undergraduates are from state schools and 38% are from black and minority ethnic groups. Depending on how the indicators are cut and used, the figures are 41% from the lowest socio-economic groups and 21% from the lowest participation neighbourhoods.

Stephen Williams: Flipping the figures around, 26% of undergraduates at Exeter are from a privately educated background compared with 3% at Coventry. How does that fit in with the targets that the Office for Fair Access sets you as institutions? Presumably Coventry is meeting them comfortably?

Professor Atkins: We exceed our benchmarks, yes.

Professor Smith: You will be pleased to know that Exeter now exceeds its OFFA benchmark. When I arrived five years ago, there were 66.9% from state schools and, from recollection, the OFFA benchmark by 2010 will be 73%. We are ahead of that, which is a deliberate policy of the institution.

Q258 Stephen Williams: I guess that for some subjects your admissions tutors have an over-supply of good candidates coming through. How do you differentiate between those candidates? Is it from their predicted grades or do you interview people?

Professor Smith: We have 24,000 applicants—we are 12th in the country on applications by place. We do not interview everyone because it would be practically impossible to do so. We take predicted grades. An important point about predicted grades is that 67% of grades predicted are accurate within one grade either side. For a student who is taking three A-levels, that means that predicted grades are not perfect. However, out of all A-levels, 84% of predicted grades are accurate within two grades either side. In a sense that is pretty accurate and we use those. The problem we have—different institutions will have different problems and Madeleine will have one story and we will have another partly reflecting our hinterland—is that the crucial determinate of all the issues you are putting your finger on is the entry grades required. As you know, 25% of students get an A and 4% get three As.

Q259 Stephen Williams: Out of the whole population?

Professor Smith: Out of the whole population of 660,000. Those figures stand. The problem is, in some major areas of work, the average entrance grade is A, A, and A. Thus a major issue is how to discriminate between the applicants who have three grade As. Of course, as you also know, 31% of students who get an A come from the independent schools sector even though they educate only 7% of the population. Going back to Susan's point, in the shortage subjects—in the STEM subjects and in some of the languages—well over 50% of As come from the independent schools sector. We have a very nice dilemma to deal with. As an institution, the more we use A-levels and predicted A-levels, the more we move towards certain social groups.

Professor Atkins: There is a slightly different picture at Coventry. First, our ratio of applicants to places has been on average 5:1. We use predicted grades as the basis on which to decide whether or not to make

a standard offer for the course. We find that very low predicted grades correlate with difficulties in completing the course and with drop-outs. We find that there is less correlation between medium and good predicted grades and the ultimate outcome of degree classification. Over the last four years, as A-level grades have gone up, we have raised our grade ask. To give you a precise example, four years ago, for business studies we would have been asking for something like 160 UCAS tariff points. We now ask for 260 UCAS tariff points. That is a change over a four-year period. We put most weight on the predicted grades and actual grades achieved at GCSE and AS. We obviously look with care at the reference and at the personal statement, but the predicted grades and actual grades carry the most weight. We find that the predicted grades where there are external and independent assessments, such as AS and A2, are more accurate for our purposes in predicting whether the student will do well on the course than some of the vocational qualifications, where there is a much higher proportion of internal assessment and there is not necessarily the same guarantee of coverage of the syllabus. We find that GCSE, AS and A2 are better predictors for our purposes, but we do not differentiate our standard offer; once we decide to make an offer, it is the standard offer—we do not differentiate by board or widening participation category.

Q260 Stephen Williams: Let us take a practical example. In either institution, I am sure that there is a subject you could cite—perhaps you have 20 places for English and you get 80 or 100 applicants, all of whom actually meet your entrance criteria, whatever they are. How do you then select which students will be admitted to Coventry or Exeter? Do you interview at that point? What other criteria do you look at?

Professor Atkins: For those subject areas where we are a selecting institution, for example in design, which is one of our major strengths, we look at portfolio. We audition, interview and require portfolio evidence, and that would be the major way in which we would discriminate at that point. In other subject areas where we are a recruiting university, we will make the standard offer.

Professor Smith: At Exeter we use A-level grades to drive up A-level grades/UCAS tariff points as the way of discriminating. The problem, therefore, with many of our subjects, comes when we have a large number of people who are predicted to get three As. We will at that point use the personal statement. We will take into account school performance because that actually is quite a good indicator of where the person fits in that group, but really, of course, every university has some courses for which it selects and some for which it recruits, but for the vast majority of universities, there is always room to move by upping the offer each year. Similarly to Coventry, we have moved our standard offer now so our intake comes in on average with about 395 UCAS tariff points, which of course is three As at A-level and a bit more.

Q261 Stephen Williams: Professor Smith mentioned in his introductory statement that he was involved in a PQA group. What difference do you think that that would make to applicants, if their A-levels were certain rather than predicted? Would it make your admissions tutors' jobs easier?

Professor Smith: It is a great example of a nice technical fix to a problem that actually ends you up in another problem. PQA looks very attractive—what could be fairer than someone coming to admission with their grades? The difficulty is that that militates against widening participation activities. We find that a lot of the issues that we have in driving up the percentage of students coming to us from the lower socio-economic groups is a lot of painstaking work—over two years often—with the school, encouraging them to come and visit the university. Some 68% of our widening-participation of students come from our partner schools in the south-west, which means that we have at least two to three years' engagement with the students. Our worry is that PQA might seem technically nice, but what it might do is specifically disadvantage those who have not got the confidence. The ultimate thing that you find with people's choice with A-levels is a large percentage of students who could go to institutions that demand higher grades but who chose not to because of lack of support, lack of aspiration and lack of ambition. I think that PQA could, if we were not careful, make that problem more difficult.

Q262 Stephen Williams: I know that widening participation is probably the remit of the other Select Committees rather than this one, but none the less we have discovered already that 9% of people get three As at A-level and a significant proportion of those will come from state schools. At least they would be known, I suppose, at the point of application, and maybe they could be sought out by some universities, rather than waiting for them to apply, so perhaps the dynamics would change. I will ask a final question on A-levels and one more question about the introduction of the A*. Do you think that that will enable you to differentiate between top-level candidates, or will it lead to a new set of problems?

Professor Atkins: On the face of it yes, of course, it gives us a greater degree of discrimination. The extent to which certain kinds of school will be able to coach for the A*, as opposed to more general FE colleges, which may not have that facility or staff who are as able to coach in that way, remains to be seen. I have to say that, for example, in mathematics, where the A* will be on core 3 and 4, as I understand it, we will welcome that in mathematics itself and the subjects that require mathematics at A-level. It will mean that we can see how students are doing with the more difficult mathematics. At the moment, because many repeat their AS modules in order to pump up their marks in the first year of the sixth form, we find that the ultimate A-level grade in mathematics can hide a grip that is not that good of the more difficult subject matter in the second year of the syllabus.

Professor Smith: The core issue is the distinction between people getting As and Bs, Cs and Ds, and Madeleine referred to that earlier. That is a very good predictor of their ability to cope. A* will undoubtedly allow us to introduce stretch and to have another tool to measure. However, the core issue from the report that we published in the group that I chaired last week is that the predictions are that 3,500 people will get three A*s, 11,000 will get two A*s, and 29,000 will get one, compared with the 24,000 who get three As at A-level now. The issue between now and A*s coming in is to make sure that we do not see a move up from 31% of As coming from the independent school sector, because as Madeleine has just said, one worry that we have all got, which I shall pose as an open question, is: which schools do you think might decide that their job is to coach people to make sure that they get the A*? That would be a very unfortunate outcome if, in three years' time, we were sitting here saying, "Oh gosh. A lot of the A*s have gone to the independent school sector." What does that do when universities come before you again, and we talk about widening participation? Universities that are pushed to having as many A*s as possible are clearly going to run into that issue.

Q263 Stephen Williams: One final question—we have talked a lot, but it is my fault because I have asked the questions. We have talked purely about A-levels so far. Of course, a lot of people applying to university at the moment, and hopefully in future as well, will have other qualifications, such as the new Diploma, the Baccalaureate, and so on. How many people from both your institutions come in with anything other than an A-level, and are their success rates at application any more or less than with an A-level?

Professor Atkins: I cannot give you the exact proportion, but I can find it out for you and let you have it.

Q264 Stephen Williams: In general terms.

Professor Atkins: In general terms, we accept a high proportion of students who have vocational qualifications, often with one A-level, and sometimes just vocational qualifications. We also take, as you would expect, a fair proportion from access courses. We find that the commitment to study and to be successful is an enormously important part of predicting whether or not those individuals are going to be successful. There are some aspects of vocational qualifications that fit well for the kind of assignments that students have to do with us, and there are some that do not. The main problem is one of coverage. It is often quite difficult for us to know the extent to which a particular syllabus has been covered on some of these vocational qualifications. There are knowledge gaps and inconsistencies from college to college, and often school to school, so we have to do a little more work in getting all students to the same base point. That does not mean to say that they are less good as qualifications; just that they put a slightly different requirement on to us in order to be successful with those students in the first year.

Professor Smith: At Exeter, well over 90% come in with A-levels. The interesting finding that we have is that those who are coming in with the International Baccalaureate do better in firsts and 2.1s than the average, by about 6%, and no IB student has yet dropped out of university. We think that that is worth noting. However, because we need to spend time supporting the Government's agenda of reaching out to people from backgrounds that are under-represented in HE, we are enthusiastic supporters of things like Diplomas, and we will be taking people with Diplomas, and we want to go down that route. Also, as was announced today, we will be working with Flybe as one of the companies on this new skills training, precisely because we have got to attack the problems across the piece. The fundamental problem in the UK is the percentage of kids who leave school at 16 without five GCSEs including Maths and English grade A to C, which is currently 54%. If you read Leitch, you see the problems that those kids are going to have in future.

Q265 Lynda Waltho: I would like to speak directly to the CBI. You will be receiving school leavers, and they will have a range of qualifications: A-levels, the Baccalaureate, national vocational qualifications. Do your members like that, or would it be simpler if there were one qualification to choose from?

Susan Anderson: Clearly, GCSEs and A-levels have a good track record. Employers understand what a GCSE grade C in English and maths means. They know that it delivers a certain standard of literacy and numeracy, and, similarly, they know what an A-level means. In the case that Steve and Madeleine were talking about, where a young person presents themselves for employment with three As at A-level, that is not a problem for employers because we interview people. We interview everybody and would not offer someone a job without doing so. We have no problem distinguishing between able people because we interview them. As people get older and more experienced, their A-level or degree or their level of qualification and experience is supplemented by all sorts of work experience. Employers are used to GCSEs and A-levels. That said, many companies, such as those in hospitality, catering or hairdressing, could see real value in vocational Diplomas. That is why we have supported vocational Diplomas. We think that they enable young people to realise how to apply their literacy and numeracy skills. If you can see how that will be applied, you can see how important it is. So, yes, for employers in those sectors, the Diplomas will be valuable qualifications, as long as it does what it says on the tin—that is the key test for employers. However, because we always interview, we do not have a problem differentiating in the way that our universities do, as they frankly do not have the resources to interview every candidate.

Q266 Lynda Waltho: What about the original outline for the Diplomas as being more useful? Carrying on the Chairman's point about the

possibility of the CBI possibly stymieing the original outline—what would be your answer to that? Did the CBI stymie the original Diploma?

Susan Anderson: If I can return to my opening remarks, we said that the CBI's priority was to raise standards of numeracy and literacy, particularly at 16 but also 18, and to raise the number of students doing STEM degrees. That was the most important priority for us. We felt that the upheaval of replacing GCSEs and A-levels with a new, untried and untested Diploma did not seem to be the right focus when we had so many employers saying that they were concerned about the literacy and numeracy of our young people, particularly those who leave school at 16. The strategy that has been undertaken, to develop the Diplomas and do that in tandem with GCSEs and A-levels, has been the right one. As my colleague Richard said, we now have the Diplomas. We have good Diplomas and those employers who have been closely involved in designing the curricula are satisfied that those Diplomas will deliver the skills and competences that employers need in those sectors. We must ensure that the new system, which has been designed to deliver the Diplomas, teachers with specialist skills and the coming together of the colleges and universities, does deliver. There are some big asks there. We feel confident that they will deliver in the various areas where they are being piloted and trialled, but there are concerns about delivery. Therefore, it was entirely right and appropriate to have a twin-track approach: to retain A-levels and GCSEs as well as to develop the new vocational lines of Diplomas. It was absolutely the right thing to do.

Q267 Lynda Waltho: What about today's announcements—somewhat disparaging in some cases—about what are called "Mcqualifications"? You referred to the Flybe input. What is your view on that? Will that be another complication, or does it show less confidence in the system that you believe your members have?

Susan Anderson: I will be absolutely clear: these are not A-levels and GCSEs; they are workplace qualifications and as such they will be very valuable. Organisations that provide high-quality training such as Flybe, Network Rail and McDonald's, have been delivering very effective training. But part of the problem—this is a common problem in business, where we are spending about £33 billion on training every year—is that only about a third of that training is recognised by qualifications. The qualifications do not reflect the needs, the competences and the skills that business needs. What is happening in a number of initiatives is that the qualifications are reflecting the business needs rather than the qualifications being out here somewhere and not being helpful either to employers or to their employees of whatever age.

So they are very different. Of course we need to ensure that quality is assured, and the various organisations and companies are going through very comprehensive quality assurance and will have to meet exactly the same criteria as an Edexcel or a City and Guilds. That is an important point to make, but the point that I cannot emphasise enough is that they are delivering workplace skills to meet workplace needs. Therefore it does not matter whether they are 16, 18 or 60; a person who arrives with a good qualification reflecting business needs will always be employable. That is our key objective in this initiative.

Q268 Fiona Mactaggart: We have had a bit of evidence that suggests that what our present examination system tests is people's capacity to pass those examinations rather than what I think Professor Atkins was at least hinting at, when she referred to a lack of synoptic understanding of subjects among some of the students who enter Coventry. I am quite interested in an issue that was raised in the CBI evidence and which reflects that. Does our present testing system at A-level and elsewhere properly enable teachers to teach concepts and students to reflect them? If not, what would you change?

Chairman: Who is that to?

Fiona Mactaggart: Professor Atkins, I think. She walked into this in her earlier remarks.

Professor Atkins: What I was hinting at was, indeed, a possible tension which we feel is potentially developing between an assessment system through the school period of a young person's life, where there is a great deal of teaching to the test and the ability to repeat AS modules, in particular, again and again to try to improve on grades, and what we would wish to provide as a university-level education. In vocational courses in particular we are trying to achieve graduates who are very good indeed at problem solving, with messy, real-life problems. That requires a deep understanding and deep learning, and the ability not just to have a selection of techniques that you have learned by rote and learned to apply by rote, but to select the appropriate tools and methodologies for that particular problem because you can understand the connections between them and you can see that that might well be relevant in trying to tackle the issue that is the subject of your group work, your assignment or whatever it might be. We are slightly anxious that the atomisation of AS/A2 and potentially of Diplomas—although the extended project will be extremely helpful there—means that some students arrive with us believing that their university life will be chunked up like that as well and that we are going to teach them to the test, whereas we are taking live projects from employers, business and the private and public sectors and encouraging them to work in teams on that kind of activity-led curriculum. They find that transition quite difficult. It does not mean to say that that they cannot do it, but it does mean that we have to teach in a rather different way to begin with in order that that synoptic understanding is developed and that understanding of connections between tools, techniques and methodologies is really in place.

Q269 Fiona Mactaggart: Does the CBI want to say anything about that?

Richard Wainer: Our members have not raised that concern with us. As Susan said, their issues are about ensuring that young people are literate, numerate and employable. With support from the former Department for Education and Science, we looked at exactly what is meant functional literacy and numeracy. What skills, activities and tasks do they want young people or any of their employees to be able to perform to the basic level? We defined literacy and numeracy and we are glad the Government took that up and that they are now developing functional literacy and numeracy modules to be included in the GCSEs and the new Diplomas. Their main concern is making sure that young people can exhibit those competencies.

Q270 Fiona Mactaggart: I wonder whether that is one of the reasons for the shortage of students succeeding in the science subjects, because they are linear subjects, where prior understanding is very often required in order to do further work. That is reflected in what Professor Atkins was saying about anxiety and about people beefing-up the first bit of their AS results and having a grade that overall does not reflect their real capacity in mathematics, for example. I am concerned that physics, mathematics and modern foreign languages are all subjects in which the present way of teaching to the test might not help the development of students' understanding of that subject. If that is the case, it could be a reason for the low number of entrants. If you think there is any truth in that, what would you do about it?

Professor Atkins: We would say rather less testing through the sixth form years and the nature of the testing to be more problem-based and synoptic rather than chunked up into units, with just that bit of knowledge tested and then put to one side.

Professor Smith: Very briefly, I think the issue goes back a lot earlier. A lot of work must go on about what happens pre-14, when students make choices about what they are studying. Our concern about A-levels is that they tend to benefit the middle class because those parents know how to make sure their children are re-taking the modules, so you see an effect there. The problem we have with A-levels is that students come very assessment-oriented: they mark-hunt; they are reluctant to take risks; they tend not to take a critical stance; and they tend not to take responsibility for their own learning. But the crucial point is the independent thinking. It is common in our institution that students go to the lecture tutor and say, "What is the right answer?" That is creating quite a gap between how they come to us with A-levels and what is needed at university.

Fiona Mactaggart: May I ask Susan Anderson that question?

Chairman: We are running late because of the Division. I will therefore ask my colleagues to speed up a little, because we are still on section 1.

Susan Anderson: I want to go further back than A-levels. When we asked our employers what they understood by literacy and numeracy, some of it was pretty fundamental stuff. Some workers who come into the workplace cannot do mental arithmetic, do not know their multiplication tables and cannot work out fractions and percentages, but those things are easy to test. Sometimes schools assume that having learnt those things once it is for ever in the brain, but that is not always so. Many companies tell us that people cannot spot errors or rogue figures because they cannot do mental arithmetic. Those are things that you can test pretty accurately and, linking to the Diploma point, if you can apply them and understand why they are applied and see why they are important it makes them more relevant. That is especially so if you can apply them in a construction or a retail environment, for example, if children know that they want to take that particular vocational route. Those skills are pretty basic and if children are not getting them right at 16, 14 or 11, that is a good indication for employers of whether schools are getting the basics right.

Q271 Fiona Mactaggart: I wonder whether the responses that Professor Smith and Professor Atkins gave imply that we should not allow retakes for part 1 of AS-levels.

Professor Atkins: Some universities do not permit that. I think I am right that some medical schools do not permit the retake mark to be included, but that would have to be checked. There is concern about this culture. In a sense, you are saying that students work very hard in sixth form. They absolutely do and that is partly because tests come up every three or four months. The pressure for those tests is enormous. In the first year of sixth form, students are told that they have to get close to 100% in maths because next year they will not get such a high mark and they have to get their average up. That is an enormous pressure. We do not doubt the amount of sheer hard work that is going on. However, that would perhaps be better directed if students were saying, "This subject is really exciting; we have time to think and to get enthusiastic about it," rather than, "We have to do another test paper on Friday." There does not seem to be as much time to explore the subject as there used to be.

Professor Smith: That is why a lot of universities do additional specialist admissions tests that measure competence. At our university, we use several of the major national tests. The Department for Children, Schools and Families is aware of this problem. There are proposals to reduce significantly the ability to retake too many things. I would personally welcome that, although it is not as easy as it sounds at first. Clearly, it is good to allow students to improve their grades by increasing the work load and retaking exams. I think that there is a difficult balance to find on this issue.

Q272 Mr. Heppell: I have a brief question. There is a big demand for students studying STEM subjects, and particularly for science graduates. You seem to be saying that we should tackle the problem earlier on in school, with more people taking triple science and so on. Could this be part of the problem? I think that this is more of an issue for the CBI than the universities. Perhaps employers do not give enough status and reward to people in those jobs. If they were getting more rewards, those jobs would be

more attractive. At some stage, people are making choices. Would they be more likely to make choices towards science if they thought that the rewards were better?

Susan Anderson: Can I answer that question in two parts? First, the starting salary for someone going into the City is about £38,000. The sorts of people who can work for large investment banks or large accountancy firms are engineers and physicists. We need physicists and engineers in manufacturing, but they are also in demand in occupations and professions that need highly numerate people. On average, the starting salary of an engineer is about £24,000. Compare that with the starting salary of somebody with a general arts degree. They could not do those jobs. They might get very good jobs, but they will be in sectors such as retail, which ask for general graduates. Graduates in physics and chemistry are getting very good starting salaries. The problem is that there are not enough of them, so employers in the engineering sector see a number of potential recruits going off to investment banking. I do not think that that is a bad thing. It tells us that we need more STEM graduates, not that people in engineering ought to pay them more. Secondly, you are absolutely right that we in business need to get those sorts of facts out to young people so that they are aware, when making their choices at 14, 16 or 18, of the very well-paid jobs that are open to those who do STEM degrees. Employers must convince students that they offer good jobs and convince able students that they might like to do an apprentice programme rather than a degree.

Chairman: I will make myself very unpopular because we only have two more sections. We have covered some parts of the other sections, but we must move on. I take your point about needing good mathematics to go into the City; I understand that you have to count in French up to 4 billion, at least, to qualify these days. Dawn, you are going to ask some leader questions on the knowledge and skills deficit.

Q273 Ms Butler: Some of this might have been answered in the first section, but, in the CBI report, you have said that too many people leave school without necessary literacy and numeracy skills. Does the current qualifications system for school leavers provide sufficient opportunities for candidates to equip themselves with the skills and attributes that business people or universities are looking for?

Susan Anderson: As we have said, we have sought to feed into the design of qualifications by defining what literacy and numeracy means in an employment situation. We think that a basic level of literacy and numeracy probably equates to a Level 1. So, the new functional skills elements of the Diplomas, GCSEs, and A-levels will help employers to have some confidence that young people are coming out with basic literacy and numeracy skills. However, that is not enough. We must have, as employers, the equivalent of a grade C in English and maths. We must have much higher standards of literacy and numeracy than the very basic levels that we have talked about. I think that we have seen progress over the last 10 to 20 years in improving literacy and numeracy levels, but we are saying that they are not anywhere near where we would want them to be. When 54% of school leavers do not have 5 A*–C in both English and maths, we cannot afford to be at all complacent in looking at what is happening with literacy and numeracy. There are improvements, but it is the rate of progress that we have more concerns about.

Q274 Ms Butler: Do you think that the current plans will improve the rate of progress for school leavers?

Susan Anderson: It is very important that we are measuring what is happening so that we know how many children get A to C grades in English and maths. It is important that we measure schools according to that key target. The literacy and numeracy modules for the new Diplomas will help, and the fact that they are being delivered in an applied way will help more young people to see the value of those essential skills.

Q275 Ms Butler: Professor Smith, I see that you are agreeing?

Professor Smith: Absolutely. Diplomas provide a great opportunity of another route to attain those skills, which appeals to students who might be put off by the "academic route". Diplomas are a very good attempt to deal with that, because they are designed with the involvement of employers, so skills are relevant. It is another way of trying to deal with the 54% of school leavers who do not leave at 16 with five GCSEs at grade A to C including Maths and English. As the Leitch review pointed out, the jobs for those people will simply not be there by 2020. I see Diplomas as a very good way of creating parity of esteem, but by a different route.

Q276 Ms Butler: You talked earlier, Professor, about those people who have A*s being able to develop personal and study skills further. Do you think that Diplomas will also help young people to develop them?

Professor Smith: The truthful answer is that we have to wait and see. The good news is that with everyone—employers, universities, colleges and schools—being involved in the design of the Diplomas, there is a good opportunity for awareness of those matters. The skills modules and sections of the Diplomas offer a clear indication that that is on the agenda from the start. So, with regard to design, it is hopeful that those people have been involved from day one and are trying to achieve those ends. Clearly, we will not know whether they have succeeded until the first people complete the Diplomas, but they are being designed very much with meeting that deficit in mind.

Q277 Ms Butler: Do you think that anything could be added to the design to make us feel more secure that we are equipping young people with those study skills?

Professor Atkins: From what we have seen so far—it is still early days for the Diplomas—some of the Diplomas have got the mathematics right. The engineering Diploma is an example of that, as is the manufacturing Diploma that is coming through at the moment. I am not sure how widespread that will be in the other Diplomas as they are developed. The point that I would make on the IT side is the one that I made earlier: it is not that young people need greater facility with IT, but that being able to appraise IT sources critically will be increasingly important, whether in employment or higher education. I am not sure that the Diplomas have quite pinpointed that skill as important for the future.

Q278 Ms Butler: That is interesting. Thank you. My final question relates to the CBI's report, which states that employers often find that they need to remedy deficits in the basics. Do you engage in any remedial activity to bring school leavers up to the necessary standards after entering university or a programme—that question is to all on the panel?
Chairman: One from each group, because we are running a bit late.
Richard Wainer: Our surveys show that around 15% of our members have to provide remedial basic skills training. Although that is probably through skills for life courses, which the Government provide, there is certainly a lot of frustration among our members about having to provide that sort of training.
Professor Atkins: Yes.
Professor Smith: Yes.
Chairman: As I listen to this, I am thinking about how some of your members go to the media and say that children who have not got A to C in mathematics and English are illiterate. It is not true, is it? I sometimes think that members of this Committee should sit down to take the tests to see what our levels are and bring us back to reality. I would be interested to know what my own skills are in that department—I was never any good at maths.

Q279 Annette Brooke: Picking up on that point, could you please clarify whether the CBI's members are still concerned about the literacy and numeracy skills of those who have A to C in maths and English, and will they have what you require if they have their five A to Cs?
Susan Anderson: Generally, the answer would be yes, but perhaps it is a bit similar to what we are saying about engineering. For some employers—engineering is a good example—we might well find that they need more depth when it comes to particular engineering applications. That sort of issue can arise. The other thing that I would say from the employer community is that employers are already always training people, whether on or off the job. Therefore, it is not that you have put those skills in a box and never used them, but they are always being applied by people in their work places.

Q280 Annette Brooke: Do you have particular concerns about the lower grades of GCSEs, which is a qualification that you want to carry on with? Are we not providing what would be best for the individuals who are going to get Fs and Gs?
Susan Anderson: As Professor Smith said, many of those individuals might find that a Diploma is a much more appropriate and engaging type of qualification. At the end of the day, however, literacy is literacy and numeracy is numeracy. They have not changed much over the last 20 years. IT has changed considerably, but I do not think that employers demand the same skills for IT. In some cases, employers' demands are going up because, as Steve has said, there will be fewer jobs for people with very poor literacy and numeracy skills. Employers' expectations are rising, but they do not necessarily think that standards have been slipping or that qualifications are not now worth what they were 20 years ago. Certainly, expectations are rising.

Q281 Annette Brooke: I think that that is a really important point, which I will come back to in a moment. I just wanted to look at the table that was in your paper, which shows employers' dissatisfaction with school leavers' key skills. In fact, self-management scores more highly than dissatisfaction with basic literacy and numeracy. What is wrong with our current examination system that leads employers to be dissatisfied with self-management skills?
Susan Anderson: I have to say that self-management skills can be sorted out pretty easily in the workplace. Employers do not see basic literacy and numeracy as their responsibility, however, and they are rather harder to fix. Helping someone manage their tasks on a weekly or daily basis is something that employers can pretty readily fix. It is rather hard to retro-fix when it comes to literacy and numeracy.

Q282 Annette Brooke: I am a bit surprised that it is highlighted so much in this table. It seems as if we are saying that these young people cannot do this, that and the other and you are saying, "Well, of course that is the case." Perhaps we could apply that same point at university level. I suppose that self-management is also very important. How do you take that on board when you are assessing to whom to make your offers?
Professor Atkins: When we look at the personal statement and the reference we get some idea of whether the applicant has those skills. On the whole, we find that references from schools give a slightly more authoritative view of the young person than those from colleges. School teachers seem to know their sixth formers slightly better. When we go through induction and the first month of the first year we spend a lot of time with all our new entrants going through self-management, study skills, balance of activities and so on. We regard that as part of our responsibility and it is important to do that. I do not think that we regard it as hugely problematic, but we do think there is a transition from school sixth forms particularly, where most of the day is divided up on a timetable and there is very

little choice for the students in what they do, to a university setting, which is very different. There is a transition that has to be worked through there, but we think it is our responsibility to help young people to do that.

Q283 Annette Brooke: When examination results come out we celebrate rising standards and the achievements of our young people, but immediately the cold water is poured on it by claims that things are easier and that standards are not really rising. Could I ask both halves what comment you would make about standards 20 years ago? In things like the self-management aspect, has there been a change over that period in that we are perpetually testing our young people in school and not leaving them to their own devices so much? I am slightly answering the question myself.

Professor Smith: I think the answer is mixed. There is some evidence in the public domain that in many of the sciences the amount "learnt" is of a different order, and there is some work coming out saying that a grade D 20 years ago would get a grade B now. That has to be balanced by the fact that people come along with a different skill set and the ability to apply knowledge. In that way we now know what they come with in terms of deficits. We offer them subject-based teaching. We also offer every student a series of study skills programmes and individualised support, if necessary. As long as we know, we can deal with it. It is not as simple as saying better or worse. The content has changed and it certainly is the case that more students are getting As. The number is rising by about 1% a year. That is in part because they can retake, and in part because the teaching in schools is probably now more linked to exams. It is a complicated issue and a genuinely mixed picture.

Q284 Annette Brooke: May I ask the CBI whether that has had any impact on self-management and independent thought?

Susan Anderson: Students are right to celebrate because they work very hard. The vast majority achieve very good results. That said, employers' expectations are rising. We have to recognise that there will be less demand for people with very poor literacy and numeracy skills. Similarly, there will be a growing demand for people with high-quality STEM skills. We can over-obsess about self-management, but it is an important employability skill. However, as my colleagues from the university have said and as I would say, these skills can be learned relatively easily. They are not as fundamental as not having enough physicists, chemists or, indeed, students with high levels of literacy and numeracy.

Q285 Annette Brooke: I have one final question. To return to a point you made, Professor Smith, you did not appear terribly enthusiastic about A*s and PQA in relation to widening access. Hard-working pupils from independent and grammar schools feel quite hard done by when somebody from a comprehensive school has a place that they might have had? How do you get the right balance between widening access and being fair?

Professor Smith: That is a very difficult issue and I do not think there is a technical solution. Admissions tutors make judgments. There is some stunning evidence from a report in 2004 that shows that of a cohort of 76,000 students tested, students from independent schools who got 3Bs are outperformed at final degree result by state school students who got 3Bs on entry, so much so that you would predict that the students from the state school system came in with two As and a B. For us, it is a difficult one. A-level grades are a given, but we have to take into context the situation in which the student has acquired those grades. For universities, it is a constant battle to try, first, to work with schools to encourage students from poorer backgrounds to apply and then to try to do what we can to ensure that there is a level playing field at the margins in terms of taking into account school performance. I do not think we will ever get it right but clearly we cannot say, on the one hand, that aspiration is the key because we have A-level results to contend with and, on the other, that A-level results are very highly correlated with social class. It is the job of universities at the margins to try and find a way through this so that people who can benefit most from going to university get offered the place. I do not think we ever get it right; we never will, but there is no nice, neat rule that we can use to say, "Well, it's just A-levels or just the school type."

Chairman: Lynda, I think you have a quick question on 14–19 performance.

Q286 Lynda Waltho: Wish-list time. We have heard from many witnesses about the reforms they would like to see. What about you?

Professor Atkins: We would like to see less testing through GCSE and the sixth form or college years 1 and 2, more synoptic testing and opportunities in those subjects that do not have it at A2 for a more extended piece of work that enables the students to engage in more complex areas of the subject than perhaps they do now. We wish the Diplomas well; we hope that the mathematics and numeracy content remains strong across the piece and that they do not become too atomised as they go through the design phase.

Susan Anderson: Raising literacy and numeracy must be the key and running that a very close second would be increasing the number of STEM graduates, which we recognise must be addressed at school level. It is not just about what the university is doing; primarily, action at the school level is needed.

Professor Smith: Overwhelmingly, I want to support the Government's raising of the target for five GCSE A-Cs, including maths and English. The last 10 years have seen a very healthy improvement of about 10% in those figures. But still, comparatively, it does not put the UK in a very good place. It is 24th out of 29 on OECD figures for the number of kids still in school at 17. That is the issue. Remember, 92% of students who get two A-levels go to university and 39% of 18-year-olds take A-levels. You cannot get

that much higher when only 46% get five GCSEs A to C including maths and English. For me, that is the No. 1 requirement of the system in order to have social equity, which is precisely why Diplomas have the support of organisations such as the CBI and the universities.

Chairman: This has been a very good session. I would like it to go on as there are many more questions that we would like to ask such a good group of witnesses, but the Division did rather disturb the pattern of questioning and answering, and we have to go on to the second section. I thank you, but please keep in contact with the Committee. If there are issues that you think were not covered sufficiently in the oral session, will you write to us? Often the best answers are the ones that you might think of as you are going home on the train or on the bus. Thank you.

Memorandum submitted by the Department for Education and Skills (DfES)

Introduction

1. National Curriculum assessment, with public examinations at 16–19, provides an objective and reliable measure of the standards secured by pupils at crucial stages in their development. The focus on English, mathematics and science in the curriculum and assessment at Key Stages 1–3, and in future on functional skills in the curriculum and assessment for 14–19 year olds, reflects their importance to the future prospects of children and young people in education and the world of work beyond.

2. The system of National Curriculum assessment was developed on the basis of expert advice at its inception to complement arrangements already in place for public examinations at 16 and beyond. It recognises that any assessment system makes trade-offs between purposes, validity and reliability, and manageability. It places greatest emphasis on securing valid and reliable data about pupil performance, which is used for accountability, planning, resource allocation, policy development and school improvement. In short, it equips us with the best data possible to support our education system.

3. The benefits brought about by this system, compared to the time before the accountability of the National Curriculum, have been immense. The aspirations and expectations of pupils and their teachers have been raised. For parents, the benefits have been much better information not only about the progress their own child is making but also about the performance of the school their child attends. And for the education system as a whole, standards of achievement have been put in the spotlight, teachers' efforts have been directed to make a difference and performance has improved. The public has a right to demand such transparency at a time of record investment in education.

End of Key Stage Assessment and Public Examinations at 16–19

4. The National Curriculum sets out a clear, full and statutory entitlement to learning for all pupils, irrespective of background or ability. It determines the content of what will be taught, promoting continuity and coherence throughout the education system, and defines expectations about levels of attainment and progression throughout the years of compulsory education. It promotes public understanding of, and confidence in, the work of schools by establishing national standards for all pupils. It provides a framework for the continuous monitoring of pupils' performance which is a key feature of many successful schools' strategies for improving teaching and learning.

5. Until end of key stage assessment arrangements were introduced, the only measure understood and accepted by the public was general qualifications examinations, taken by most pupils at the end of compulsory education. There were no objective and consistent performance measures which gave the public confidence about expected standards in primary schools or the intermediary years.

Key Stage 1

6. Assessment arrangements at Key Stage 1 have, since introduction, always been more flexible than those at other key stages, and changes in 2005 developed that approach further. The focus on teacher assessment recognises that, at age seven, children's performance on a specific occasion may be less representative of what they can do, and that wider evidence of attainment is better gathered over a longer period. Teachers make assessments for reading, writing, speaking and listening, mathematics and science. They administer national curriculum tasks and tests in reading, writing and mathematics, which provide evidence to inform their overall assessment of a child's progress. But only the teacher assessments are reported. Teacher assessments are moderated by local authorities to ensure that they are made on a consistent basis and so provide a suitably reliable measure of performance. Schools publish their Key Stage 1 performance data. The Department publishes national summary Key Stage 1 results but does not publish school-level data.

Key Stages 2 and 3

7. National Curriculum assessment at Key Stages 2 and 3 comprises tests in the core subjects of English, mathematics and science, and teacher assessments in core subjects at Key Stage 2, and in core and foundation subjects at Key Stage 3. Schools report teacher assessments alongside test results. There are no moderation arrangements for teacher assessments at these key stages as externally-marked tests provide the basis for assessing all pupils in the country on a consistent basis.

Assessment at age 16 and over: GCSE, A levels, Diplomas and the International Baccalaureate

8. The General Certificate of Secondary Education (GCSE) and Advanced level (A level) qualifications are internationally respected qualifications at the heart of our assessment system at 16 and after. We have been very clear that both are here to stay. They are a proven and valuable method of recognising achievement, particularly at Level 2 (GCSE)—the level which the Leitch Review recommends that 90% of the adult population of the UK need to reach if the UK is to provide growth, productivity and social justice in a rapidly changing global economy.

9. The GCSE is currently the principal means of assessing standards at the end of compulsory schooling at age 16. Although it is not compulsory for pupils to be entered for GCSEs, approximately 96% of 15 year olds are entered for one or more full courses each year. They are assessed through a combination of coursework and terminal examination and are graded on an eight-point scale from A*–G. GCSE grades are awarded on a criterion-referenced basis: to be awarded a grade C, the candidate must demonstrate the qualities associated with that grade. The A* grade was introduced in 1994 to recognise outstanding pupil achievement.

10. A levels are one of the main routes into higher education and employment. They were radically changed in 2000 when a completely revised advanced level curriculum was introduced to increase breadth (particularly in the first year of study) and allow students to monitor attainment and make informed decisions about their future learning.

11. A levels are modular qualifications and are assessed as young people proceed through the course, rather than only being examined in a single examination at the end of a two year course. Some A levels also have a coursework element. They are graded on a five point scale from A-E. We recently announced the introduction of a new A* grade for A level from 2010, which will reward excellence in a similar way to the A* at GCSE.

12. We recognise that GCSEs and A levels are not right for every student. By 2013, every young person will have the choice to pursue one of 14 Diplomas, which will provide a new way of assessing standards at Levels 1, 2 and 3. And by 2010, at least one maintained institution in every local authority (fewer in London) will be offering the International Baccalaureate.

Assessment and Accountability

13. Test and exam results are published every year in the Achievement and Attainment Tables, which are an important source of public accountability for schools and colleges. The publication of threshold measures of performance is a strong incentive for schools and colleges to ensure that as many pupils/students as possible achieve the required standard, particularly, at Key Stages 1-3, in the core subjects of English, mathematics and science. This emphasis on the "basics" has been strengthened beyond age 14 by the inclusion of a measure incorporating English and mathematics in the Key Stage 4 tables.

14. The inclusion of Contextual Value Added (CVA) in the tables provides a more sophisticated measure for comparing the performance of schools, based on what we know about the effect of particular pupil characteristics on attainment and progress. CVA brings us closer to a common indicator allowing us to contrast schools' effectiveness.

15. Used together, threshold and CVA measures provide a powerful tool for school improvement and raising standards across the education system, enabling us to track changes in performance over time nationally and locally, and at school and individual pupil level:

— At national level, performance data derived from tests and public exams enable Government to develop policies and allocate resources. For example, the first Key Stage 2 test results showed that 49% of pupils were not reaching the standard expected for their age in English, while 48% were failing to do so in maths. This provided the impetus for the National Literacy and Numeracy Strategies which have done so much to raise standards in primary schools.

— At local level, performance data have been integral to target-setting and have enabled us to focus on areas of particular under-performance.

— At school level, performance data have provided a basis for the judgements of inspectors and challenge from School Improvement Partners. School-level data published in the tables and School Profile also help parents to make informed choices about their children's education. Used in this way, performance data is an important lever for driving school improvement. As described below, performance data and supporting information can also provide an important tool for schools in devising their strategies for improvement.

— At pupil level, assessment provides clear and widely-understood measures of progress for every child, supporting the personalisation of teaching and learning by enabling children's future work to be planned to meet their needs and help them to fulfil their potential.

Standards over Time

16. The strength and validity of the accountability regime requires us to ensure to the best of our ability that tests and public exams measure pupil performance against standards that are consistent over time. QCA is responsible for ensuring that standards are maintained over time and its processes for doing so in relation to National Curriculum Tests were found by the independent Rose Panel (1999) to be robust and to bear comparison with best practice in the world.

17. For public examinations, the 1996 SCAA/Ofsted report, and subsequent follow up reports, showed that standards at A level have been maintained for at least 20 years. These reports concluded that the overall demand of subjects studied remained broadly the same between 1975 and 1995, although an increase in breadth of coverage led to a reduced emphasis on some topics.

18. The Independent Committee on Examination Standards chaired by Dr Barry McGaw, Director for Education at OECD, published its findings about A levels in December 2004. The report concluded that:

— no examination system at the school or other level is so tightly or carefully managed;

— strategies for maintaining comparable examination standards across awarding bodies are adequate to the task;

— the awarding bodies have broadly consistent and well-regulated systems for setting question papers, managing marking and awarding grades; and

— QCA has robust systems in place to monitor and regulate the work of the awarding bodies.

Assessment and Teaching and Learning in Schools

19. Schools are the point where the key strands discussed so far—end of key stage assessment, accountability and standards over time—come together with assessment for learning (or formative assessment) to drive improvement in the performance of individual pupils and groups of pupils. We provide schools with a range of tools and support to enable them to make the most of this interface.

20. RAISEonline is a web-based data analysis system developed jointly by the Department and Ofsted. It enables schools to see how their pupils' performance in tests compares with pupils in similar schools, to track the progress of individual pupils, groups of pupils and year-groups, and to evaluate the effectiveness of teaching. That information is a powerful tool for planning development and improvement, and enables schools to set challenging targets for improved future performance across the school. RAISEonline also supports the school self-evaluation process, which is a key element in Ofsted's school inspection framework.

21. The value of the end of key stage assessment (or summative) data published in the tables and in RAISEonline can be enhanced by the quick provision of marks for each question and sub-question to support analysis against attainment targets. These data can be analysed in RAISEonline to identify the curriculum areas where pupils performed better or worse than the national average.

22. Test scripts are returned to schools and they use evidence from them to identify strengths and weaknesses in individuals, teaching groups and subject teaching. This summative evidence at the end of each key stage should complement continuous teacher assessment. Pupils should be tracked against attainment targets and progressively achieve individual learning goals which relate explicitly to National Curriculum levels.

23. Teachers can use a number of tools to support their summative assessment judgements, such as tasks provided through the Assessing Pupil Performance (APP) materials, optional tests and 'P scales' for SEN pupils working towards Level 1 of the National Curriculum. Summative assessment is in turn a key piece of evidence for effective assessment for learning (AfL). Effective AfL is about teachers working with pupils and their parents/carers, using a range of evidence to establish where pupils are in their learning, where they need to go and how they get there. Where this is effective, the pupil and their teachers share an understanding of the progress they have made and their objectives for moving forward.

24. To help schools deliver AfL in the round, the APP materials are supported by guidance on the techniques that support AfL (for example setting targets with pupils or supporting peer and self-assessment). We have also developed more tailored Intervention materials and "Progression Maps" designed for use with pupils who are under-achieving. The Intervention materials and maps enable teachers to refine their summative judgements of pupils below expectations to pinpoint specific weaknesses and then plan teaching

that will tackle those issues. APP, AfL and Intervention materials are all delivered through the National Strategies, who develop the materials in partnership with the QCA. AfL guidance has been available in Secondary Schools since 2004, APP materials for English and maths are available in Secondary Schools and will be released in primary schools in 2007–08. Intervention materials were rolled out in 2006.

Criticism of Testing

25. There are three criticisms which are most often levied by public commentators at the current approach to assessment and accountability. The first is that it leads some teachers to teach to the test. The second is that it causes schools to narrow the curriculum. The third is about the burden testing places on schools and pupils.

Teaching to the test

26. The best preparation for any test is to ensure a pupil has the deep knowledge and understanding of a concept, or the extended experience of practising a skill, which permits them to demonstrate that knowledge, understanding or skill proficiency and in response to a variety of possible test items.

27. At the same time, and as with any exam based around set criteria, it is certainly the case that a child who is helped to understand what the markers are looking for and how to present answers accordingly is likely to do better than a child who faces the test unprepared. Preparation for assessment should be wholly integrated into the classroom experience for pupils and not at the expense of teaching the wider curriculum. Teachers should agree with pupils their targets for the next stage of progress, discuss with pupils what they need to do to reach the next stage and constantly reflect that progress in their teaching. There should be continual pupil tracking based on such assessment, benchmarked periodically by formal tests such as the optional tests provided by the QCA. There should then be no "mad dash" to the tests in year 6 and year 9 to make up for lost time. Pupils know what they can achieve and have the right teaching and support to achieve it in the statutory assessment.

28. The teacher who prepares pupils for a test without developing the deeper understanding or more extended experience required may be fortunate enough to enjoy some short-term success, but will not be likely to maintain that performance over time.

Tests and the curriculum

29. Within the assessment regime, we make no apology for the focus on the core subjects of English, maths and science, which encourages schools to prioritise these subjects, because they hold the key to children's future success in the classroom and in the world of work beyond. For example, pupils going to secondary school at the age of 11 can be seriously disadvantaged if they lack secure standards (Level 4 or better) in English and maths. It is not simply a matter of the value of these subjects for their own sake. Pupils' study of subjects like history and geography, for example, will be hampered if they cannot "show understanding of significant ideas, themes, events and characters" through reading or "develop their own strategies for solving problems and use these strategies both in working within mathematics and in applying mathematics to practical contexts" (as set out in the English and maths Level 4 descriptions). There is nothing that narrows a pupil's experience of the curriculum so quickly as a poor preparation for the level of literacy and numeracy that the subject demands.

30. Science should be maintained as the third national priority. It helps pupils to explore the world around them and understand many things that have relevance to daily life. It is a key element of preparation for life in modern society and is essential to our future economic prosperity.

The burden of assessment

31. The statutory assessment system demands comparatively little of the child in the 11 years of compulsory schooling. Assessment at KS1 should be carried out as part of normal lessons. The child will not necessarily recognise a difference between the formal tests and tasks s/he completes for other classroom exercises. The tests can be taken at any time throughout the year and just to the extent necessary to enable the teachers to be secure in the judgement of which level the child has reached. Assessment at KS2 involves one week of tests in May, most lasting 45 minutes and in total amounting to less than six hours of tests under secure examination conditions. Assessment at KS3 is similar, involving one week of tests normally an hour in length and amounting to less than eight hours. Although most children take the tests in the years they turn 11 and 14, the requirement is that they are assessed at the end of the completion of the programmes of study in the key stage. That means that schools can enter children for tests at an early or later age depending on their development.

32. We know that some children can find examinations stressful, as with many other aspects of school life. All effective schools will help anxious children to address the demands of expectations of performance in tests.

33. Throughout the evolution of the assessment system we have kept in balance the costs of assessment and the workload expectations on schools and teachers. Parents have a right to expect that the standards of their children and the school are assessed in an objective way, free from bias or influence. External assessment minimises the workload demands on classroom teachers. Other approaches to ensure objectivity and reliability introduce workload burdens.

34. At GCSE level, we are responding to criticisms that coursework is often repetitive and burdensome, does not add educational value, is open to cheating and is difficult to verify. As there have been changes in education and technology since GCSE was introduced, QCA has assessed each subject to determine whether coursework is the best way to assess the learning outcomes. In the future we intend to assess some subjects purely through external examinations, and develop controlled assessments in other subjects. QCA will be consulting with key stakeholders on what controlled assessment will look like, and will be consulting more widely on the GCSE criteria this summer.

35. We are also taking steps to reduce the burden of assessment at A level on students and on teachers, without compromising the standards of the qualifications. Now that early concerns about the impact of Curriculum 2000 have subsided as the modular approach has bedded down, we are reducing the number of units for A levels from six to four in the majority of subjects. This will reduce the assessment burden by a third, reduce costs and address exam timetabling difficulties: but the reduction is only in assessment, not in content or standard. QCA has also considered the burden of coursework in individual subjects and the cumulative effect across A-level programmes. From 2008 only those subjects containing particular skills and subject knowledge that can not be assessed through an examination will continue to be internally assessed.

Evolution of Assessment

36. As our response to criticisms about GCSE and A-level assessment shows, the system has constantly evolved to meet changing needs and it will continue to do so. The two key areas of development at present focus on progression and on further strengthening of 16–19 assessment.

Evolving assessment policy to drive progression

37. At present our assessment and accountability system is largely based around the achievement of thresholds. While we will continue to look for improved performance at threshold levels, our increased emphasis on a personalised approach to learning and on ensuring that every child progresses well in their education suggests we should also hold schools accountable for the rate at which pupils progress. The *Making Good Progress* pilot, which will begin in nearly 500 schools in September, will look at how we best we can do this. An important part of the pilot will be to investigate whether different approaches to assessment and testing can encourage sharper use of assessment, better pupil tracking leading to better planning by schools, and improved progression to the next key stage. At the same time it will need to maintain the accountability, consistency and reliability of the current system.

38. In pilot schools, Key Stage 2 and Key Stage 3 English and mathematics teachers will be able to enter children for an externally written and marked test as soon as they are confident (through their own systematic assessment) that the pupil has progressed to the next national curriculum level. The tests will be available twice a year at all levels to pupils in all year groups so, for example, an exceptional pupil in year 4 could be entered for a Level 5 test, whilst for a small minority of pupils in year 6 a Level 3 test would be appropriate. The tests will be shorter than the existing end of Key Stage tests because they are confirming whether a pupil has reached a specific level rather than testing across a range of levels.

39. A key feature of the pilot will be to improve the way teachers track and assess pupils' progress. It provides a valuable opportunity to explore how we can make formal testing arrangements work together even more effectively than at present with assessment for learning by looking at how we can better exploit the formative potential of tests. The pilot tests will be taken only when a teacher assesses a pupil as ready to achieve the next national curriculum level, placing the emphasis on where a pupil is throughout a key stage, rather than the end of the key stage. The tests should become part of a continual learning process, engaging parents, pupils and teachers in focusing on the next steps learning. An important area we will be looking at in monitoring and evaluating the pilot is how close teachers' judgements about the level a pupil has reached are to the test results. In principle, if teachers' assessments are accurate, and if they are only entering pupils when they are ready, pupils are much more likely to experience success in the tests.

40. At a national level, the model being piloted will generate rich data about pupil and school performance. It will provide a much clearer picture than at present about how pupils progress through key stages, including which groups of pupils fall behind at which points. For schools, the combination of termly teacher assessments, and the opportunity to have those assessments confirmed by tests as pupils progress through levels, will be a valuable means of tracking and improving the progress of individual pupils.

16-19 Assessment

GCSE

41. From this September, we will be piloting new GCSEs in English, maths and ICT which include functional skills in these areas: once these are rolled out from 2010 it will not be possible to achieve a grade C grade or above in English, maths or ICT GCSE without demonstrating the relevant functional skills. This will mean some significant changes in assessment techniques and we have a programme of activity in hand to ensure that the workforce is appropriately trained. Both assessment and workforce development will be developed further as part of the pilot, taking into account the burdens on learners and teachers.

A levels

42. Around 4% of the age cohort achieves three or more A grades at A-level. We believe that more can be done to stretch and challenge A-level students, particularly our brightest students, and to provide greater differentiation for universities which have large numbers of applicants for popular courses. Following consultation with HE representatives, QCA began pilots of tougher questions from September 2006. QCA has revised its A-level criteria to require more open-ended questions and reduce atomistic assessment, and is currently in the process of accrediting new specifications against these criteria. The standard required to obtain an A grade will remain the same but the strongest performers against the new assessment will be rewarded through the introduction of the A* grade.

Diplomas

43. The phased introduction of Diplomas will begin in 2008 and will offer greater choice to learners by providing a strong focus on applied learning within a broad educational programme. They are designed to improve both participation and attainment at Levels 1, 2 and 3 with clear progression routes through these levels and beyond into employment and higher education.

44. Diplomas will require innovative forms of assessment to reflect the blend of practical and theoretical learning within them while ensuring the rigorous standards expected of a national qualification. Assessment will be a combination of locally determined and standardised external assessment that will provide formative and summative data to inform the progress of individuals and the performance of educational institutions. Diplomas will be graded on the same scale as GCSE and A levels.

The extended project

45. QCA has developed, in consultation with HE institutions and employers, the criteria for models for an extended project which will both form part of the new Diplomas and be a free-standing qualification which can be taken alongside A levels. The extended project will be in an area of the student's choice and will have the potential to stretch all young people and to test a wider range of higher levels skills, such as independent research, study and planning. It is currently being piloted and will be available nationally from 2008.

46. We do not want the extended project to add to the assessment burden on students and institutions and it is our expectation that students pursuing an A level programme of study will normally complete an extended project instead of a fourth or fifth AS level and not in addition. For Diplomas, the project will form an integral part of the programme of study. We will review assessment and grading of the extended project as part of the evaluation of the pilots to ensure that burdens are minimised and the assessment and grading arrangements are fair and robust.

Roles and Responsibilities

Qualifications and Curriculum Authority (QCA)

47. The QCA is a Non-Departmental Public Body (NDPB), with a Board appointed by and accountable to the Secretary of State for Education and Skills. The Secretary of State sends an annual remit letter specifying his priorities for the QCA and setting out its budget, and holds the QCA to account for delivery of these priorities and QCA's statutory functions. This accountability works differently for different areas of QCA's work: the Authority has statutory responsibility for regulation of qualifications, so it would be inappropriate for the Department to intervene in that; but in other areas, such as the 14–19 reform programme, the QCA is implementing Government policy, and the Department and QCA work closely together on delivery.

48. We believe that the accountability works well, striking a balance between respecting QCA's arm's length status and ensuring it makes its contribution to delivering the Government's education priorities and provides good value for the public funds it receives.

49. Separately from its formal accountability, QCA also has to maintain its credibility with the wider education sector, particularly the learners, employers and others who rely on the qualifications it regulates and the curriculum materials and tests it produces.

The respective roles of the QCA and Awarding Bodies

50. QCA is responsible for the standards of National Curriculum assessment and for the delivery of National Curriculum tests through its National Assessment Agency (NAA), which contracts with external providers for the various stages of the process, eg test development, external marking, data collection. The QCA regulates the tests to ensure that standards are maintained and that assessments are fair and effective.

51. Different arrangements apply in relation to external qualifications, such as GCSEs, A levels and the new Diplomas. In those cases, a range of awarding bodies develops the qualifications, organise assessments and make awards, regulated by the QCA.

52. This difference properly reflects the different nature and purpose of tests and qualifications: QCA has to secure provision of statutory tests in core subjects linked to the National Curriculum; but for qualifications, we need to have a range of different options available, offered by credible, independent awarding bodies with assessment expertise, to reflect the needs and aptitudes of learners of all ages, including those outside the state sector.

July 2007

Supplementary memorandum submitted by the Department for Children, Schools and Families (DCSF)

INTRODUCTION

1. In May 2007, the Department submitted a memorandum to the Education and Skills Committee's inquiry into testing and assessment. This supplementary memorandum provides further information relevant to that inquiry. The Government remains committed to the assessment system as described in the original memorandum.

NEW REGULATOR FOR QUALIFICATIONS AND TESTS

2. On 26 September 2007, the Secretary of State announced plans to establish a new, independent regulator of qualifications and tests. The regulator will be the guardian of standards across the assessment and qualifications system and will report to Parliament on the tests and qualifications system and the value for money it offers the taxpayer.

3. *Confidence in Standards: Regulating and developing qualifications and assessment* (Cm 7281) was published jointly by the DCSF and the Department for Innovation, Universities and Skills on 17 December 2007, and sets out the Government's proposals in greater detail. A consultation on these proposals is running until 10 March 2008.

4. The proposed reforms have two goals: firstly, to ensure that we have a regulatory system that continues to maintain the high standards of qualifications and assessment in England; and secondly, to ensure that learners, employers, higher education and the general public have full confidence in these standards and their consistency year on year. The reforms will achieve these goals in the following ways:

- The scope, powers and functions of the new regulator reflect the Government's principles of good regulation and build on the experience and expertise of the QCA, which has managed the qualifications and assessment system successfully for the past 10 years. A number of changes to the existing regulatory approach are proposed in Chapter 2 of *Confidence in Standards*. The proposed regulatory system will allow the regulator to maintain standards in the current system and to operate effectively in the changing landscape of the coming years;
- The creation of the new regulator will separate regulatory activity from the work of the QCA on development and delivery of public qualifications. This will resolve the perceived conflict of interest that exists with these functions sitting in the same organisation; and
- The new regulator will be independent of Ministers and will report to Parliament on its work. This will allow it to demonstrate that it is carrying out its regulatory activity independently of government, which will help to ensure public confidence in the standards that it is regulating. The Secretary of State wrote to the Chairman of the Select Committee on the day the consultation document was published, inviting the Committee to consider how it wished to monitor and review the work of the new regulator.

5. The Office of the Qualifications and Examinations Regulator will be a Non-Ministerial Government Department. It will have its own chief executive, chair and non-executive board. We will legislate to establish the new body at the earliest available opportunity following the outcome of the current consultation exercise. In the meantime, an interim regulator will be set up under existing statutory powers in advance of next summer's exams.

Single Level Tests

6. The Department's previous memorandum referred to the *Making Good Progress* (MGP) pilot, which is trialling new ways to measure, assess, report and stimulate progress in schools (paragraphs 37–40). It involves pupils in Key Stages 2 and 3 in over 450 schools across 10 Local Authorities. Since the submission of that memorandum the pilot has now begun. It started in September 2007 and will run to July 2009. Its five elements are:

- one-to-one tuition of up to 20 hours in English and/or maths for pupils behind national expectations who are still making slow progress;
- a focus on assessment for learning;
- the introduction of single-level tests which pupils can take "when ready";
- school progression targets (in 2008–09 this is to increase by at least 4% points the proportion of pupils making 2 levels of progress in the key stage); and
- an incentive payment for schools which make good progress with those children who entered the key stage behind expectations.

7. Within the MGP pilot, we are looking at how testing might more effectively support a personalised approach to learning and encourage every child to make good progress throughout their school careers. In December, 22,500 pupils took 43,000 single level tests in reading, writing and mathematics. An independent evaluation of the pilot being undertaken by PricewaterhouseCoopers (PWC).

8. Single level tests are shorter than the current end of Key Stage tests and each covers a single level of the National Curriculum, from Level 3 to Level 8. They are aimed at pupils aged between seven and 14. They are intended to confirm teachers' judgements and are designed to motivate pupils by focusing on the next step in their learning. Pupils will take a test which is pitched at the level at which they are judged by their teacher to be working, rather than a test which spans a range of levels. If they are unsuccessful, they will be able to take the test again.

9. During the pilot, which runs until summer 2009, single level tests will be available in December and June each year. Pupils will also take the current tests in English, mathematics and science at the end of Key Stages 2 and 3.

10. In the Children's Plan, which we published on 11 December, we signalled our intention to implement single level tests in reading, writing and mathematics on a national basis at the earliest opportunity, subject to positive evidence from the pilot and endorsement of this approach from the regulator. Those tests would replace the National Curriculum tests at the end of Key Stages 2 and 3. We will also explore new options for the assessment of science. In the meantime, the current National Curriculum tests for science will continue.

Assessment for Learning

11. The Children's Plan has also cited the aim to make the use of tracking and Assessment for Learning (AfL) tools and techniques truly universal across all schools, extending them beyond the core subjects of English and mathematics. The English and mathematics Assessing Pupil Progress (APP) materials, already developed by the QCA, should become universally used in schools, and we want to expand those tools into more subjects, starting with science. Schools will be expected to have a systematic approach to AfL and intervention as a key strategy in helping both pupil and teacher understand where they are in their learning and what their next steps should be. AfL, and especially the use of tracking tools such as APP, also ensures that teachers' assessment judgements against curricular targets are consistent, robust, and build on a solid evidence base. This will be particularly important if single level tests are introduced on a national basis, so that teachers are able accurately to enter children for the appropriate test when they are ready. £1.6 billion has been committed for personalised learning over the next CSR period (2008-11) to ensure that schools are able to do this.

GCSE Controlled Assessment/Coursework

12. As our previous memorandum explained (paragraph 34), changes are being made to GCSE to respond to concerns expressed about coursework—tightening and strengthening assessment arrangements—as well as to ensure that specifications develop to reflect the recent secondary curriculum review and the requirements of the Disability Discrimination Act. Following a series of consultations and reports, QCA recommended that controlled assessments should replace coursework in the following subjects: business studies, classical subjects, economics, English literature, geography, history, modern

foreign languages, religious studies and social sciences. In addition, it recommended that in art and design, design and technology, home economics, music and physical education, internal assessments should continue with stronger safeguards. It also recommended mathematics coursework be removed from September 2007: this has now come into effect.

13. Controlled assessments are assessments that take place under controlled conditions in schools and are either set or marked by an awarding body. This approach will address the need to balance concerns about potential cheating against the added burden and cost to schools of monitoring every piece of work a student does or increasing the number of exams.

14. In the summer of 2007, QCA consulted on new GCSE qualification and subject criteria, incorporating the above recommendations. They received clear support from most of the respondents. The criteria have now been finalised and the subsequently revised specifications will be available from September 2008 for first teaching in 2009. The exceptions are science, which will remain unchanged, and English, English Literature, information and communication technology (ICT) and mathematics, for which new specifications will be available in the autumn term of 2009 ahead of first teaching in 2010. The English, ICT and mathematics criteria are being revised on a slower timescale to other GCSEs to allow time to incorporate the functional skills which are currently being piloted.

New Diplomas

15. As explained in our previous memorandum (paragraph 43), to increase the options available to young people the Government is introducing a new range of qualifications for the 14-19 phase. Diplomas will provide a new way of assessing standards at Levels 1, 2 and 3 in 17 lines of learning. Consortia of schools, colleges and other providers will begin delivering the first five Diplomas from September 2008. A further five will be rolled out from September 2009, and four more in September 2010. By 2013, all students anywhere in the country will be able to choose one of the first 14 Diplomas.

16. In October 2007, the Secretary of State announced the introduction of three new Diplomas in Science, Humanities and Languages—it is expected that these will be available for first teaching in September 2011. Suitable Diploma Development Partnership structures will shortly be established to specify the most appropriate content and assessment arrangements for each of these new Diplomas.

17. The Government has committed that in 2013 it will review the evidence and experience following the introduction of all Diplomas to reach conclusions about how in practice the overall qualification offer meets the needs of young people in progressing to further study and employment. It will consider the future of A levels and GCSEs in the light of this evidence.

18. We are currently developing a strategy for all 14–19 qualifications, to underpin our 14–19 reforms, and will publish proposals for consultation shortly.

Assessment of Diplomas

19. Diplomas are applied qualifications and the approach to assessment needs to support learners to learn how to apply their learning. Assessment will therefore be a combination of internal controlled assessment and external assessment. In designing the Principal Learning qualifications for the Diploma, awarding bodies have chosen the assessment methods that best suit the knowledge, skills and understanding required for each unit. Units that involve internal controlled assessment will typically focus on practical learning and those with a more theoretical focus will be externally assessed. This approach is consistent with our approach to other national qualifications. The majority of GCSEs and A levels have a theoretical focus and therefore are predominantly externally assessed: however, as set out above, those GCSE subjects which have a greater focus on applied knowledge and skills will include an element of controlled internal assessment.

20. QCA has recently published guidance on controlled internal assessment which is available on their website. This sets out how internal assessment may be controlled in relation to how tasks are set, taken and marked and also how teachers are trained to assess.

21. To ensure that the assessment system and assessment practice for Diplomas in schools and colleges operates to a high standard, we have given the National Assessment Agency a remit to develop a national approach which will set standards for high quality controlled internal assessment processes and practice and will ensure the professional expertise of local assessors.

January 2008

Witnesses: **David Bell,** Permanent Secretary, **Sue Hackman,** Chief Adviser on School Standards, and **Jon Coles,** Director, 14–19 Reform, Department for Children, Schools and Families, gave evidence.

Q287 Chairman: Can I welcome Jon Coles, David Bell and Sue Hackman to our deliberations? I apologise for the Division that has delayed the beginning of this session. You seemed to be enjoying that first session, from which we got some good information and feedback. We normally allow you to say something—it is very nice to see you here, Permanent Secretary, we were delighted that you were able to join us— so do you and your team want to say something to get us started, or do you want to go straight into questions?

David Bell: If I could just make some brief introductory remarks, I did enjoy that last session, I hope that I am going to enjoy this session just as much.

Chairman: I would not guarantee that.

David Bell: This year is an interesting year because it marks the 20th anniversary of the passage of the Education Reform Act 1988, which gave us, among other things, national tests. The key purposes of testing and examinations more generally have stood over that period. We want them to provide objective, reliable information about every child and young person's progress. We want them to enable parents to make reliable and informative judgments about the quality of schools and colleges. We want to use them at the national level, both to assist and identify where to put our support, and also, we use them to identify the state of the system and how things are moving. As part of that, both with national tests and public examinations, we are very alive to the need to have in place robust processes and procedures to ensure standards over time. What it is important to stress, however, is the evolution of testing and assessment, and that is partly for the reasons that you have just heard—that demands, employers' expectations and society's expectations change. However, we also want to be thinking about better ways to ensure that testing and assessment enables children and young people to make appropriate progress. Many of the issues that I know are the subject of the Committee's inquiry are close to our hearts and are issues that we are looking at. For example, in areas like single-level testing or the assessment arrangements around Diplomas, we are trying to take account of change in expectations and demands. Throughout all this, rigour is essential in assessment and testing, and I assure you that both the Ministers and the officials of the Department are completely behind any changes that will improve the testing and assessment of children and young people in this country.

Q288 Chairman: Thank you for that. Shall we get started with the questions now? Jolly good. My question comes from the last remark by Professor Smith, from Exeter university. Why do you think we are not getting above this still-restricted number of people getting to A to C in GCSEs? He finished by saying that that there is one thing that restricts it; 90% of people who take A-levels get into university. Why is it that there is this restriction; why are we not getting more young people through? As you said, this is the 20th anniversary this year of the Education Reform Act, so why do we still seem to be performing lower than the demands of a modern economy would suggest we should be?

David Bell: It is important to stress that we have made considerable progress. If you take the percentage of youngsters who achieve five A* to C grades at GCSE, we know that that figure has improved over time with English and maths. The percentage of young children leaving primary school with the appropriate level of education has improved. What Professor Smith argued, and what we would argue, is that you have to keep the pressure on all the time to ensure that more and more young people achieve the required standard. There is not a single, simple answer. For example, trying to improve the teaching of reading in the earliest years of primary school is as important as ensuring that those youngsters who have not achieved the appropriate standard at age 11 are given support to get into secondary education, or ensuring that youngsters have the right kind of curriculum choice, whether that is Diplomas, GCSE, A-level or whatever. This is a story of progress. As you have heard from other witnesses in this inquiry, we have gone from below average to above average, but there is a long way to go. This country's economic success depends on more people becoming highly skilled and we know that other nations, both developed and developing, are increasing their rate of progress in educational attainment.

Q289 Chairman: Earlier I said that I had been at Southfields Community College. One of the complaints that I heard there was that it was still getting a substantial number of children coming through at 11 who were not able to read and write properly. We have had numeracy and literacy programmes for a considerable time now. Is it the quality of teaching? Is it the way we teach? What is holding things back?

David Bell: Again, I would make the point that there has been considerable progress over that period. However, there is no getting away from the fact that we must make even greater progress. There are some things that are crucial: the quality of the teaching that children get is important in the primary years as is the content of the curriculum and what they are actually taught. Again, we have made changes to the teaching of reading in recent years. In the new arrangements for Key Stage 3 at the beginning of secondary education, we have tried to make more time and space available for schools who are still picking up youngsters who have not made sufficient progress in primary education. That is all very important. When we get on discuss the single level tests—I know you will want to talk about them—one of the principles underpinning the pilot is the placing of greater emphasis on the progress that students are making and really ratcheting up the teacher's capacity to assess where students are, and to help them to make appropriate progress. The youngsters who are falling behind will be given additional support in English and maths. There are lots of ways in which we can continue to improve

performance, but we need highly skilled teachers, head teachers and school leaders who are really focused on raising standards.

Chairman: Let us get on with the rest of the questions.

Q290 Mr. Heppell: You started by speaking about the different things that tests and so on are used for. We are all aware that the national curriculum tests and the public examinations at 16 and beyond are used for different methods. Some people might say that if they are designed initially to put a greater emphasis on pupil performance but are then used for accountability reasons, they will be less accurate and less relevant. Do you recognise people's concerns about that?

David Bell: I do recognise that, but I have never found it a particularly persuasive argument. It seems to imply that you can only use tests or assessment for one single purpose. I do not accept that. I think that our tests give a good measure of attainment and the progress that children or young people have made to get to a particular point. It does not seem to be incompatible with that to then aggregate up the performance levels to give a picture of how well the school is doing. Parents can use that information, and it does not seem to be too difficult to say that, on the basis of those school-level results, we get a picture of what is happening across the country as a whole. While I hear the argument that is often put about multiple purposes of testing and assessment, I do not think that it is problematic to expect tests or assessments to do different things.

Q291 Mr. Heppell: Will you accept that the use of test results for that purpose means that there is a tendency for teachers in schools to concentrate on improving their statistical performance and working for the tests at the expense of children's education in general?

David Bell: I would obviously want to make the point that enabling youngsters to be well prepared for a test or a public examination is quite important. Actually some schools had not previously given sufficient attention to that. Before we had national curriculum tests, the first time that some youngsters took a structured test or anything approaching a structured test examination was at the age of 16 when they came to sit their first public examinations. It is important as part of preparation for a test that youngsters are given some experience of that. We also hear some people arguing that it skews the whole curriculum—that teachers become completely obsessed with the testing perhaps because they are concerned about the school's performance and the whole curriculum is changed. Our evidence, which I think is supported by inspection evidence, is that schools that are confident and know what they want children to learn can comfortably ensure that children are prepared for the test in a way that does not distort teaching and learning. I do not think that anyone would say that they want everything to stop just for the test. Again, examples from the best schools suggest that they can comfortably marry good progress made by children in tests with a rich and varied curriculum.

Q292 Mrs. Hodgson: I want to ask a couple of questions about teaching to the test. We have received significant evidence that schools and teachers are now so motivated to meet targets and do well in league tables that they resort to widespread teaching to the test. Consequently, there is a narrowing of the taught curriculum. What are your comments on that?

David Bell: Perhaps, Mr. Chairman, I could begin to bring some of my colleagues in?

Chairman: It would be a pleasure to hear from them—we know them well.

Sue Hackman: First of all, it is important that literacy and numeracy are robust because that is what the rest of the curriculum depends upon. There was never a great student in design and technology who did not have good maths as well. Someone cannot easily study history and geography unless they can read and write well. Literacy and numeracy are the cornerstones of the whole curriculum. Personally, I do not think that there is a huge problem with focusing on the core subjects. At the same time, we support a broad and balanced curriculum. That is part of what should be taught. We have researched how much time schools spend in preparing for tests. In the four years of Key Stage 2 it is 0.14% and during Key Stage 3 it is 0.2%.[1] That does not seem too much. In life, we need some experience of being challenged and stretched as well as of being supported and coached. That is part of the rich experience that we must provide. Having said that, we do give guidance to schools that they should not over-drill to the test. They should prepare for the test so children can show what they are capable of, but the biggest reason why children should not be trained just for the tests for long periods is that it does not work. What works is effective, consistent teaching and learning throughout the three or four years of the key stage. It is certainly not our policy to drive schools to spend too much time on teaching to the test.

Chairman: Leave David to one side; he will come in when we need him. I really want to get Jon and Sue to pursue that matter.

Jon Coles: Can I just add something on that question of teaching to the test, which is an important issue? Broadening that in relation to public exams, which is my side of this discussion, it is obviously important to get the curriculum right, and to have the right curriculum that prepares young people in the right way. A lot of what you were hearing from the previous witnesses was very much about how well the curriculum prepares young people for life.

[1] *Note from David Bell, Permanent Secretary:* In answer to question 292, Sue Hackman indicated that the Department holds research evidence on the amount of time schools spend preparing students for tests. I would like to clarify that the evidence we hold actually relates to the time schools spend administering tests. The time spent administering tests at Key Stage 2 is 0.14% and during Key Stage 3 is 0.2%. I recognise that this is a meaningful difference and apologise for the mistake.

Then, of course, you need to test the whole of that curriculum, and you need testing that is valid and reliable in relation to that curriculum, and that tests the skills that people were talking about, not just a very narrow set of skills and abilities to answer the test. That means that the test needs to be properly trialled and robust; it needs to have the right assessment instruments for the sort of learning. The testing that you do for mathematics is very different from the testing that you do for construction or engineering. Evidently, some things can be assessed through written external tests, and some things cannot be. You need the test not to be too predictable, so that it is very difficult to narrow down on the questions that you need to prepare people for. At that point, good quality test preparation and preparation for exams is making sure that young people can display what they know and can do in the assessment that they are about to face. I, for one, am glad that my teachers went to the trouble of preparing me to do the tests and exams that I had to do, because having taught me the curriculum properly, it was important that I was able to demonstrate my understanding of it in the exam. I think that that is what we would want for all our children—that they should be taught the curriculum properly and then taught to display what they know.

Q293 Mrs. Hodgson: What we have been trying to get to the bottom of in the evidence that we have taken so far is whether teaching to the test is a good or a bad thing. When we had the exam boards here, we asked them the same question, and the feeling from them was that it was more of a good thing. However, we have all had a detailed e-mail from Warwick Mansell,[2] who is an author and reporter for *The Times Educational Supplement*.
Chairman: Warwick Mansell has written a very good new book on testing and assessment. We are getting a percentage.
Jon Coles: Like your previous witnesses, I have not read the book either.

Q294 Mrs. Hodgson: Warwick Mansell has obviously done a lot of research and interviewed a lot of people about the issue of teaching to the test. You mention literacy and numeracy, and that has such an impact across the whole curriculum. However, repetition is very important, because that is how you learn a subject. Biology and chemistry, however, are quite different and there is one example here of a sixth-former who says, "Most of my chemistry class excelled at chemistry exams, but knew very little about chemistry as a subject. The same was true in biology." My only worry is that children are being taught to pass the exams, which is good, because they need to be able to pass the exam, but are they developing independent thinking so that when they get to university, they will know how to study a subject fully on their own and to develop their own strand of thinking, rather than studying the best way to pass the exam?

[2] Memorandum TA 48 (on Committee website)

Jon Coles: The point about A-level that your previous witnesses were making—that they worry that assessment has become too atomised and not sufficiently synoptic, because it does not test people's ability to make connections across the subject—is something that we have heard a lot. That is why we are making some significant changes to A-level, in particular, reducing the number of assessment units from six to four, which goes straight to the point about how much testing is going on. It also makes sure that people have to make connections across a bigger range of the subject, and changes to the forms of question in A-level will make sure that people are asked questions that have more variety and that require more extended writing, more analysis, and independent thought and study. A number of things have strengthened A-level and made it an even stronger qualification since the introduction of Curriculum 2000. The A-levels that I suspect that most of us in this room did were possible to pass by being taught seven essays, and revising five of them, and expecting them to come up. That was a pretty good strategy for passing A-levels in the past. You cannot do that now because A-levels require you to know and understand the whole of the syllabus. They test the whole of the specification at every sitting. The issue is whether A-levels have gone too far in testing all of the knowledge in small chunks. Do they do enough to test an understanding of the connections across the subject and an ability to analyse in depth? That is the purpose of the reforms that are now in train. The A-levels that will be taught from this September will be assessed in a new way with four units rather than the six that currently exist. There will be a bigger variety of question stems, more analytical questions and more extended writing. That is an important set of reforms. Set alongside the new extended project, that means that the sixth-form experience that people will get in the future will demand much more independent thinking and analysis of the subject. Nobody would condone for a second the idea that students should be taught to answer exam questions and nothing else. Just as Sue said that that is a bad strategy at Key Stage 2, I want to emphasise that it is a bad strategy for teaching A-level because there is already a significant amount of synoptic assessment. It is not possible to get the best grades at A-level without demonstrating an understanding of the wider subject. I do not want you to take quite at face value the quotation that you read to me. We do not see the scenario across the country that people can pass exams, but have no understanding of the subject. I do not think that that is the reality in our schools and colleges.
Chairman: We will come back to that point.

Q295 Annette Brooke: I have two quick questions. I want to return to how much time is spent on teaching to the test and the percentages that were given about that. If we were to do a survey of primary school teachers and we asked how they approach teaching year 7 compared with how they approach teaching

year 6, do you not think that our survey would show that they feel constrained in their overall teaching of year 6?

Sue Hackman: In year 6, teachers prepare for the tests. That is the reality of life and it is appropriate that they prepare pupils for those tests. I dare say that there would be a difference. We might also reflect that teachers in year 7 should have the same sense of urgency in teaching their pupils to get them ready for the rest of that key stage. There is a difference in teaching every year group because every year has a slightly different purpose; there might be a test, pupils might be starting out in a key stage or they might be consolidating what they have already done.

Q296 Annette Brooke: But is this not like one big hiccup, which interrupts the flow of general education? When a test is coming up, it is a critical time and the whole school will be judged on it.

Sue Hackman: Hopefully, with the single level tests, we are moving towards children being tested throughout their careers as they become ready for it, so that we can see how they develop as they go along. That would spread some of the pressure through the years.

Annette Brooke: I will leave my colleagues to ask about that.

Chairman: David Bell wants to come in on that point.

David Bell: I just want to make the point that a number of the folk in this room spend quite a lot of time visiting schools. As Sue mentioned, you do hear people saying that the pressures on youngsters get greater at year 6. People often tell you that they are teaching to the tests, that all of the imagination is gone and that there is no room for anything else. However, on talking to them further and on talking to the children, you will hear about the huge range of activities that are going on. That somehow gives the lie to the argument that the curriculum has become completely narrowed as a result of testing. On this issue it is quite hard to get to what people actually do, as opposed to what they think. It is very unusual to go to schools where everything has been turned over for a large amount of time to focus just on the tests. I can speak from very considerable experience, having visited hundreds of schools.

Q297 Annette Brooke: I guess that it is a matter of proportion. Running through our evidence, we have heard about the wide degree of variation in the giving of grades in these assessments; 30% of the grades awarded being wrong is the figure that was mentioned. In most walks of life, when there is a potential degree of error, or perhaps a particular variation in the whole year cohort, there is a little note under the table explaining that. Do you think that we should have some footnotes when you publish your league tables?

Jon Coles: On the public examination system, I do not know where this figure of 30% comes from. I have seen it appear in your transcripts and I have not been able to track down its source.

Chairman: One of our witnesses gave evidence on that.

Jon Coles: They said that that was the case, but I do not know where they got their information from.

Chairman: Okay.

Jon Coles: Just in relation to our public exam system, I simply do not accept that there is anything approaching that degree of error in the grading of qualifications, such as GCSEs and A-levels. The OECD has examined the matter at some length and has concluded that we have the most carefully and appropriately regulated exam system in the world. You did not ask the chief executives of the award body whether they accepted that figure.

Q298 Annette Brooke: I did ask the Qualifications and Curriculum Authority what it was doing to investigate the matter. I was not very satisfied that it was checking out the figure. I think that it is important to check it out. Perhaps you could ask them to do it, and then you can put your hands up and say that there is not a 30% error rate.

Jon Coles: I can say that without asking QCA to do any further work, because QCA regulates the exam system robustly. I can say to you without a shadow of a doubt—I am absolutely convinced—that there is nothing like a 30% error rate in GCSEs and A-levels. If there is some evidence that that is the case, I would really welcome knowing what it is.

Chairman: Jon, let us agree that together we will get to the bottom of this.

Jon Coles: That would be excellent.

Chairman: We all want to pursue it. It was a figure that was given to the Committee and we will pursue it.

Annette Brooke: That is fine.

Chairman: I would like to move on to the broader question of standards.

Q299 Stephen Williams: Perhaps I could direct these questions at Sue Hackman as this is her area of responsibility. In his introductory remarks, the Permanent Secretary said that results have gone up. He was talking about those getting five GCSEs at A to C grades, including maths and English, and those getting A-levels. Is that the same as standards going up?

Sue Hackman: Statisticians would draw a distinction between the two. There are mechanisms to ensure that tests each year are anchored by means of pre-testing against the performance of the year before. There are also efforts to anchor standards at the level setting point, against the level that was achieved in the previous year. Therefore, there are safeguards in the system for anchoring the tests each year.

Q300 Stephen Williams: You think that standards have been held, but not necessarily improved?

Sue Hackman: I definitely think that standards have improved. We can corroborate that by the patterns that exist in other surveys of pupil abilities, and in the PIRLS and PISA tests, which also give us another take on how standards are developing as do Ofsted reports. We are pretty certain that standards

are rising. However, having said that, no one will say that every test is perfect. There must be year on year fluctuations. It is the task of the QCA to ensure that they are regulated and watched.

Q301 Stephen Williams: If Sue Hackman thinks that standards are improving, how does that stack up with the evidence from Professor Smith in the earlier session? He said that universities increasingly have to lay on remedial classes, particularly in maths, or for anyone who wishes to do a physics degree, because the standard of entrance—even if a pupil has three straight As—is not sufficient to be a first year undergraduate. Therefore, how could standards have gone up?
Sue Hackman: National tests test every single child in the cohort. University intakes vary from year to year, so it did occur to me that university fluctuations might be to do with which pupils are choosing to come to your university over time. Having said that, I agree that when pupils are studying a science degree at university, they need a high level of mathematics, and we must strive to produce pupils who can do that.

Q302 Stephen Williams: But if you look at the number of people applying to do a subject such as physics, it has gone down considerably over the last 20 years. So it is safe to assume that the people who are applying to do physics now are the hard core who want to do that subject and are committed to it. Yet university departments say that the standard coming in is not what it was—not only at Exeter, I have also heard it from many other admissions tutors around the country.
Jon Coles: There are some important considerations relating to how many people are doing a STEM subject, which is broadly science, technology, engineering, or maths, and there is a need to increase that number. You will know from previous inquiries, and from documents that the Department has submitted to the Committee, some of the background to what we are doing about the matter. One thing that has been happening over the last 20 years at A-level and post-16 years old, is a significant widening of the choices available to young people and a significant increase in the range of options that they can take. A feature of that has been a decline in the number of people doing some of the traditional subjects, particularly science subjects. That is unquestionably a cause for concern. In recent years we have seen some reversal of that trend in relation to some key science subjects and to maths, which was affected after curriculum 2000. A range of issues needs to be addressed. I do not think that any of what I have said suggests that standards themselves have fallen, but rather that there is a wider range of young people doing a wider range of subjects. Therefore, people who might once have chosen to do science subjects are choosing other subjects. That is a feature of our system, and you have to judge at every stage to what extent to allow people free choice of subjects or to attempt to constrain their choice.

Q303 Chairman: If you took students from the independent sector and grammar schools out of the number of young people qualifying to come through in STEM subjects, you would be very worried indeed. So something is going on in the state sector that you surely need to take seriously.
Jon Coles: That is absolutely right. I do not have the figures in front of me for numbers of school leavers doing STEM subjects by different types of school, although the Department certainly has them. That concern is an important policy direction. As you know, a huge amount of work is going on to raise the numbers of students doing the sciences, particularly the physical sciences, maths and other technology subjects. That is important work. The good news is that, in the last couple of years, some reversal in that decline has begun and numbers are beginning to come up again. Numbers in maths have now recovered to above their pre-curriculum 2000 levels, which is serious progress. There is more to do, and of course the new science Diploma will be part of encouraging more and a broader group of young people into science.

Q304 Stephen Williams: I think that we are drifting on to the choices people make at 16, rather than the standards once people have exercised those choices, which is what we are asking about. Looking at A-levels in the round and going back to some of the comments made in the earlier session by Professor Smith again, I wrote down what he said: increasingly students mark hunt; they do not have independent thinking and are afraid to give critical answers; and they play safe. Do you accept that it is a fault with the existing modular A-level system that students are simply trying to leap each hurdle in order to get the grade at the end, rather than having an in-depth understanding of the subject?
Jon Coles: As I was trying to say in an earlier answer, the A-level system tests people's knowledge of the whole syllabus better than before, and ensures that more young people have to know more from all of it. I think that we agree there is a need to have more questions that test the abilities to think independently and critically, to analyse in depth, to give extended answers and so on. That is the purpose of the reforms in train at the moment, which will reduce the number of A-level modules from six to four and increase the amount of synoptic assessment. That is absolutely a driver of policy to get more of that into A-level.

Q305 Stephen Williams: I have one and a half questions. This is the last time that I will ever ask a question in this Committee.
Chairman: On this day.
Stephen Williams: First of all, it has been put to us that as a result of this teaching to the test, teachers have become more skilled at getting children or young people through exams, but the downside is that the students themselves become deskilled in the area I have just asked you about. Given that universities have these worries and employers have these worries, as we heard from the CBI, and given that universities are increasingly introducing or

reintroducing their own supplementary exams and tests, are A-levels fit for the purpose for which they are perceived to be designed?

Jon Coles: Yes, I think they are. The changes we are making now will make them more fit for that purpose. The tougher questions that follow from the changes that I have been talking about and the introduction of an A* will address the concerns of those universities which struggle to discriminate between candidates. Those changes, together with the extended project, will mean that more and more young people are required to learn independently, to study in depth and to pursue their own thinking and ideas. Those are crucial things. All of those things are being designed into Diplomas from the outset. So the size of Diploma assessment unit will be the size of the new, bigger A-level assessment units rather than the existing smaller ones. They will be designed to encourage critical thinking and reflection. That broad set of personal learning and thinking skills—the things about self-management you were pursuing in the last session—the ability to work in teams, to learn in depth and to research critically are all built into the Diplomas and into the new extended project. So all young people pursuing that route will be required to pursue in depth and under their own steam an area of learning that is particular to them. It will force them to become independent learners in order to be successful. That is a set of changes that we have under way which are very important.

Q306 Chairman: The awarding bodies said right at the end of our session last week that they thought an A* would certainly favour the independent sector in getting into the research-rich universities. Does it not concern you that they said that?

Jon Coles: We have analysed that. It was a concern that we all had. It was something that was important to look at. At the moment the analysis I have suggests that just over 9,000 of the 26,000 of the candidates for A-level—some may be over 18—who get three As come from the independent sector. That is just over a third. Our analysis based on last summer's exam results of what proportion would get three A*s, suggests that about 1,150 of just over 3,050 candidates who would have got 3 A*s had that grade been introduced last year would be from the independent sector. Again, that is just over a third. It is a slightly greater proportion from the independent sector, but it is not dramatic.

Q307 Chairman: But that is not the point, is it? The point is that what we know in this Committee, and you must know it well, is that the longer you extend and the higher the hurdles you put up in that process of coming through education to higher education, the more kids from less privileged backgrounds drop by the wayside or lower their sights. That is why we are so obsessed as a Committee in all its incarnations that, for example, Oxford and Cambridge still get away with having a different application system. As you said, David Bell, we all go to schools and we know that as soon as the kids from the poorer backgrounds find out that there is this posh route to Oxford and Cambridge they say, "I'm not going to be different from my friends and I'm not going to do that." So you lose them there. Then you are going to give them A*s, and that will put them off even further. That is the problem, is it not? It is more psychological than the careful analysis of figures you have just given us.

Jon Coles: I think that the point the awarding bodies were making in your discussion—it is absolutely right to raise it as an issue—is whether you would have a disproportionately greater proportion of people from the independent sector getting three A*s. This analysis suggests there is a slight increase in that proportion, but it is very slight.

Q308 Chairman: Why not just publish the results, give them the figures and let them judge? What is wrong with the figures? Why do you have to have an A*? Give them the figures. They can judge who has got what in their A-levels at present.

Jon Coles: Well—

Chairman: David, you are looking very quizzical.

David Bell: I will come in in a minute.

Jon Coles: At the moment, we certainly allow universities to have information about performance in AS modules and to have the marks in individuals AS modules if they want them—that is available to them.

Q309 Chairman: In AS?

Jon Coles: Of course, it is only AS marks that exist at the moment of making offers, which is the crucial moment for many of the selecting courses in HE. So, they do have that information available to them. The issue, of course, is that the rules around AS and cashing in mean that not every candidate has their AS grades at that point. We make available all the information that exists in the system now, but the point you are making is that some universities would like more information and a greater ability to make choices. This is an important issue. The crucial point about the A* is that it is designed with respect to the more stretching assessments that we are putting in place; it is designed to make sure that the things that we are trying to do to stretch students and the broader range of skills that you are rightly focusing on are taught, rewarded and recognised in the assessments. It is no good putting in those more stretching questions, which require people to show that they can analyse in depth, if, having demonstrated that they can do so, they do not get the reward for that. That is an important part of why having an A* makes a difference educationally, as well as in terms of selection.

Q310 Chairman: We do fear that there will be an A** and an A*** on the way shortly.

David Bell: Very briefly, I just wanted to defend the honour of many universities, which actually go out of their way to put into place programmes—some are funded nationally and some are introduced on the instigation of universities themselves—to open up access as best they can. But, of course, the universities themselves rely on the supply of students from schools and colleges, which is why a lot of our

attention has equally got to focus on encouraging schools to ensure that their youngsters get the right opportunities. The assessment is an important issue, as Jon has said, but a lot is being done across the system to ensure that students are given the best possible opportunities to go on to higher education.
Chairman: David, we take that point, but you know how this Committee has felt historically about the dual application process. We move on to single level tests.

Q311 Lynda Waltho: I think that we are expecting the results of the first round of the pilot single level tests on 18 January. I have a quote. Apparently, a DCSF spokeswoman said: "there are some differences between subjects, levels and performance in different key stages that we need to understand better before we are confident about releasing results". Are you any closer to understanding the results and publishing them?
David Bell: I am pleased to confirm that that quote was correct, and we said in a letter to the schools that participated that there had been some unexpected patterns in the results. The first thing to say is that we would not and should not be surprised that, when you pilot a new form of testing, you might need to see what actually happened. We are doing some further work, and we have asked the National Assessment Agency to do some further work. We are not ready yet to come back with the results of that analysis to say what has happened. To reinforce the point, however, we will make these findings public in due course. Obviously, we want to enable those schools and students who took part in the pilots to receive their results in due course. I cannot say any more than that at the moment. It would be unfortunate, to say the least, if I misled the Committee by starting to speculate on the results of that kind of analysis.

Q312 Lynda Waltho: It is about the time frame, really. The Government have actually stated that the single level test will be rolled-out depending on positive evidence from the pilot, but what are we talking about when we say, "positive evidence"?
David Bell: We said two things: first, as you suggested, there was positive evidence from the pilot, to which I will return in a moment; secondly, there is an endorsement of the approach from the regulator. It is a very important second point that we would need to ensure that the tests passed the standard. In terms of positive evidence, there are a number of things that you want them to do. You ask whether they are robust, reliable and valid, what is their impact on students' and teachers' behaviour, and so on. We have to consider a range of things in this round of testing. We are very fortunate in having 40,000 students taking part at this stage, so there is a good base. We also said that we would expect at least four rounds in the pilot phase, so that we learn as we go—I did a bit of homework on this before I came in—and we should not be surprised that it takes time to do that, because if you look back to the introduction of the national curriculum tests, it took quite a bit of time to get them right. We are quite careful in saying that we are doing a national pilot; we will examine it carefully; it will be independently reviewed; and we will need the advice of the regulator before we go forward nationally. However, as Sue said earlier, we think that the basic principle is sound in offering youngsters a test when they are ready to take it, to build it much more naturally into the flow of teaching and learning. You have to go from what seems the right principle to something that works in practice, hence the need for the national test in the pilot.

Q313 Lynda Waltho: As the head of the whole thing, do you have a time frame that you think would be a good idea to aim at—a time when you would want it to be finished?
David Bell: We have the four rounds of the pilots, and perhaps Sue will give more detail on the timetable. We are cautious about saying that it will be done by this summer and that we will try to roll it out and so on, because we want to be very clear that we have the evidence to make a decision should Ministers decide to go in this direction.

Q314 Chairman: Should Sue's head be on the block for this? Sue, is it your section's responsibility that this project is not making sufficient progress?
Sue Hackman: Team work is my responsibility.
David Bell: I must come in and defend a member of staff. In the end, the Permanent Secretary has to take responsibility in the civil service.
Sue Hackman: We have given ourselves two years. There are two test windows a year for test runs. At the end of that time, we would know if this thing is workable, but then there is a run-in time to prepare future tests, so it would not automatically happen that, at the end of two years, you would go straight into a single testing system. We would commit to saying that, after two years, we would take a decision on whether to carry on or not.

Q315 Lynda Waltho: You were quite definite about it being possible by the 18th—
David Bell: Just to be clear, we had given a date for the release of the results to individual schools and pupils on the basis of the December pilot. That is the work we need to do, based on the first pattern of results. As Sue said, in terms of making a decision about the future of the whole programme, we are working to that longer time scale.

Q316 Ms Butler: So is it intended that, at the end of two years, the single level test would replace all the other Key Stage tests nationally?
Sue Hackman: Yes. The pupils in the pilot are doing both; they are testing the new tests and sitting the Key Stage tests. If it were successful, the aim would be that it would replace the current testing regime and there would be single level tests for pupils as they move through the system.

Q317 Ms Butler: What safeguards are in place—I ask this in the context of Bernard Coard's 1970 report on the education of black boys—to ensure that black boys will not be held back by teachers' perception that they are not ready to take the test?
Sue Hackman: Underpinning the single level test is a big project to secure assessment for learning in schools based on a piece of work that QCA is doing for us called APP—Assessing Pupil Progress. That is a ladder of progress in reading, writing and mathematics, and in due course we think it will apply to other curriculum subjects. Those criteria will help teachers to arrive at accurate rather than impressionistic judgments, and they are the same criteria that underpin the single level tests, so there will be a close tie between teacher assessments. Single level tests are attractive and attract a lot of attention, but the progression pilot is a project that starts with classroom assessment with strong strategies for knowing exactly what children can do and what they need to do next. It also helps with periodic judgments based on the criteria that I have just described. When the teacher and the pupil feel that the pupil is ready for the next level, they would be entered for the next level, a little like they do in music tests. The test will confirm the teacher assessment. With this model, teacher assessment will have more credibility and more importance than currently. The answer to your question is that assessment for learning is really important, and the test and the teacher assessment go hand in hand on the ground as part of the pilot.
Chairman: A lot of people are worried about that comparison with music tests. The Secretary of State uses it often—sorry Dawn.

Q318 Ms Butler: May I drill down on that a little? Are you saying that you are introducing further smaller tests to assess whether pupils are ready for the single level test?
Sue Hackman: No. I am describing the single level test. There is just one set of tests.

Q319 Ms Butler: So what is the APP?
Sue Hackman: The APP is the material to help teachers to arrive at accurate judgments in their everyday assessments. It is classroom assessment—ongoing assessment.

Q320 Ms Butler: Sorry, I am not clear exactly how it will monitor the pupil, as opposed to the teacher's perception of how the pupil is doing.
Sue Hackman: When a pupil arrives at a level—let us say Level 4, Level 5 or Level 6—it means something: the child has certain competencies. For example, at Level 4 English, children do not just read aloud and literally; they can read between the lines. That is how you know that they are at Level 4. That is the marker. At Level 5, they can use standard English and write in paragraphs. What is in the APP that measures progress? It spells out those markers and competencies showing that single children are at that level. When the teacher is sure and has assessed in the classroom that a child can do those things—for example, can use standard English and paragraphs—and knows and has seen on several occasions that the child can do that, they say that the child really is at Level 5 in their everyday work, so they can be entered for their single level test to get external confirmation that that is the level that they are at. It is a formal confirmation of the teacher's classroom assessment.
David Bell: May I come in on that? Miss Butler is also concerned about what you might call depressed expectation and some youngsters never being considered ready for the test. It is important, alongside all that Sue has described, that teachers and school leaders, such as head teachers, ensure that all youngsters, irrespective of their background, are suitably judged ready at the right time. We want a system with the potential to enable youngsters who are not doing well to make better progress and, as part of the pilot, we are also encouraging schools to look at two levels of progress that students can make, but we do not want all those good intentions to be undermined by some pupils not being considered ready when they actually are. That is an issue, first and foremost, of classroom practice—of teachers being really skilled at judging when students are ready to move on—and, secondly, and perhaps equally important, of school leaders asking the question, "Are we sure that every youngster, irrespective of background, has been properly assessed by the teacher and taken forward?"
Sue Hackman: May I add that one of the purposes of introducing the single level test is to introduce motivation into the system so that the child has some short-range goals to be going for during those long key stages that last four years? We think that it will add some interest and motivation for pupils who are facing those tests. For the most able, the test will give stretch—if they do well in their end-of-Key-Stage test,[3] it will give them additional challenges to move on to—and it will allow the least able, or those who perhaps move more slowly, to move at a pace that is most suitable for them while giving them something to go for. When a less able child enters the tests, they will enter at a level that is suitable for them, and at which the teacher is confident that they will achieve. The tests will build their confidence and enthusiasm for learning. It was partly those children who we had in mind when we introduced the tests.
Chairman: We have one last section and only nine minutes to get through it. Fiona, would you lead us through?

Q321 Fiona Mactaggart: We have talked quite a bit about the reliability of tests—the 30% figure—but I am also really interested in the validity of tests. It seems that in the hunt for more reliable testing—a test that will produce the same result every time—we are making tests that cut down the curriculum. In a way, that was the point that I was getting at earlier with the universities.I feel a bit like Mr. Gradgrind, who asked for the definition of a horse. Sissy Jupe knew everything there was to know about a horse but did not know that it was a gramnivorous quadruped, as I recall. I worry that we are

[3] *Note by witness:* This should read 'single-level test' instead of 'end-of-Key-Stage-test' .

overstressing reliability at the expense of validity in assessing young people's learning. Is there any truth in that worry, and, if so, what are you doing to try to overcome it?

Jon Coles: Should I respond in relation to the public exam system? I think that that comes to the point about the A-level specification, for example, and the extent to which being very tight means that the assessment objectives are clear and can be assessed precisely and reliably. That is one of the reasons why we have a reliable system of public examination. That then goes to the question that universities and employers were raising in your last session and, in all of that, the desire to ensure that there is the full range of knowledge, that the tests are reliable and that they can be confirmed—whether we have managed to capture sufficiently all the broader sets of skills. They range from self-management through to independent learning, the ability to work in teams and all the rest. The set of reforms that we have in place now, which range from the new Diplomas, which are designed with some really different approaches to assessment in places—

Q322 Fiona Mactaggart: OCR suggested in its evidence to us that seeking parity with GCSE has actually limited the range of assessment and made it too much like previous examinations.

Jon Coles: I do not think that that will be the case. I believe that in their oral evidence to you, several of the awarding bodies said that they hoped that the Diplomas would give them the ability to introduce the broader range of assessment methods that would better test the broad range of skills that all the universities and employers who appeared before you earlier said were so crucial to them, and that is indeed what the Diplomas are designed to do.

The extended project is another example of the system changing in quite a significant way to ensure that some of those higher-order intellectual, personal, practical and thinking skills can be developed and tested in ways that enable young people to pursue their own learning and to learn and research independently. So there is a set of changes in train, and it includes the introduction of the new forms of controlled assessment at GCSE, which are designed to ensure that we have a system that is at once valid and reliable in testing the skills and knowledge that the syllabus is designed to develop. I think that you are right to say that that is something that we have to keep working at over time, and at any one point in time there is a risk that one is stressed more than the other, but I think that the set of reforms that we are implementing at the moment really focuses on getting the set of things that employers and universities want, while ensuring that we have a very reliable system underpinning that.

Q323 Fiona Mactaggart: Would you expect a new regulator to publish data on the reliability of particular examinations and qualifications?

Jon Coles: I am not sure what data they would be.

Q324 Fiona Mactaggart: Everybody has an assessment of the reliability of a particular testing and examination system. As I understand it, the Key Stage 2 test was assessed as having a variation factor of some 30%, and I suspect that that was where the 30% figure originated. Do you not think that these figures should be public?

David Bell: The potential of an independent regulator enables it not just to work behind the scenes as it were, but also to make a report. In fact, there will be an annual report, and it will be for the independent regulator to determine what kind of evidence to put into the public domain. I would have thought that if one of the underpinning principles of having a new regulator was to build on the good work of the QCA and, crucially, to help to reinforce public confidence in tests and examinations, it will want to put their work out. In fact, we have also said that we expect it to have quite a strong research function. We expect it to be looking at international data, data from across the four nations in the UK and so on. So, I think that that would be for it to decide, but I would expect it to be putting a lot of their findings, including their technical findings, into the public domain.

Sue Hackman: We could design tests that would deliver fantastic rates of reliability, but I do not know if they are the kind of tests that we would want because they would be made up of those very small, reliable, atomised kinds of questions that do not deliver a sense of what the child can do and their ability in, as I think the vice-chancellor from Coventry described it, synoptic or conceptual understanding. With regard to national curriculum tests, we do our best with QCA to ensure that a range of skills are tested. There will be some very tight, specific questions and, at the other end of the paper, there will be some wider questions. I think that with any testing system, there is a compromise between having atomistic and reliable questions, and having wide questions that allow pupils with flair and ability to show what they can do more widely.

Q325 Annette Brooke: I wonder whether you could tell us, now that more and more work is being done on the Diplomas, what you believe will be the long-term future of GCSEs and A-levels. After all, you have put in the children's plan that as long as you have got positive results back on your testing of the single-level tests, that is the way forward. What are your confident predictions, give or take the feedback and piloting of the new Diplomas? What is the long-term future for GCSEs and A-levels?

David Bell: The Secretary of State made it clear in the autumn that he was not going to carry forward a review of A-levels in 2008, but was going to wait until 2013. The answer is, let us wait and see. By 2013, not only will we have seen the effect of the reforms to A-levels that Jon has described and other changes to GCSEs, but we will have all the Diploma lines up and operational. Do not forget that as we expand the apprenticeships programme, we will see more young people, we hope, following that

particular route. We are building a system that we hope will be increasingly good for each of those different qualifications, but that will also provide a wide variety of choice for young people and their families and meet a wide variety of needs. I cannot sit and speculate about what will happen.
The most important thing, surely, is to have a system of qualifications, tests and examinations that meets the individual needs of every young person but that, at the same time, continues to build the economic and social strength of our nation.

Q326 Annette Brooke: But you would keep the two routes that Tomlinson thought it would be a good thing to get rid of?
Jon Coles: Probably the best thing we could do is quote directly from the Secretary of State's statement on the day of the launch of the three new Diplomas and the formal launch of the first five. He said, "If Diplomas are successfully introduced and are delivering what employers and universities value, they could become the qualification of choice for young people. But, because GCSEs and A levels are long-established as well valued qualifications, that should not be decided by any pre-emptive Government decision, but by the needs and demands of young people and parents, schools and colleges, employers and Universities." I am not sure whether we could add to that. We think that they could become the qualification of choice, but that will ultimately be decided by people's choices and the qualifications that they value in future.

Annette Brooke: I should probably leave it there, although I am tempted to say a bit more.

Q327 Chairman: I think that we should leave it there. I have one last thing to say to David Bell. There is still a view among teachers and foreign commentators that we still rely on testing and assessment, that the pendulum must start swinging back at some stage to take the pressure off, and that we have gone as far as we can go on testing and assessment. Do you share that feeling?
David Bell: I think I would be much more concerned if we were sitting here saying, "No, we are prepared to defend everything and we are not prepared to consider any change. Everything must go on as it always has". I hope that you have heard today is that we are very open to the sorts of comments, questions, views and opinions that you have expressed. Much of what the vice-chancellors and the CBI have said in public has been said to us privately, and much of what we have done has been a response to that. I do not accept that we can ever have a system without good and robust national testing and public examinations, the results of which are made available to the public. At the same time, we must meet changing demands, as one of your earlier witnesses described, to ensure that we have the best system.
Chairman: Thank you. This has been a good session—it could have gone on, but we are late already. I am only slightly disappointed that I did not get a Lancastrian head on the block. Apart from that, I thank you very much.

Monday 18 February 2008

Members present:

Mr Barry Sheerman, in the Chair

Annette Brooke
Mr Douglas Carswell
Mr David Chaytor
Mr John Heppell
Fiona Mactaggart
Mr Andy Slaughter
Lynda Waltho

Letter to the Chairman from Jim Knight MP, Minister of State for Schools and Learners, Department for Children, Schools and Families

UPDATE ON THE MAKING GOOD PROGRESS PILOT

I would like to take this opportunity to update you on the *Making Good Progress* pilot which started in September last year.

The *Making Good Progress* pilot is testing new ways to measure, assess, report and stimulate progress in our schools. It involves pupils in Key Stage 2 and 3 and is running between September 2007 and July 2009 in 455 schools. We are providing more than £20m for the pilot in the academic year 2007–08.

The cornerstone of the *Making Good Progress* pilot is developing and improving teachers' assessment skills to focus on moving children on in their learning. Sharper use of assessment, better pupil tracking and better planning by schools to help each child to progress perfectly illustrates assessment for learning, which is central to our drive to raise standards for all children. We are investing a substantial amount in this, £150 million will be spent on improving assessment for learning practice in all schools between 2008–09 and 2010–11.

The main focus is on assessment for learning. Other elements of the pilot are:

- one-to-one tuition of up to 20 hours in English and/or maths to pupils behind national expectations who are still making slow progress;
- the introduction of single-level tests which pupils can take "when ready";
- school progression targets; and
- a payment for schools which make outstanding progress with those children who entered the key stage behind expectations.

This combination of measures is intended to ensure that all pupils, from the very brightest, to those who enter school below the expected standard for their age, are making the progress they should. It should encourage regular and better use of assessment to identify what each child needs to move on. Schools can call on extra support for those who need it, through one-to-one tuition, and there are incentives to schools to keep every child progressing to the next level throughout the key stage.

The pilot provides an important opportunity to make sure that all children, no matter what their starting point or background, progress as they should in our schools. Our strategy is to continue to raise standards for all while ensuring those who have the furthest to go, or those who are struggling, get the support they need, when they need it. The focus on progression will help all pupils, regardless of background or circumstance, to achieve their potential. It is crucial that we ensure more pupils make the expected progress through the key stages at school, especially pupils from disadvantaged backgrounds who have not kept pace with their peers.

SINGLE LEVEL TESTS

Single level tests are only one part of a coherent package of progression activity, though much of the attention on the pilot has focused on the new tests. As you know we committed in the Children's Plan to roll out single level tests nationally, subject to positive evidence from the pilot and to endorsement of this approach from the regulator.

The tests are part of a personalised approach to learning to ensure every child progresses at their own rate, rather than having a snapshot of attainment at the end of a key stage. They are used to confirm teachers' assessment, when they judge a child is working at the next level. We've just finished the first round of tests and, as with any new assessment, the QCA is reviewing the outcomes to refine the model.

Local Authorities and schools are positive about the principle of Single Level Tests. But the pilot is crucial to enable us to get the model and delivery of it right—this is only the first of four test windows in the pilot.

First Round of Tests

The first round of tests took place in December. 22,543 pupils took 42,833 new tests at Levels 3–6. There have been some unexpected patterns in the results which we needed to investigate. This is not unusual in piloting new tests, but it did mean a short delay in releasing their results to participating schools whilst the National Assessment Agency (NAA) reviewed the outcomes.

The results were made available to schools on 7 February, together with some explanation of the factors that may have caused unusual patterns of performance. Although we do not at this stage have a full explanation for this we are clearer about what the likely factors are. NAA are doing further work to feed lessons learnt into the next round of tests.

NAA's investigation found that a number of factors are likely to have combined to produce unusual outcomes, with variations in performance between KS2 and KS3 pupils taking the same test and at some of the levels tested:

- inappropriate entry, with a number of pupils entered who had a teacher assessment that was lower than would have been expected if the pupil was secure within the level being entered for;
- the new style of test with questions pitched at a single level, rather than a range of three levels, will have been unfamiliar both to pupils and to markers; It may be that pupils perceived the test as harder without the initial "warm up" questions to ease them into the tests, as they are used to in National Curriculum tests;
- pupils may have been less motivated—research suggests that pupil motivation for new tests taken in a pilot may be lower than for National Curriculum statutory tests, and that this factor may be more marked for pupils in Key Stage 3 than in Key Stage 2;
- a number of pupils did not get to the end of their papers, particularly on higher level papers; we are mindful that this is the first time pupils would have experienced national tests of this type and this could have had an impact on their performance; and
- markers are unaccustomed to marking scripts at a single level from pupils in two key stages.

Learning from the First Round of Tests

Of course, one of the main purposes of this first round of single level tests was to test the tests. It is not unusual for early trials of new tests to show unexpected results and inconsistencies.

It is important to recognise that NAA developed these tests on a much shorter timescale than is typical for test development, and that this did not allow for the usual pre-testing that would take place. Nor did they have the benefit of data from previous tests of this kind. The test development process for the June tests, whilst still on shorter timescales, will more closely match that for National Curriculum tests.

We are also working with participating schools and Local Authority pilot leaders to learn from their experience, and in particular to understand more fully issues related to teacher assessment judgements and test entry decisions.

As the two-year pilot develops schools, markers and all those involved in the administration of these tests will become more familiar with their use. There are four rounds of single level tests. Each of them will provide evidence on which to base an analysis of their effectiveness, and each may be used to improve and refine the test model. With the June tests we expect to establish better comparability with the end of Key Stage tests. At the same time, we are developing assessment for learning to support teachers in making sound judgements, and in entering pupils for the tests when they are truly ready. We intend to publish an interim report on the assessment model in the autumn, reflecting the experience of schools over the first year of the pilot, and after two rounds of tests.

February 2008

Witnesses: **Jim Knight MP,** Minister for Schools and Learners, and **Ralph Tabberer,** Director General, Schools Directorate, Department for Children, Schools and Families, gave evidence.

Q328 Chairman: Now that people have had time to settle down, I welcome the Minister for Schools and Learners, Jim Knight, and Ralph Tabberer to our proceedings. Our inquiry into testing and assessment is getting particularly interesting. We sometimes say to ourselves that we know that we are getting under the skin of an inquiry when we feel that we are more dangerous than we were when we started, because we have a little knowledge. We have had some very good evidence sessions, and we hope that this one will be of even more value than the others. Do either of you want to say anything before we get started?

Jim Knight: As is traditional, I will not make a statement, because I do not want to delay the Committee. On the letter that I sent to you today and circulated to other members of the Committee, as certain portions of the media have shown an interest in this subject, some clarification might be helpful so that you have more facts to get you beyond some of

the fiction that you may have read in the newspapers. The letter sets out the timetable for publishing an interim evaluation of the single level tests in the autumn. In general terms, we are very pleased with the progress of that particular pilot. Obviously, I will be delighted to answer your questions on that and anything else that you want to ask.

Chairman: Ralph?

Ralph Tabberer: I have no introduction.

Q329 Chairman: May I start us off by saying that this testing and assessment all seems to be a bit of a mess? We have taken evidence, which you must have read—your officials will certainly have summarised it for you. We have had so much evidence that shows that people are teaching to the tests and using the tests inappropriately, and for outcomes that were never intended. A lot of people have criticised the level of testing and assessment, and we are looking at whether it is fundamentally effective in improving the life chances of the children in our schools.

Jim Knight: As you would expect, I do not agree with you that it is a mess. Naturally, I have heard a lot of the evidence. I cannot be accountable for what your witnesses say, but I can offer you a bunch of other people who might say something different. In respect of teaching to the test, there is a yes and a no answer. In general terms, we are pretty clear about our priorities in testing. We want people to focus on maths, English and science and to get them right, which is why they are the subjects that are tested. In that regard, we want people to teach to those priorities. However, the vast swathe of teachers and schools up and down the country use tests appropriately. In order to help those who do not and to improve best practice generally, we are investing £150 million over the next three years on assessment for learning to improve the way in which the tests are used. In respect of the charge that tests are used inappropriately or for too many different things, it could be done differently. As some people argue, you could judge national performance on the basis of some kind of sample test. I am sure that that would be fine with regard to judgments around the national performance of the school system, but testing is about not only that, but parents being able to see how well their child is doing in the school system, pupils being able to see how well they are doing against a national comparator and parents being able to see how well individual schools are doing. If you want to do those sorts of things, some people would argue that alongside sampling you would have some form of teacher assessment. However, using teacher assessment to hold schools accountable would put quite a significant burden on teachers and assessment, so there would need to be some form of accreditation on how the assessment is done to ensure that it is fair and transparent and that it compares nationally. When I look at the matter and begin to unravel the alternatives and think about how they would work in practice, I find that the current SATs are much more straightforward—everybody would understand it. They are used for a series of things, and there might be some compromise involved, but the system is straightforward and simple, and it shows what our priorities are and gives us accountability at every level. I do not think that it is a mess at all.

Q330 Chairman: If you look internationally, you will see how such a system looks like an English obsession. Most other countries in the world do not test and assess as much as we do. The Welsh and the Scots do not do so, and nor do most of the countries with which we normally compare ourselves.

Jim Knight: I visited Alberta in November and found that it tests just as much as we do. In fact, we have shifted on the Key Stage 1 test in the past 10 years, whereas Alberta has continued with externally marked tests that are conducted on a single day. Alberta is one, but we could include Singapore.

Q331 Chairman: My brother and sister were born in Alberta, so I know a bit about it. It is hardly comparable to England, is it?

Jim Knight: In terms of international comparisons, which is what the question was about, Alberta is out there alongside the usual suspects—

Q332 Chairman: I meant countries like ours, such as Germany, France, Spain, Italy or the United States.

Jim Knight: Some parts of the United States, such as New York, do quite a bit of testing. Every education system is slightly different, and it is difficult to draw such international comparisons and say that this or that is exactly the same from one to the other. We have a system of accountability and testing; some countries test, such as Singapore or Alberta, but others do not. We think that we have struck the balance. Ofsted inspects schools, the Qualifications and Curriculum Authority independently monitors and regulates the tests, and the Office for National Statistics independently publishes the results of the tests, so the process is perfectly separated from Government. There is evidence that standards are consistently improving as a result of the use of the tests and there is good accountability to parents, which is important.

Q333 Chairman: The Government's watchword when it comes to education and other policies is "evidence-based". When you look at the evidence, are you sure that the testing and assessment method, which seems to have been uniquely championed in this country, is effective? Do you have any doubts at all about it? Is it successful? Does it give children in our schools a better experience and education than that provided by our competitors?

Jim Knight: I think that it is successful. When I look at how standards have improved since tests were introduced and since we increased accountability through tests and tables, I can say that they have worked. That is not to say that the situation cannot be improved. The Government are piloting, through *Making Good Progress*, single-level tests and testing when ready. As we signalled in the Children's Plan,

finding that those pilots are working may mean that we can evolve SATs one step further. That does not mean that we want to retreat from tests.

Q334 Mr. Slaughter: Picking up on what the Chairman has said, if I understood you correctly, you said that in relation to national policy or, indeed, national standards, which is whether overall standards of education and learning are rising, there are alternatives to testing every pupil in every school—in other words, it could be done by inspection, sampling or the like. A valid criticism might be that there is too much testing, which distorts the learning process, and that you could do it another way as regards national results and policy. Are you are defending testing every school on the basis of the effect on that school? Did I understand that correctly?
Jim Knight: Yes, I think that you probably have understood correctly. It is worth saying that no pupil spends more than 2% of their time taking tests.[1] Assessment, including tests, is and always will be part of teaching. The question then is whether we should have national tests and whether the amount of time spent taking and preparing for national tests is too stressful. I do not buy that. I think that life has its stresses and that it is worth teaching a bit about that in school. I do not get the argument. I visit enough schools where tests are used extremely well by teachers to drive forward and progress learning. In the end, I flatly reject the argument that there is too much testing.

Q335 Mr. Slaughter: It is the Government, not us, who are thinking of relieving the burden of vivas in foreign languages. Obviously, you are sensitive to the stress on the poor dears.
Jim Knight: I am delighted that you have brought that up. Certainly, the move that we are making on the oral examination for modern foreign languages is because of not stress, but standards. Ron Dearing has said that hit-and-miss, one-off, 20-minute tests in which you are coached to rote-learn bits of French, or whichever subject is being studied, are not serving us well. Controlled teacher assessment during the course that tests different scenarios in which people use languages is likely to improve standards significantly, which is why we want to do it. It is not because of stress.

Q336 Mr. Slaughter: I meant the stress on the examiners—the native speakers listening to their languages being mangled in those exams. Let us talk about individual schools. You gave the example of information for parents, so that they can look at league tables and select a school. That is free-market education, is it not? It would aid parents in selecting and migrating to schools, particularly if they have the time, knowledge, access to the Internet and all that sort of business in order to get hold of such information. Is that not part of the postcode lottery for schools or of the segregation or decomprehensivisation of schools?

[1] See the answer to Q 368 for correction and clarification of this figure.

Jim Knight: A lot of implicit values are tied up in that. I will not say yes, but it very much informs parents, which is a good thing. We explicitly want to move to a position in which parents choose schools, rather than schools choose parents, and I have debated that with the Committee in the past. We believe in parental choice—we can rehearse those arguments, if you like—but phrases such as "postcode lottery" involve separate issues from whether we should publish data about schools. Quite frankly, if we did not publish such data, there would be an outcry that we were hiding things, and the media would publish them anyway. I think that it is better that we put them out in a controlled and transparent way so that they can be scrutinised by bodies, such as this Committee, rather than leaving it to the vagaries of how newspapers choose to publish them.

Q337 Mr. Slaughter: Looking at the positive side of that, as far as the Department and the inspectorate are concerned, do you think that part of the role of testing in individual schools is to identify the performance of schools and of the teaching staff within them in order to alert you to failure or underperformance in particular?
Jim Knight: I write to the top 100 most-improved schools in the country every year. Testing helps me to identify success. I also keep an eye on those that are not doing so well, and my colleague, Andrew Adonis, does the same—perhaps he is the bad cop to my good cop. However, the data help us to manage the system. We are accountable to Parliament and are elected by the public in order to continue the improvements of the past 10 years in our education system.

Q338 Mr. Slaughter: I suppose that what I am getting at is that if—you might not be with me on this—one of the effects of publishing data is that parents who are savvy enough gravitate towards or even mutate certain schools, which results in more of a division between good schools and bad schools in an area, that would at least allow you, or professional educationalists, to identify underperforming schools and to do something about them through, for example, the Academy Programme.
Jim Knight: Yes, it allows us to identify areas where we need to intervene. If we did not have the tests and tables, something would be missing from the body of information that we recommend that parents look at when making decisions about which schools to choose for their children, but they should not be seen in isolation. They are very simple and easy for people to understand—they are easier than leafing through Ofsted reports, which we also recommend—although perhaps not as easy as chatting to other parents in the neighbourhood or going to visit the school itself, which are the sorts of things we expect parents to do. However articulate parents are, and however much technology they have at home, those are the sorts of things that we expect them to do when choosing schools for their children.

Q339 Mr. Slaughter: One aim of the Academy Programme, as I understand it, is to target underperforming schools, particularly in areas of deprivation, and to put good schools—whether new or replacement schools—into such areas. Do you see tests in the same way? Do they enable you to focus resources on areas of deprivation or underperformance, rather than simply to present information to third parties so that they can deal with such things?
Jim Knight: Undoubtedly, they are an indicator that we use. They are not the only indicator—we, too, look at Ofsted reports and other factors, such as attendance rates, when assessing how well a school is doing—but they are undoubtedly the prime indicator. We have explicitly set targets for the number of schools with 25% or fewer pupils getting five A*s to C at GCSE and we now have the targets for 30% to get five higher-level GSCEs, including English and maths. Ten or 11 years ago, half of schools did not have more than 30% of pupils getting five higher-level GCSEs including English and maths. That is now 21% of schools, but we have further to go. That measure helps us to target schools, and we are doing work on that right now.

Q340 Mr. Chaytor: Minister, the Department's submission to the inquiry describes the arrangements at Key Stage 1, saying that, "The child will not necessarily recognise a difference between the formal tests and tasks he/she completes for other classroom exercises." If that is important at Key Stage 1, why is it not important at Key Stages 2 or 3?
Jim Knight: I will let Ralph make a contribution, because he has been sitting here very patiently, but I would say that there is a difference. When you look at the age at Key Stages 1, 2 and 3, there is clearly a significant age difference, and with that, in general terms, there is a difference in maturity. There comes a point when it is appropriate to start introducing young people to the pressures of assessment. Those are pressures that we all live with throughout our educational careers; we have to start getting used to that at some point, and I think 11 is a better age than seven.

Q341 Chairman: Ralph, I hope that you do not feel neglected. The Minister has said that you have been very patient. Will you catch my eye if you want to say something, and we will welcome you in?
Ralph Tabberer: Thank you, Mr. Chairman. I endorse what the Minister has said. We try to take decisions about assessment that suit the context and the particular teaching and learning environment. We will perhaps look at the use of more controlled assessment and teacher assessment, where they offer us a better alternative. That might, for example, be when young people are younger; there may be more variation in their performance on particular days, and such assessments may be more sensitive. We would also look at the use of more controlled assessment or teacher assessment in areas such as applied learning. There are aspects of applied learning in Diplomas that will not be as susceptible to an external test.

Q342 Mr. Chaytor: When we get to Key Stage 2, the judgment is made on tests that last about 45 minutes. How does that equate with the Minister's criticism a few moments ago of what is now the old system of language orals? You said, "We want to move away from the hit-and-miss, 20-minute test in which you are coached to learn." How can it be wrong that there is a hit-and-miss, 20-minute test, but right that there is a hit-and-miss, 45-minute test?
Jim Knight: In respect of the oral examinations for GCSE, those are the qualifications that you take with you through your life. I cannot remember whether I got a B or a C for the oral.
Mr. Chaytor: I am sure it was a B.
Jim Knight: Well, I got a C for the written one and a B for the oral or *vice versa*. I cannot remember which way round it was, but I do remember the oral exam. You carry that with you. I cannot imagine that many people remember their SATs scores—I do not reckon many of us were young enough to take them.

Q343 Mr. Chaytor: But the tests determine the primary school's position in the league tables and the pupil's self-esteem when they enter secondary school. My question is why are the Government so hung up on the single test at the end of Key Stage 2.
Jim Knight: It may be that we are not. It may be that if testing when ready and the single level tests prove effective in raising standards, we will be able to move to a position in which you have a number of test windows during a year—there are currently two, but we might be able to develop that further—and it is not necessarily all about how everyone did on a rainy Monday afternoon in English and a rainy Friday afternoon in maths at the end of Key Stage 2; it can be throughout that key stage.
Ralph Tabberer: I add to that that the judgment we are making is about the context—the type of learning taking place—and an oral assessment looks to us better placed as a teacher assessment rather than as an external exam. In relation to the end of Key Stage tests, there is also an issue in every assessment of manageability. If we go back far enough in the history of the testing regimes, at key stages there was experience of using teacher assessment. That proved, in the early days of the national curriculum, very unmanageable. It meant that we were losing huge amounts of teacher time to moderation, which was not proving terribly effective. It is about trying to judge the best kind of measurement—the manageability of the measurement and the validity of the measurement.

Q344 Mr. Chaytor: In getting an accurate picture of a child's ability in year 6, is it not more valid to have the result of teacher assessment throughout the year as well as an external test, rather than relying simply on the external test?
Chairman: What is the point of personalised learning?
Jim Knight: That is the excitement of the progression pilots. The current situation with SATs is that everyone takes the test and then the examiner decides which grade you are at based on your response in that test, whereas the single level test is a

scenario whereby the teacher assessment informs whether the child is ready and what level the child is put in for, so the test is used as part of teacher assessment for learning, rather than sitting alongside it as it does at the moment.

Q345 Mr. Chaytor: My other question is this. In moving away bit by bit from the regime that was inherited in 1997, will you accept that there is a link between a very rigid testing regime and disaffection and demotivation among children who do not perform well under that kind of regime?
Jim Knight: I think that it would be a very tenuous link. You see schools that are performing very well in very difficult circumstances. Obviously, part of what they are doing in performing well is that a large number of their pupils are doing well in tests. Why are they doing well? Which comes first, the chicken or the egg? I think in this case it is getting the behaviour, ethos and atmosphere in the school right, and getting people focused on their learning, which means that they are not disengaged. What then subsequently happens is that they do well in their tests, but my own feeling would be that you would be getting it the wrong way round if you said that because they are not doing well in tests, they are disengaged.

Q346 Mr. Chaytor: In any high-stakes testing system, 80% pass, but 20% fail. I am interested in whether there is any link between the sense of failure and loss of self-esteem of those who so publicly fail, and disaffection in the early years of secondary—Key Stage 3—which is a major concern of the Government.
Ralph Tabberer: First I question the premise that these are, in conventional terms, high-stakes tests. We normally talk about high-stakes tests as determining for pupils the school they go on to within a selective system. Within our assessment history, if we go back 20 or 30 years and look at tests such as the 11-plus, those might legitimately be called high-stakes tests for children, because they were so determining of the next stage. We have got medium-stakes tests for our students that allow them to show what they can do and give them and their parents a sense of where they are. They also happen to give us very useful information, as Mr. Slaughter has indicated, for policy development and accountability. The Minister is right to point to the progression tests as an interesting experiment. What we have been keen to do is to offer Ministers alternative approaches. We have listened like you to comments over the years about the possible downside of what you termed rigidity. We have listened to people talking about the possible effect on the year 6 curriculum and the possible effect on the pace of the whole key stage. So looking at a progression test as an alternative, the idea of actually being able to draw down a test to be ready on time for pupils may give teachers and pupils a different environment. We think it is appropriate to pilot that, but we do not think it appropriate to rush for that solution. We want to give Ministers alternative options, so they can deal with just that sort of question.
Chairman: We now move on. John Heppell will lead us on the notion of centralised control and validity versus reliability.

Q347 Mr. Heppell: Looking at how the Government are addressing A-levels and Diplomas, people might think that there has been a subtle change, in that whereas we were moving towards greater reliability at the expense of validity, there has been a slight move the other way. Many universities have said to us that people come to them without a sufficient breadth of skills in a particular subject. We have heard from examination boards that, instead of being closed, the questions are now, if you like, opened out, or open-ended. Obviously there is a consequence to that, and I know that such things are finely balanced, but can you confirm that there is a move to ensure that validity goes up a bit in the rank, as against reliability?
Jim Knight: We want both, obviously, and we will continue to evolve and improve the A-level as we introduce the Diplomas. We are mindful of the criticism, which we have heard from both employers and universities, that people may be well versed in the particulars of their subject, in which they perhaps took an A-level, but they need to do better in terms of some of the wider, softer skills. That is why we are introducing measures such as the extended project into A-levels. That is also why we have introduced personal learning and thinking skills and why the work-related learning that runs through the Diplomas in order to offer what both universities and employers are saying they want more of from our young people.

Q348 Mr. Heppell: Moving on from that slightly, you now have what is called controlled internal assessment. However, we are told by our advisers that there has always been controlled internal assessment. You have not actually taken the coursework into account in assessing that. You are taking an add-on to the coursework and assessing that. Is that not the case? Are you not interfering, from a centralised position, with what should be a creative process? I understand that you set down the guidelines fairly rigidly in respect of what the controlled internal assessment—the very name says it—does. Is it a lack of faith in teachers?
Jim Knight: No, I do not think that it is a lack of faith in teachers. Again, we had to design an examination system that retains the confidence of everyone that it is serving and those involved, including the pupils—most importantly—the teachers and parents, the employers and those in further and higher education. We found that an over-emphasis on coursework in some subjects was problematic, so we have moved away from that. The use of controlled internal assessment is, perhaps, a halfway house between the examination hall one afternoon and the continuous assessment of coursework. Ralph, do you want to add anything to that?

Ralph Tabberer: Again, I think it is a question of looking at different subjects and seeing which is the right design that suits them. There are some subjects for which coursework is a natural or highly desirable assessment method—art, for example. There are other subjects for which external assessments work almost fully. For example, we have moved more of maths into that realm. We have been trying, with the controlled assessments, to create a more controlled environment where it is more likely that the assessments made by one teacher will be replicated by another. That addresses public questions about coursework, as the Minister suggests, and about the possibility that there is variability, which affects GCSE results.

Q349 Chairman: Is this an endless search for an accurate method of evaluating the teaching and the quality of knowledge that the child assumes? It is endless, is it not? Does it not squeeze out the thing that John is pushing you on: the creativity—the breadth, depth and the imagination of it? The Welsh have got rid of it. Are they struggling because they have got rid of this testing?
Ralph Tabberer: Any assessment system is a design where you are trying to balance validity, reliability and manageability. You try to get the best design for your whole system. I think that we have been very consistent, actually, with the principles that were set out at the start of the introduction of the national curriculum assessment. We have tried to stick to those principles in measuring and giving parents and pupils information about what they can do. We have made changes when there has been a build-up of concern and we have felt that it has not been possible to answer that. So we have moved when things have been unmanageable. We have not been inflexible. The basics are still there. Again, if we find better ways of assessing, we will put those options to Ministers. I suppose that one of those areas in future will be IT-delivered testing. We should certainly keep our eyes open for alternatives that give us the best balance.

Q350 Chairman: Is there a horrible generation in the Department that read, as I did as a young man, a thing called "The One Minute Manager", the central theme of which is that, if you cannot measure it, you cannot manage it? It seems to me that the Department is still desperate to measure all the time. They do not measure so much in the independent sector, do they? That is not the way they get good results, is it? Not through constant measurement.
Jim Knight: I am a product of an independent school where I was tested an awful lot. That is part of a traditional elitist education, I think. It is an endless pursuit because the economy is ever-changing and we are ever-changing socially. The environment in which schools and the education system are operating is ever-changing, and education has to respond to that. It therefore has to keep changing. The environment that I grew up in and in which I went to school was one in which, if you were lucky, 10% went to university. However, skills needs have changed, as we discussed in other evidence sessions. We therefore need to change the qualifications to respond to that change—and as you change the qualifications, you change the forms of assessment.

Q351 Mr. Heppell: Have you actually come across problems on the ground? Has somebody that is studying or somebody that is teaching said, "Look, it doesn't work like this. We need to have more flexibility in the way we deal with it"? Are there problems on the ground?
Jim Knight: Specific problems?
Mr. Heppell: Problems specific to the controlled assessment rather than just an assessment of coursework.
Chairman: Ralph, perhaps you should answer that. You used to be in charge of teacher training.
Ralph Tabberer: I am trying to think of any particular cases where people have brought up additional problems relating to controlled assessment, but I cannot think of a piece that does that. In general, though, I am clear that we do monitor the impact of assessment; we monitor not only the impact on pupils and schools but the opinions of different parties. We keep that in view, and as I tried to say earlier we are willing to change, and when we can we put alternative proposals to Ministers. We think we have something with the progression tests that might give an alternative approach, and Ministers have been quick to say, "Well, let's pilot it. Let us not implement it until we know more about how it might impact." That is all evidence of us being intelligent and open. We keep on looking for improved solutions, but not moving away from the basic principles on which the current model was developed.

Q352 Chairman: The desire to measure is not driving out the imagination and joy of education?
Ralph Tabberer: It should not, no.
Jim Knight: I thought of this Committee when the announcements were made last week around the culture of entitlement and the importance of greater partnership. Similarly, we have a commitment to sport in the curriculum that we are developing, and we have a cooking announcement. Some of these things are not that easily measured. The pilots on culture will test how easy it is to measure the five hours in the curriculum. We are ever-evolving about this. Some things are much easier to measure than others, but the pressure that I was talking to John about in respect of employers and universities around the softer skills is more difficult to measure—but that does not mean that we are not committed to trying to work harder to develop that.
Ralph Tabberer: Of all the schools that I have visited, I cannot think of one that is immensely creative that is not also interested in the tests and doing their best by them. I cannot think of a school that I have visited that does well in tests that does not have a strong creative side as well. Sometimes we set these aspects as alternatives, but I do not think that that is always fair. There are plenty of schools that manage to do both very well indeed. They are well led, and they know what they are doing. There is plenty of evidence that you can have creativity, a lot

of autonomy and a lot of self-determination by teachers and that you can have a properly assessed system that gives parents a good account of what is happening in the school.

Chairman: Let us drill down into testing and school accountability with Annette.

Q353 Annette Brooke: I want to look at whether the test is fit for all the purposes that we try to use it for. I do not think anyone would disagree that there should be some form of measurement of pupils' progress. But is not the difficulty that the Government are placing so much reliance on a test that was designed for one purpose but which is now being used to measure whole school performance? Do you not have any concerns about that?

Jim Knight: If it were the only measure of school performance and the only aspect of accountability, one would have to be concerned that we were putting all our eggs in one basket. But we are not. We have inspection and we look at performance in other areas. We only test a few subjects through the SATs. We are still looking at how people are doing on other things. We also have national strategies working in some other areas. So I would say that it is a critical measure but it is not the only measure. Therefore, I am happy with how it sits.

Q354 Annette Brooke: Many parents will focus only on this as a measure. Personally, I can feel fairly relaxed that a local authority is looking at the whole school, because it might set off some warning signs that need to be dipped into. However, the majority of parents are not going to dip below what they see in their local newspaper. Therefore, do you not think that this measure is harmful to the idea of parental choice?

Jim Knight: It goes back to what I was saying before. We do not publish ranked tables; the newspapers choose to rank locally what we publish. I have spoken to various people who work in the media who were a little sceptical about whether they should publish the tables, but when they saw how well rival newspapers sold when they published them they soon leapt at it and published them too. There is no doubt in my mind that if we did not publish the tables someone else would. As I said before, if we as a Department do it, we can be scrutinised; the process is carried out independently using national statistics, and we know that it will be done objectively and fairly, rather than someone who is not subject to as much scrutiny being able to do it. Of course, our local newspaper, *The Dorset Echo*, regularly reports the Ofsted results of schools as and when they achieve them. They usually publish the successes and the pictures of celebrating pupils, teachers, head teachers and governors, rather than those schools that get satisfactory ratings, but obviously they will occasionally run stories on those schools that are not doing so well. I think that those stories, along with what parents say to each other, are as informative of parental choice as the tables.

Q355 Annette Brooke: I would still disagree that parents have full information. For example, someone told me recently that a certain selective school has 100% five A to C grades at GCSE, and said, "Isn't that fantastic?" That is the perception, is it not, because of the way that league tables are presented?

Jim Knight: Again, I cannot be accountable for how league tables are presented in every single newspaper. We publish the contextual value added as well as the raw scores. We are now moving to progression targets, so that we are looking at the proportion of schools that progress children through two levels for each key stage. So we will be reporting a number of different things. We are putting a science and language into the indicator, so that all that data will be available in the attainment and assessment tables. However, we cannot tell newspapers how to report the tables.

Q356 Annette Brooke: May I just dig in to the Contextual Value Added measure? We have had varying evidence on this measure. I think that there was one throwaway comment that it was really just a measure of deprivation. There is also the issue that not all parents will understand the significance of the measure. Is there more that you could do as a Department to make the measure stack up better and be genuinely more informative for parents?

Jim Knight: I would not say that the measure is perfect, and I will let Ralph give the technical answer on CVA in a moment. However, one of the reasons why I have been pushing on the progression measure is that it is slightly easier for people to get their head round, as to how every single pupil is progressing. So it is not a threshold but something that applies across the board. I think that that will help in respect of the concerns that you raise.

Q357 Annette Brooke: I would like the answer to my question. However, I would like to pick up on that particular point about progress. An admirable school may have a high percentage of children with special educational needs. Up to 40% of its children may have special educational needs. It is quite likely that those children will not be able to progress more than one level in the standard assessment tests over a given period. If you use that information across the whole school it will add even more distortion to the picture.

Jim Knight: I am not sure whether there will be any more distortion than there is at the moment. It is a reasonable criticism. When we introduce the foundation learning tier, which can accredit progression and learning below Level 1 in national vocational qualification terms—it is very confusing having national curriculum and NVQ levels—we may be able to look at whether that can be reflected. At the moment, if you have a high proportion of children with SEN, you will not do as well in the raw scores as those schools with a lower proportion.

Ralph Tabberer: The most important thing is to see the Ofsted inspection as the top of the tree. For parents, who are your concern here, the Ofsted inspection is probably the most rounded, richest and

most comprehensive assessment that they will get of a school's strengths and weaknesses. I would always point parents to that assessment as the best thing to consult. When we publish results, we try to ensure that the public can see raw results and that they can look at comparators and benchmarks. We have had a lot of discussion about which way to approach value added. In our consultations on that, we have settled on contextualised value added as the most fair. In trying to publish series of data, we are following the principle of being transparent about all of the analyses so that parents can access the information that they understand or the information that they want. I have to say that we get very few complaints about the testing regime.

Q358 Chairman: That is because they cannot understand a word of it. You have to understand it to be able to complain about it.
Jim Knight: They could complain that they cannot understand it.

Q359 Chairman: That is true. Come on, look at your site. Get a group of parents to look at the site and evaluate how much they understand the contextual value added score. It is very difficult to understand. Why present it in that form?
Ralph Tabberer: Equally, why hold it back? What I am saying is that we know that many parents consult Ofsted reports. We know that in putting those Ofsted reports together, the school, in its self-evaluation, and the inspectors will draw on all those analyses. There is no reason for us to hold back those analyses. What we do is make them transparent.

Q360 Chairman: You are missing the point. We represent a broad swathe of population in our constituencies and we want the information to be intelligible to people with higher education, lesser education and very little education. We want them all to be well informed. In the way that you present CVA scores, what you have set up is a system that is only understandable to people with higher levels of qualifications. That is unfair.
Jim Knight: I want to challenge that if I may. I do not think that it is that difficult to understand that in CVA terms, 1,000 is the norm. If you are above 1,000, you are adding value better than the norm. If you are below 1,000, you are adding value lower than the norm. If that is all people understand, then it is pretty straightforward.

Q361 Mr. Chaytor: Surely, the real point is that the significance of the degree to which it is below 1,000 is unclear. What is the Government's resistance to a simple banding system or a five-point scale, from excellent to poor, to rate a school's value-added performance? Would that not be easier? We use simple banding systems to describe most other public institutions. Yet this value-added concept is a—
Jim Knight: A star system.

Q362 Mr. Chaytor: What parents want to know is to what degree their school differs from what could reasonably be expected. As it presents at the moment, they just cannot work that out.
Jim Knight: The problem is that you would be lumping lots of different judgments together. We would have constant Select Committee inquiries into whether it was a fair way in which to lump everything together.

Q363 Mr. Chaytor: That is what the Ofsted report is.
Jim Knight: The Ofsted report gives a series of judgments under a series of headings.

Q364 Mr. Chaytor: It comes together under one scale.
Jim Knight: Yes, the Ofsted report is the thorough, authoritative reflection on a school, whereas, finding ways to lump things together using year by year assessment and attainment tables would make us vulnerable to criticism and questions such as whether we left out soft skills, science, foreign languages and so on. There are many ways of making judgments about a school. You could say the same about a hospital to some extent, but a hospital's performance is rated by inspection.

Q365 Annette Brooke: Can I ask you to have a look at how informative CVA is for the vast majority of parents? The vast majority of parents do not really appreciate it and do not take it on board. They still see, in a selective system, that school X must be better than school Y, because it has better figures according to raw results. School Y might be doing fantastically well, but that is not the message that comes out.
Jim Knight: Annette, we would always look seriously at the Committee's recommendations, and I shall look out for that one in particular.

Q366 Chairman: It is not rocket science. Do a quick little test—anyone could do it. Get the Institute of Education to see how understandable it is. You would not have to take a big sample—you could simply test how many people easily understand it, as Ralph said, and sample by class and educational background. You could do it in a week.
Jim Knight: I will reflect on the wishes of the Committee.

Q367 Annette Brooke: Finally, even if the league tables really work and if they convey something or other to parents, you often argue that they are driving up standards. What is the causal link between league tables and the driving up of standards? Do you genuinely have evidence for such a statement?
Jim Knight: We have discussed this afternoon the nature of testing and the publication of results in tables, and their use in making schools accountable. Part of the debate is about whether the tests are high stakes. Schools take how well they are doing in the tests really seriously, which drives forward their literacy and numeracy priorities. Getting things

right in English, maths and science is a priority. There is evidence that such sharp accountability has driven things forward in those subjects.

Annette Brooke: I shall leave the drilling on that question to my colleagues. You write to your 100 most improved schools, but are they most improved on raw results, on CVA or on both?

Chairman: Let us move on. Fiona, do you wish to ask a question about the unintended consequences of high-stakes testing?

Q368 Fiona Mactaggart: Minister, you said earlier that no individual pupil spends more than 2% of their time on taking tests. That might have been a mis-statement: Sue Hackman told us that no one spent more than 0.2% of their time preparing for tests, but David Bell said in writing that that meant taking tests. Do you have any evidence to show how long pupils spend on revision and preparation for tests?

Jim Knight: I do not have those statistics. Key Stage 2 tests take a total of five hours and 35 minutes in one week in May, so the amount of teaching time taken away so that pupils can sit the test is 0.2%. For Key Stage 3, seven hours and 55 minutes, or 0.3%, is taken away. Those are the figures that Sue quoted. When I recalled that it was 2%, I should have said 0.2%. However, I do not have any exact statistics on the average amount of time anyone spends preparing for the test, which would be hugely variable. With some schools—and I think the ideal is that they would just integrate it into their learning—there would be a certain amount of preparation for taking a test, because it is just good practice to instil in young people the belief that when they are about to take an examination they should prepare for it. I prepared a little bit for this hearing, believe it or not. However, I do not know exactly how the figure might average out across the country.

Q369 Fiona Mactaggart: Would you be surprised to learn that the Qualifications and Curriculum Authority did a survey of primary schools that showed that at Key Stage 2, in the four months before the test, it was 10 hours a week on average per pupil? That is nearly 50% of the teaching time available.

Jim Knight: I have not seen that research. I do not know whether it is something with which Ralph is familiar.

Ralph Tabberer: I am certainly familiar with the QCA research. It goes back to what I said earlier about the problems that have surfaced from time to time regarding the possible impact on the year 6 curriculum. That is why we are looking at—and we always listen to questions raised by the profession and by non-departmental public bodies—the impact of the current system.

Q370 Fiona Mactaggart: What effort is the Department making on this? Of course public examinations require rehearsal and revision—I have no doubt about that—but the Key Stage tests were not originally conceived as that kind of test for pupils. If half the time of a Key Stage 2 pupil is taken up with revision for the test, that is time when they are not learning new things.

Jim Knight: I shall let Ralph come in on this in a minute, but I dispute that, and would be very surprised if there are young people who are just sat there revising when they know it all. If they are spending some time making sure that when they complete year 6 they have the necessary maths, English and science skills to be able to prosper when they get into secondary and move into Key Stage 3, I do not have a problem with that. I do not have a problem with their being taught the things they need to be able to pass the test, even if that means more catch-up classes, or even if it means a shift in the amount of time being spent on the priority subjects in their final year in primary.

Ralph Tabberer: I am sorry if I nearly interrupted my Minister, but it was only to be just as quick in disputing the premise that revision is wasted time. It is enormously important that young people are as prepared as possible for Level 4, particularly in English and maths, so that they are ready to access the secondary curriculum. I am not concerned if that is a prime interest for teachers teaching in year 6. We know there is a tremendously strong relationship between pupils who attain Level 4 in English and maths at that age and their results at GCSE, and if we can give more young people access to that level we are sure that they make a stronger transition into secondary schools and are more likely to succeed. The threshold is not there by accident, and I do not think we should treat revision as necessarily a negative.

Q371 Fiona Mactaggart: I do not do so, but I am concerned about standards. I have only one concern, and it is about standards. There is a balance between testing and standards, and testing is the way in which you assess whether a child has achieved a standard. You might want to anchor a child into that standard, but I am concerned that our focus on testing may—I am not saying that it does, but there is a risk that it will—interfere with the drive for standards. In a way it can become a substitute for standards. For example, Mr. Tabberer, your experience is as a teacher educator and the head of teacher education. Would you say that well-implemented teaching for learning, and that kind of in-teaching assessment, has the capacity to drive up standards faster than the constant testing of children?

Ralph Tabberer: Yes, I believe that assessment for learning is an immensely important and continuous part of the teaching and learning process. I also believe in occasional external assessments to lift performance and give young people the opportunity to perform at a higher level. Both have a beneficial effect, and I have never said that one is more important than the other—both are of value. The helpful thing in your distinction is that we do not want to see children being drilled so that they can just repeat low-level processes accurately and get marks for that—we are all clear that we do not want that. This is where I turn to the professionalism of

teachers, and when I talk to them, I do not hear that they are engaged in that process—they are trying not to drill, but to prepare pupils so that they can do their best in these tests. We have a good balance. The evidence is saying that if there is anywhere in our overall system where we need to invest in assessment, it is in more assessment for learning, so you are right.

Jim Knight: Which we are doing, with £150 million over the next three years.

Q372 Fiona Mactaggart: So far, our efforts to implement it across the board have not been as good as they should have been. Is that not the case?

Jim Knight: We can always do better.

Q373 Fiona Mactaggart: Let us briefly take this back to the point that Annette raised about the results of tests being used for other purposes. I do not think that most teachers drill pupils, but some do, and my anxiety is that that is partly because we use the tests in the ways that we do. There is a risk—I would like your views on this—that drilling might become more, not less, prevalent in the single level tests, although there is a possibility that it might become less prevalent. However, I would like your opinion on the fact that the research shows that about 30% of pupils at Key Stage 2 are misallocated levels just because that happens. About 15% are allocated a level below that which they should have and about 15% are allocated a level above—I might have got the margins slightly wrong, but that is how I read the research. One anxiety about the single level tests—this is great for the individual pupil—is that once you are through the gateway, you are safe. My anxiety is about pupils getting unreliable success results, although that would be good for those pupils and might motivate them, with good consequences. However, because we use the tests to measure schools, there is a real incentive for teachers to push children through the gateway. I have not seen any evidence that the Department has addressed the damaging consequences of what the evidence suggests is really going on.

Jim Knight: We are obviously in the early days of the pilot on the single level tests, and we have had only the December round of testing. It is worth noting that we made the decisions about changing the measures on the tests in November, before the December tests were even taken, let alone before the results were known—I say that for the benefit of any media representatives listening. We will see what happens in those pilots, but one thing that was a bit weird about the patterns from the December tests was the number of entrants who were put in at the wrong level. As things bed in, teachers will understand the importance of the assessment of their children's learning and the fact that these are pass/fail tests. There is a big difference from the SATs as they stand, where the examiner makes the assessment of which level the pupil is at. In this case, the teacher makes the assessment of which level the pupil is at, then the examiner checks whether the teacher is right. That changes the terms of trade quite significantly. It puts more emphasis on the teacher's own assessment. That is why assessment for learning is built into the *Making Good Progress* pilots, alongside one-to-one tuition, progression targets and incentive payments.

Q374 Fiona Mactaggart: I am trying to refresh my memory about your letter to the Committee. As we saw it only this morning, I might be wrong, but one of the things that you were wondering about—this was on the third page of your letter—was whether part of the reason for the wrong entry might have been that pupils were not being focused and prepared in the way that they have been. I am quite interested in this issue. Let me explain why. I am an MP for an area that has the 11-plus. When it was originally conceived, the 11-plus was said to be an assessment of where pupils were at and not a consequence of drilling and so on. Of course, ambitious parents drill their children extensively. Of course they do—they pay for tutors if they can afford it—because it makes a huge difference to pupils' performance, as a result of which it is not the kind of assessment that it was in the days when I did the 11-plus. I went into school one morning and the exam was stuck in front of me; I did not know that that was going to happen that day. Today, it is quite a different experience. I think it would be a good thing if we could, in the single level tests and in Key Stage 2 tests, make that the norm. It would create a blip in results in the short term, because of the lack of preparation, but it might tell us better truth about what pupils know, and mean that teachers could focus on getting the pupils' understanding strong rather than on getting pupils through examinations.

Jim Knight: I think I am with you on this. However, I would sound a note of caution, in that I do not want to take the pressure off.

Fiona Mactaggart: Neither do I.

Jim Knight: I know you do not. We are explicitly designing this to drive progress for every single pupil, regardless of their starting point, because of the concern that some people have expressed about the current situation, in which, let us say, at the end of Key Stage 2 there is too much focus in some schools on people on the margins of a Level 4 and not on the rest, because that is where the measure is. This is about every single child making two levels of progress within each key stage and being tested when they are ready, so the testing is the culmination of learning when the child is ready to take the test, rather than everyone being pushed and drilled for an arbitrary date in June or May or whenever it is. That feels like a good model to explore in the pilot. In relation to ambitious parents, I think the answer is to try to get every parent as ambitious as the next for their child. I would love an aspect of this to be parents asking at parents evenings or in e-mails to teachers when their child will be ready to take the next test in reading, writing or numeracy, so that there is a bit of a push in the same way as there is with the music grading exam. In the days when my daughter was young enough to take her cello exams, when I saw the cello teacher I would ask when she

would be ready for grade 4, grade 5 or whatever. Just that little bit of push in the system for each individual is not a bad thing.

Ralph Tabberer: I agree entirely that whenever you pilot or look at an alternative approach to assessment, the thing you have to do, as you are suggesting, is look at the changes that causes in teacher behaviour, pupil behaviour and parent behaviour. That is precisely why we want to go through this pilot exercise. When you are looking at a pilot and weighing it against the strengths and weaknesses of an existing system, you are asking yourselves questions about whether it might cause less time in year 6 to be devoted to revision, less time to be devoted to drilling and so on. I think we have to go through a couple of these rounds of the progression tests and look quite closely at whether those are the problems, or whether we get a new set of behaviours. At the moment we are very open to those possibilities. You can always set out a theory that a new assessment form could cause this or that, and there are plenty of people out there with experience of assessment—and some without—who will proselytise for different theories about what might happen. We clearly want, with the progression test, to change some behaviours, say earlier in Key Stage 2—to get some of the questions that are being asked about the progress of all pupils asked earlier in the key stage than may be the case now. If we can make it more obvious that children are perhaps progressing more slowly than we wish in year 3 and year 4, that could be a good effect, but if we misjudge the accountability to the point where everybody feels that they have got to drill in order to prove worth, then we have gone too far. These are very subtle things, but these are the crucial questions that we have got to ask.

Q375 Chairman: I have to say that your language worries me, because it is you, and then the Minister, who have kept talking about drilling—drilling, driving and pressure. That language seems to me all wrong in terms of the educational process that Fiona is probing on, in the sense that you set up a system that has an enormous amount of testing in it; you incentivise teachers to achieve on that business of testing and being successful; and you seem not to be able to step outside that and say, "But what does this achieve? What do independent assessors, researchers at whatever institution, tell us?" It is an internal world; you seem to glorify testing and believe that it is going to lead to a better quality of education for the children. It worries me tremendously when you talk about it, and when you brush aside the fact that it could be 50% of the time in school spent on trying to get the kids to achieve on the test. In the school where 30% or 40% of children have special educational needs, and there a lot of poor kids, I have a feeling that the percentage in those schools, particularly perhaps the ones that might just make it with extra drilling, would be much more intense than 50%. It worries me that this is not the world that I see when I visit schools—the world that you describe. They seem to be lost in this drilling, driving and pressure.

Jim Knight: Fiona started it with the drilling.

Chairman: No; you guys came up with drilling.

Ralph Tabberer: I have clearly given the wrong impression if you think that we just, in your words, drive this as an internal system. Far from it. We do not just sit in Sanctuary buildings and imagine what we think will be an effective system. We do a lot of work looking at what the research tells us is working, and what is not working. I would say there is an immensely powerful trail of evidence that our approach to assessing has been very effective over 20 years.

Q376 Chairman: So there is no evidence that the marginal student gets even more pressure to get to that next level? Is there any evidence that those people who teachers think are never going to make the standard are just left in the wilderness?

Ralph Tabberer: I accept that there are real questions about where the onus of attention goes with any external tested system at the end of a key stage. Again, that is why within the Department we have been so interested to move towards progression as a new model, looking at the progress that pupils make across the key stage. It is just as important to us that a child who is working at Level 1 gets to a Level 3 as that a child who is working at Level 3 gets to a Level 5, through the key stage. Far from being locked into just one approach we are much more concerned with the overall range. Where I perhaps disagree with you, I am not sure, is that I believe that measurement helps us to understand where a child is and gives us a sense of where they are going. It helps to give the parent and the child that sense, and it helps to give the school that sense. That is itself worth having. I think that helps to open up the secret garden.

Q377 Chairman: Would not a qualified and perceptive teacher give you that?

Ralph Tabberer: Yes, but when you are trying to use a system also for public accountability, as we are doing, you are looking for a manageable solution. I believe that our system of external testing creates the best possible model. Indeed, in terms of the international investigation of different models, we get a trail of people coming here to look at the way our system works, to look at the power of the data that we have available and to look at the willingness with which we have been able to confront areas of failure—areas of failure for the disadvantaged as well as the advantaged. Indeed, if you look at the recent comments from colleagues in OECD they point to our system as having probably more of the components of a modern and effective education system than any other they know.

Chairman: You have been very patient, Annette, in getting to your questions.

Q378 Annette Brooke: I am still rather concerned about children who would only be able to progress one level being ignored. In many cases, it is probably a cause for great celebration that those children do progress that one level. The teacher will be trusted to celebrate that, I guess. Why, on perhaps the easier

aspect of going up two levels for children of a certain ability, are there financial incentives? It seems to me rather topsy-turvy, and we may be in danger of not putting enough incentives in the system for children who, however high the quality of teaching, will find it much more difficult to progress.

Jim Knight: We have yet to announce exactly how the financial incentives within the pilot will work. My inclination is to focus around those young people who have not been making the sort of pace of progress they should, so rather than just paying money for those who you would normally expect to do well, you focus the incentive around those who have not been doing as well as they should, who have not been making the pace of progress that you want and being able to reward those schools—not teachers, but the schools themselves—if they manage to achieve that. We will make some announcements on that fairly shortly. As for those who are making only one level of progress during key stages, obviously there are some with special educational needs where that might apply. It is worth bearing in mind that, as Ralph said, the new system will celebrate as much someone moving from 0 to 2, or 1 to 3, as it will someone moving from 5 to 7. That is a significant step forward in terms of a system that rewards improvement across the whole ability range.

Chairman: We must move on.

Q379 Lynda Waltho: Minister, thank you for your letter, although for me it arrived a bit too close for comfort.

Jim Knight: I have apologised to the Chairman.

Q380 Lynda Waltho: I did not know what you were saying to us, but I have now had a chance to read it. It answers some of the questions that I was forming. Interestingly, you refer to some unexpected patterns in the results. Indeed, Sue Hackman said the same in her letter to schools in January. You go on to say what might have caused those unusual patterns, but you do not say what the patterns are. I wonder whether you would expand on that? You touched on the subject slightly with Fiona, and I wonder whether you would expand a little on what those patterns might have been.

Jim Knight: As I said at the outset, we will publish a proper evaluation of the December and June tests in the autumn as it can be analysed more fully than in the early stages of a pilot. We should bear in mind that it took four years for the SATs to be piloted. All of these new tests take some time, and you will inevitably have some teething troubles. We do not publish as each of the test results come out. We do not publish them in a drip-drip fashion; we tend to do an annual publication of test results. I do not think that we should do anything particularly different for this, because it might skew things and put undue pressure on those schools that are in the pilot scheme. As I said in the letter, the most significant unusual outcome was variations between Key Stage 2 and Key Stage 3 pupils taking the same test. So, let us say that they were taking a Level 4 writing test. The Key Stage 2 students were doing significantly better when they were taking exactly the same test as the Key Stage 3 students. Now, that was a bit odd. We will have to wait and see why that was the case. It might just be that the sorts of scenarios that they would have been writing about in that test were more engaging for younger children than for older children; I do not know. Maybe there are issues of motivation in Key Stage 3 that are different from Key Stage 2 around taking the test. There were some other patterns around the higher levels and the lower levels, and the expectations were different between the two. However, when we had a first look at the overall patterns that were emerging, we just thought that there were enough oddities that, although they are not out of keeping at all with early pilots, we should ask the National Assessment Agency to run some checks and make sure that the marking was right before we gave the results to the pupils and the schools, which we have now done.

Q381 Lynda Waltho: There is a perception that part of the problem might have been a higher rate of failure, if you like.

Jim Knight: No, it certainly was not about the results. It was the patterns of results and the differences between different types of students, particularly that difference between Key Stage 3 and Key Stage 2 students, that we were concerned about. The decisions that we made in November were about how we pitched this test so that the results are more comparable with the current SATs, in terms of the level gradings. We made those decisions before these tests were set or before we had the results. For those sections of the media that think that we have changed the rules because of results, they are misunderstanding at two levels: one, they are misunderstanding if they think that we are unhappy with the overall level of results in these pilots; and two, they are misunderstanding the sequence, because they are interpreting that we have made these changes in response to results.

Q382 Lynda Waltho: If we could drill down on this issue, basically what I want you to confirm is whether the passing rate was lower than you expected it to be. I think that that is cutting to the chase.

Jim Knight: Again, it is more complicated than that. In some tests, the results were better than expected and in some tests the results were worse than expected. So, it was not about the pass rate; it was about the pattern.

Ralph Tabberer: There were good results and there were some weak results, but the anomalies were sufficient to make us appreciate that there were some things that have changed within the tests. As we had set up these new tests, there was a test effect, but we do not know what that effect is yet and we will not know until we run another pilot. There are some things related to teacher behaviours changing, including which children they are putting in for the tests and at what stage. We do not know how much is down to that factor. There are also some questions about the level that we are pitching the tests at. However, it is impossible from this first pilot to

separate out which effect is pushing in which direction. You must also remember that these are self-selecting schools, so there is no sense in which they are a national representative sample of the performance across the country. So, we are having to find our way through this process quite carefully. We need another round of results before we know what is going on. We will have a chance to try another set of tests and then we will be in a position to make the review of this process available at the end of the year.

Q383 Lynda Waltho: If the problem is a higher rate of failure, it might imply that there is perhaps a discrepancy between the tool of assessment and the teacher judgment about whether a child has reached a particular stage. If that is the case, how could we resolve it?
Ralph Tabberer: First, the only thing we can do is speculate—we are not in a position to know and we cannot answer whether it is down to a change in teacher behaviour. If a child does less well in a test, say, you may deduce that that reflects that the teacher has got the level wrong. However, we do not know whether teachers behave differently or take a different approach to progression tests than they would to an external test at the end of year 6. They might, for example, be pitching for a child to show a level earlier than they normally would in order to take part. However, we will not know enough about how people behave until we review the situation.
Jim Knight: Equally, a Key Stage 2 maths teacher, for example, might be very familiar with Levels 3, 4 and 5, because they deal with those all the time. However, they might be less familiar with Levels 1 or 7, say. Making the assessment in those very early days—in December, they were only two or three months into the pilot—and making the right judgment on whether pupils were ready to take some of the tests, might have been difficult.

Q384 Lynda Waltho: The Government have put a lot into single level testing and have stated that it will be rolled out nationally subject to positive evidence from the pilot study. That shows quite a lot of confidence. Why are you backing single level tests so publicly before we have sufficient evidence from the pilots? I know that you are a confident man.
Jim Knight: Theoretically, the progression pilot, single level testing and testing when ready, accompanied by one-to-one tuition, is compelling. It would be a positive evolution from SATs, for reasons that we have discussed. Naturally, we want such a positive evolution to work. If it does not work, we will not do it.
Lynda Waltho: That was a very confident answer.
Chairman: Douglas.

Q385 Mr. Carswell: I have four questions. The first is general and philosophical. There are lots of examples in society of testing and qualifications being maintained without the oversight of a state agency, such as university degrees, certain medical and legal qualifications, and musical grades. I cannot remember hearing a row about dumbing down Grade 2 piano or an argument about whether the Royal College of Surgeons had lowered a threshold. Therefore, why do we need to have a state agency to oversee testing and assessment in schools? Does the fact that the international baccalaureate has become more popular in certain independent schools suggest that some sort of independent body, which is totally separate from the state and which earns its living by setting rigorous criteria, is necessary?
Jim Knight: Yes, it is necessary, which is why we are setting up an independent regulator that will be completely independent of Government and directly accountable to Parliament.

Q386 Mr. Carswell: But it will not earn its living by producing exams that people want to take—it will be funded by the taxpayer.
Jim Knight: Yes, but the examinations are absolutely crucial to the future of the country and to the future of children in this country—marginally more so, I would argue, than Grade 2 piano—so it is right to have an independent regulator to oversee them in the public interest. However, we should move on from the QCA as it is currently formed, which is to some extent conflicted, because it is both develops and regulates qualifications. Because it undertakes the development of qualifications, it has a vested interest in their success, which is why we thought that it would be sensible to split them. We will legislate in the autumn, but we will set up things in shadow form later this year under current legislation. That means we will have that independence.

Q387 Mr. Carswell: If the QCA earns its fees by setting competitive examinations in competition with other bodies, I have no doubt that it will be setting good tests.
Jim Knight: Would I not then appear before the Committee and be asked about the over-marketisation of the education system? People would say that valuable exams are not being properly regulated or set because there is not enough of a market to make that particular speciality commercially viable.

Q388 Mr. Carswell: Architects and surgeons seem to get on okay.
Jim Knight: Yes, but there will always be a good market for architects and surgeons, but there may not be for some other important skills.

Q389 Mr. Carswell: Without wanting to move away from asking the questions, I wonder whether you would deny the claims of those people who suggest that over the past 15 to 20 years, under Governments of both parties, standards have dropped. I will give you some specific instances. In 1989, one needed 48% to get a C grade in GCSE maths. Some 11 years later, one needed only 18%. That is a fact. Successive Governments and Ministers have claimed that exam results get better every year. However, in the real world, employers and universities offer far more remedial courses to bring school leavers up to standard than they did previously. International

benchmarks show that UK pupils have fallen behind. Does that suggest that, paradoxically, we have created an education system that is drowning in central targets and assessments, but one that lacks rigour? Central control is having the opposite effect to the one intended.

Jim Knight: You will be amazed to hear that I completely disagree with you.

Q390 Mr. Carswell: Which fact do you dispute?

Jim Knight: Ofsted, an independent inspectorate, inspects the education system and gives us positive feedback on standards. We also have the QCA, which is an independent regulator. Although we are strengthening the independence of the regulation side, the QCA still remains relatively independent. It regulates standards and ensures that the equivalence is there. It says categorically that standards in our exams are as good as they have ever been. Then we have National Statistics, which is also independent of the Government. We commissioned a report led by Barry McGaw from the OECD, which is a perfectly respectable international benchmarking organisation, and he gave A-levels a completely clean bill of health. What has changed is that we are moving to a less elitist system. We are trying to drive up more and more people through the system to participate post-16 and then to participate in higher education. Some people rue the loss of elitism in the system and constantly charge it with dumbing down, and I think that that is a shame.

Chairman: You did not answer Douglas's point about the particular O-level in percentage terms.

Mr. Carswell: In 1989, one needed 48% to get grade C GCSE maths. Some 11 years later, one needed 18%. Do you agree or not?

Ralph Tabberer: I do not agree. The problem with the statistics is that you are comparing two tests of very different sorts.

Mr. Carswell: Indeed.

Ralph Tabberer: The tests have a different curriculum, a different lay-out and different groups taking them. We cannot take one percentage and compare it with another and say that they are the same thing. That is why we need to bring in some measure of professional judgment to look at the tests operating different questions at different times. That is why in 1996 we asked the QCA to look at tests over time, and it decided that there were no concerns about the consistency of standards. In 1999, we asked the Rose review to look at the same thing, and it said that the system was very good regarding consistency of standards. In 2003, as you rightly pointed out, we went to international experts to look at the matter. We put those questions to professional judgment, because it is so difficult to look at tests.

Q391 Mr. Carswell: Quangos and technocrats are doing the assessment. The Minister has mentioned three quangos, so technocrats are assessing performance.

Jim Knight: Look at the key stage results. Look at Key Stage 2 English, where the results have gone up from 63% to 80% since 1997. In maths, they have gone up from 62% to 77%. In English at Key Stage 3, they have gone up from 57% in 1997 to 85% in 2007. There is consistent evidence of improvement in standards. It should not be a surprise, when you are doubling the amount of money going into the system and increasing by 150,000 the number of adults working in classrooms, that things should steadily improve. The notion that the improvements are because things are dumbed down is utter nonsense. The international comparators are obviously interesting and very important to us. We started from a woeful state in the mid-90s, and we are now in a much better state, but we are still not world class. We know that we have to do better to become world class, and we said so explicitly in the Children's Plan. We also know that if we do not carry on improving, we will be left behind, because the international comparators also show that more countries are entering the league tables and more are taking education seriously and doing well. Globally, the competition is out there, and we must respond.

Q392 Mr. Carswell: We may not agree on that, but there is one area where I think that we agree, because I agree with what you said earlier about education needing to respond to changing social and economic circumstances. If the centre sets the testing and assessment, it is surely claiming that it knows what is best, or what will be best, and what needs to be judged. If you have central testing, will you not stifle the scope for the education system to be dynamic and to innovate? It is a form of central planning.

Jim Knight: We do not specify things for the end of Level 4 examinations; we talk about equivalency in terms of higher-level GCSEs, so if people want to take other, equivalent examinations, that is fine. The only things where we specify are SATs, which, as we have discussed, are intended to provide a benchmark so that we can measure pupil performance, school performance and national system performance.

Q393 Mr. Carswell: I am anxious about the way in which the SATs scoring system works. I was reading a note earlier about the CVA system, which we touched on earlier. If testing is about giving parents a yardstick that they can use to gauge the sort of education that their child is getting, that is a form of accountability, so it needs to be pretty straightforward. Is there not a case for saying that the SATs scoring system and the CVA assessment overcomplicate things by relativising the score, for want of a better word? They adjust the score by taking into account people's circumstances, and I have read a note stating that the QCA takes into account particular characteristics of a pupil. Is that not rather shocking, because it could create an apartheid system in terms of expectations, depending on your background? Should it not be the same for everyone?

Jim Knight: There is not an individual CVA for each pupil, and I do not know what my child's CVA is—the CVA is aggregated across the school. The measure was brought in because there was concern that the initial value added measure was not sufficiently contextualised, that some were ritually disadvantaged by it and that we needed to bring in a

measure to deal with that. On questions from Annette and others, we have discussed whether it is sufficiently transparent to be intelligible enough. I think that the SATs are pretty straightforward. They involve Levels 1, 2, 3, 4, 5, 6 and 7—where are you at? That is straightforward enough. Obviously, you then have 1a, b and c and the intricacies within that.

Q394 Mr. Carswell: There is all the contextualising and relativising that means that you cannot necessarily compare like with like.
Ralph Tabberer: There is no difference in the tests that different children sit or in the impact of levels. You could perhaps suggest that the CVA is providing an analysis that includes a dimension of context, but that is different. The tests are constant. With the analyses, we are ensuring that—over the years, through consultations, people have encouraged us to do this—a variety of data is available, so people can see how a child or school is doing from different angles.

Q395 Mr. Carswell: A final question: living in this emerging post-bureaucratic, Internet age, is it only a matter of time before a progressive school teams up with a university or employer and decides to do its own thing, perhaps based on what is suitable to its area and the local jobs market, by setting up its own testing curriculum? Should we not be looking to encourage and facilitate that, rather than having this 1950s attitude of "We know best"?
Jim Knight: We are encouraging employers to do their own thing and to become accredited as awarding bodies. We have seen the beginnings of that with the Maccalaureate, the Flybe-level and Network Rail. We have a system of accreditation for qualifications—you might think it bureaucratic, Douglas. We are rationalising it to some extent, but a system of accreditation will remain. After accreditation, there will be a decision for maintained schools on whether we would fund the qualification, and I do not see that as going away. We will publish a qualifications strategy later this year setting out our thinking for the next five years or so. The independent sector might go down that road. All sorts of qualifications pop up every now and then in that area. Any other wisdom, Ralph?
Ralph Tabberer: I go back to the starting point. Before 1988, we had a system whereby schools could choose their own assessments. We introduced national assessment partly because we did not feel that that system gave us consistent quality of education across the system. I do not know of any teacher or head teacher who would argue against the proposition that education in our schools has got a lot better and more consistent since we introduced national assessment. We have put in place the safeguards on those assessments that you would expect the public to look to in order to ensure that they are valid, reliable and consistent over time.
Jim Knight: Obviously, with the Diplomas it is a brave new world that has started with employers and with asking the sector skills councils to begin the process of designing the new qualifications. That has taken place at a national level—it is not a local, bottom-up thing, but a national bottom-up thing.
Chairman: We are in danger of squeezing out the last few questions. I realise that this is a long sitting, but I would like to cover the rest of the territory. Annette.

Q396 Annette Brooke: We heard a view from a university vice-chancellor that it is possible that pupils from independent schools will account for the majority of the new A* grades at A-level. What is your view on that?
Jim Knight: I tried to dig out some statistics on the numbers getting three A* grades, which has been mentioned in the discussion—I am told, based on 2006 figures, that it is just 1.2% of those taking A-levels. To some extent that is on the margins, but we have done some research into whether, judged on current performance, those getting A* grades would be from independent or maintained schools, because we were concerned about that. We believe in the importance of adding stretch for those at the very top end of the ability range at A-level, which is why we brought in the A* grade. However, we were conscious of worries that it would be to the advantage of independent-sector pupils over maintained-sector pupils. The evidence that has come back has shown the situation to be pretty balanced.

Q397 Chairman: Balanced in what sense?
Ralph Tabberer: We have looked at the data, following the vice-chancellor's comment to the Committee that around 70% of those getting three A*s would come from independent schools. From our modelling, we anticipate that something like 1,180 independent school pupils would get three or more A*s from a total of 3,053, so 70% is far from the figure that we are talking about.
Annette Brooke: We have to wait and see.
Jim Knight: Yes.
Annette Brooke: That sounded a very reasonable hypothesis.

Q398 Chairman: The examination boards also said that A*s will make sure that fewer kids from less privileged backgrounds get into the research-rich universities. That is what they said. Although you are not responsible for higher education, Minister, every time you put up another barrier, bright kids from poorer backgrounds are put off from applying. You know that.
Jim Knight: Yes. We had some concerns about that, which is why we examined actual achievement in A-level exams, and we were pleased to see that particular result. The difficulty when we took the decision, just to help the Committee, was that we could see the logic in providing more stretch at the top end for A-level. The issue was not about universities being able to differentiate between bright A-level pupils, because we can give individual marks on modules to admissions tutors; it was genuinely about stretch. Should we prevent pupils, in whatever setting, from having that stretch, just because of that one worry about independent and

maintained-sector pupils? We took a judgment that we had a bigger responsibility than that, which is to stretch people in whatever setting they are in. That is why we took that decision, but we were reassured by the evidence that the situation is not as bleak as Steve Smith, whom we respect hugely as a member of the National Council for Educational Excellence, might have at first thought.

Q399 Annette Brooke: I disagree with you on that point. Many years ago, for my generation, we had S-levels, so you had the opportunity for stretch. Why do we need this so tied in? Maybe sometimes we should put the clock back.
Jim Knight: We had S-levels—I was very pleased with my Grade 1 in geography S-level. We replaced those subsequently with the Advanced Extension Award, in which I was delighted by my daughter's result. However, not many people took them, and they were not widely accepted—they did not seem to be a great success. S-levels were extremely elitist—it was fine for me, in my independent school, to get my Grade 1. I am sure that you did very well in whatever setting you were in.
Annette Brooke: I did not go to an independent school.
Jim Knight: They were introduced in the era of an elite education system. Integrating something into the A-level is a better way forward than introducing something marginal.

Q400 Chairman: Some people on the Committee have always wondered why you cannot just have the raw data—the scores that you get in your subjects—and leave it at that. Universities can judge from that; they do not need the A*.
Jim Knight: We can give that data to admissions tutors. We have said that we will do that, which is fine. The issue is about stretch, as I have said; it is not about differentiation for universities.

Q401 Mr. Carswell: You have used the word "elitist" several times in a disparaging sense. Is not testing and assessment inherently elitist, because it differentiates and breaks down people's results in a hierarchy of performance?
Jim Knight: Not necessarily, because a driving test is not that elitist.

Q402 Mr. Carswell: Between those who pass and those who fail, of course it is. I failed my driving test the first few times I took it—it was elitist, and for people who drove, thank goodness that it was.
Jim Knight: I have discussed an education system that was designed to separate people in an elitist way. We had a lot more selection, including Certificates of Secondary Education, General Certificates of Education and the rest. A few went to university, and the rest went into unskilled or skilled occupations, of which—this is no longer the case—there were plenty.[2] We cannot afford to have a level of elitism that is culturally built in. Yes, we need to differentiate, but not in a way that makes judgments.

[2] *Note by Witness:* This relates to the unskilled occupations, not the skilled.

Mr. Carswell: By definition, it is going to make judgments.

Q403 Lynda Waltho: I want to wave a flag for the poor old Diploma at this point.
Jim Knight: There is nothing poor or old about the Diploma.

Q404 Lynda Waltho: No, I do not think that it is poor, and I want to make sure that it is not ignored. The OCR has stated that in its experience new qualifications take at least 10 years to become accepted and to take root—I am not sure whether you will be pleased to hear that. It has also indicated that in seeking parity with GCSE and GCE, the main parts of Diplomas are increasingly adopting models and grading structures that mirror GCSE and GCE. What assurances can you give us, Minister, that the proposed format of the new Diploma, which I am positive about, will be given time to work and will not be subject to undue interference?
Jim Knight: As you know, the first teaching will start this September, with the entitlement in 2013 to all 14 of them—we will decide when the entitlement to the last three is introduced. That is a fair lead-in time—it is not the full OCR 10 years, but it is fair. The fundamental design of the Diplomas will not change. We are not going to move away from generic learning, which involves things such as functional skills, personal learning and thinking skills. The voice of the sector skills councils, where we started the process, will obviously be heard very loud in terms of what learning is required for someone to do well in a sector. On additional specialist learning, which is the area of greatest flexibility, there may well be some things that feel very familiar in terms of GCSE, A-level, BTEC and the other qualifications that are part of the landscape at the moment. The additional specialist learning element and the work experience element may well look familiar. We have said that no individual school or college will be able to live with those individual Diplomas on their own. We will have much stronger employer engagement and a style of teaching and learning that is related to the workplace. In assessment terms, some of that may be similar to applied GCSEs, applied A-levels and some of the BTECs—it will be a very distinctive offer—but we will always look for equivalence. We should not shy away from equivalence with GCSEs and A-levels. At the advanced level, we were pleased about the Diploma being worth the equivalent of three and a half A-levels.

Q405 Lynda Waltho: Just one more point. The supplementary memorandum from the Department states that the Government will consider the future of A-levels and GCSEs "in the light of the evidence". It is clear that both of those qualifications are here to stay. Has the Department's position on the long-term future of GCSEs and A-levels changed?
Jim Knight: No. The announcement that we made in October, when we announced the three additional Diplomas, was that we would postpone the A-level review until 2013. In the meantime, we will see what

the evidence is in terms of the experience of learners, parents, schools, colleges, universities and employers around the qualification landscape. We will continue to invest and reform A-levels and GCSEs in the meantime. We will not let them wither on the vine—far from it. Obviously, we will be putting a lot of energy into making Diplomas a success, but not at the expense of GCSEs, A-levels and, indeed, apprenticeships, which we will be expanding as well. We want to be able to assess them all to see whether they are strong offers for all stakeholders. We will have a review in 2013, and we have no preconceptions about how that will turn out.

Q406 Mr. Chaytor: Minister, may I ask about the splitting up of the QCA? Will the positions of chief executive at the two new agencies be publicly advertised?
Jim Knight: We would have to advertise under the normal Nolan rules.
Ralph Tabberer: With the new regulator, the position would be open, but with the existing QCA in its new form, my understanding is that Ken Boston will be staying on as the chief executive until the end of his term.

Q407 Mr. Chaytor: How will you ensure that the development agency and the regulator are more independent of the Department than the QCA has been in the past? Earlier, you described the QCA as the "independent regulator", and then qualified that by saying, "Well, relatively independent." What will be put in place that will make it different?
Jim Knight: The regulator will be a non-ministerial department, like Ofsted, accountable to Parliament rather than through Ministers. The new chief regulator, the chair of the organisation, will be a Crown appointment in the same way as Christine Gilbert at Ofsted. In that respect, it will clearly be more independent than the QCA as a non-departmental public body that is accountable through us and through our budget lines, subject to a remit letter. The position of the development body will be very similar to the QCA's current position: it will not be any closer; it will not be any further away; and it will still perform the development role that the QCA currently performs.

Q408 Chairman: Would you like the Committee to help you with appointments and screening the eligible candidates?
Jim Knight: Goodness me, it is very difficult for me to refuse any offers of help from the Committee, but I would have to see the form back in the Department.
Ralph Tabberer: We would have to consult civil service commissioners, who guide us on this process.
Jim Knight: There you go; that is the official advice.

Q409 Mr. Chaytor: So the new development agency will essentially be the QCA reinvented? It will not be a non-departmental public body?
Jim Knight: It will be—
Ralph Tabberer: It will remain—
Jim Knight: Yes.
Ralph Tabberer: It will remain a non-departmental public body accountable to our Ministers.

Q410 Mr. Chaytor: Not accountable to Parliament?
Ralph Tabberer: Well, it will be, but through our Ministers, whereas the independent regulator will not have a route through our Ministers. That is the distinction.

Q411 Mr. Chaytor: What will the route be for the independent regulator?
Jim Knight: As Ofsted is now, so not through Ministers but directly to Parliament.
Chairman: Through this Committee.

Q412 Mr. Chaytor: Had the development agency been in existence 12 months ago, would a policy such as the introduction of single level tests have been taken up only after advice from the development agency? Had that been the case, would it have been likely to have been introduced over a longer period? The point that I am trying to get at is this. There was an urgency about the introduction of the single level test. Your letter to the Committee mentions that one of the problems was the lack of pre-testing. Is this not almost like the introduction of curriculum 2000, when the lack of pre-testing was one of the reasons for the difficulties?
Ralph Tabberer: The pilot for the single level test has been agreed and worked on with the QCA from the beginning. The very big difference between curriculum 2000 and the single level test is the fact that it is a pilot. We are making sure that it acts as a pilot. We test it out so that we understand what is going on. We are doing that with the QCA. There is nothing in the timing of any of these decisions that will have changed the way that we approach that or the advice that we would have given.

Q413 Mr. Chaytor: May I ask about the Primary Review? Could you say a word about the time scale for that, and will it be under the aegis of the new Development and Training Agency?
Jim Knight: Jim Rose is leading that for us. Most of the support that he gets logistically, in secretariat and personnel terms, is from the QCA. They are doing that together. Jim is working from within the Department, but with the QCA, in that review. We expect some interim advice from him in a few months, and we are looking for him to conclude by the end of the year.
Chairman: Fiona has a very quick question on creativity. She has a pressing problem.

Q414 Fiona Mactaggart: I am sorry about that; I said to the Chairman that I would have to leave and that I would not be able to ask you about your welcome announcement on the cultural offer and on ways to assess creativity. Clearly, no one really knows how that needs to be done, and I wonder whether you would share with the Committee your

early thoughts on how it is to be done—and continue to share with the Committee what you are going to do?
Jim Knight: To continue to share—
Chairman: You had better make it quick, Minister, or Fiona will miss her deadline.
Jim Knight: It is obviously something that we would be happy to do. We announced last week that we would pilot how this works. We know that there are a number of schools where this already happens. We know that there is a considerable amount of culture already within the curriculum in terms of art, music, dance and drama. We know that a number of schools already do trips to access all of that, so what we need to pilot is how to extend that and use the very positive experience of greater partnerships, informed by the Committee's excellent report, and integrate it with the sport five hours, some of which, as with culture, would be out of school.

Q415 Fiona Mactaggart: What about assessing it?
Jim Knight: In terms of assessing how well it is working and the results, I turn to Ralph.
Ralph Tabberer: We would be very happy to work with the Committee on the approach we take. It is new ground for us, and it is important that we get it right. However, we are not in a position where we have a working model up and running, so given your interest we would enjoy looking at it with you.
Chairman: Honour is satisfied; thank you, Fiona. David, I am sorry about that.

Q416 Mr. Chaytor: I return to the question of the independence of the new agency. What interests me is where it is specified, where the separate functions are to be described, and how the Government are going to guarantee that independence and ensure that, at the end of the day, Ministers' influence will not determine the recommendations and the advice of the agency.
Jim Knight: We will need to legislate to achieve some of this. We can set up some interim arrangements in shadow form, effectively trying to put a Chinese wall through the QCA. Ultimately, much of it will therefore be set out in legislation.
Ralph Tabberer: I would add that the document *Confidence in Standards* is out for consultation and it sets out the detailed proposals for the bodies. That will allow you and us an opportunity to look at the careful balances that need to be achieved in order to get the regulator and the agency right. The particular opportunity that it gives us is to make what were robust internal procedures within the QCA the subject of work between the two, so one takes responsibility for development and the other holds them to account. It will now be a much more transparent process, which should allow the public to have more confidence in it.

Q417 Mr. Chaytor: The new regulator will have responsibility for regulating the quality and reliability of qualifications and assessments, but in terms of the policy surrounding assessment, will he be required to give advice on that or will it come under the aegis of the agency? The specific issue that I put to you is perhaps the question with which we started our discussions this afternoon, that of the multiple uses of assessment. Who will be advising the Government about whether it makes sense to use pupil assessment as a means of holding schools accountable and as a means of informing parental preference? Will that be the role of the regulator or the development agency, or will that stay within the Department?
Jim Knight: In simple terms, Ministers and the Department will still decide on policy. The new development agency will develop that policy into practice, as appropriate and according to its remit, and the regulator will assess whether or not standards have been maintained as a result. It will be quite important to respect the independence of the regulator. If we started to involve the regulator in the policy development too much, we would be back to where we are with the current QCA, in terms of the confliction that I talked about, if confliction is a word, between its development role and its regulatory role. Therefore, we would be cautious about that. However, we would certainly be asking the new development agency for advice on policy development and then deciding what to do with that advice.

Q418 Mr. Chaytor: If the new development agency gave Ministers advice that it no longer made sense for pupil assessment to serve so many different purposes, would you be likely to accept that advice?
Jim Knight: Given that the QCA currently performs that development role, the QCA could give us that advice and we would then decide accordingly. We have not received that advice. I have been fairly robust in my position, so I think that you can predict how I might respond if I received such advice.

Q419 Mr. Chaytor: Will the legislation be specific in giving the development agency responsibility for advising the Government on the uses of assessment and not just the specific forms of assessment?
Jim Knight: My familiarity with *Confidence in Standards* is not sufficient to be able to answer that question, but I can obviously let you know if Ralph cannot.
Ralph Tabberer: In terms of taking advice, the closest relationship will be between the QCA's development agency and the Department in putting together advice for the Ministers. I think that it would be fair to say that we will also seek the views of the regulator on the moves and their response will be public. It would not make sense just to act and have the regulator completely out of the process. However, we have to guard this process in such a way, as Jim indicated, that we do not cause conflict for the regulator in so doing. Again, I would point to the current consultation as an opportunity for us to tease out these issues. They are very good questions.

Q420 Mr. Chaytor: If the advice from the regulator conflicts with the advice from the development agency, who mediates? Will the legislation include some kind of mechanism for that mediation?

Ralph Tabberer: Ultimately, the decision in this area falls to Ministers. The advice of the regulator would be public and that advice would be subject to normal parliamentary scrutiny.
Chairman: David had to leave to attend a Standing Committee; he did not mean any discourtesy. We will have very quick questions from Douglas and Annette, and then we are finished for the day.

Q421 Mr. Carswell: Am I right, Minister, in thinking that you said that the regulator would be accountable directly to Parliament and not through Ministers?
Jim Knight: Yes.

Q422 Mr. Carswell: But at the same time, you are saying that the head of the regulatory body would be appointed by the Civil Service Commission.
Ralph Tabberer: The Civil Service Commission gives its recommendations on the process to be followed.

Q423 Mr. Carswell: Given that Gordon Brown made the excellent recommendation in his first week as Prime Minister that he was going to reform Crown prerogative and allow the legislature to make key appointments that hitherto have been the preserve of Sir Humphrey, would there not be a case for saying that the Children, Schools and Families Committee should conduct confirmation hearings for the person appointed to that regulatory role? Would that not be an excellent idea? It would ensure genuine accountability to Parliament and it would ensure that Mr. Brown's excellent idea was realised.
Jim Knight: That is a decision above my pay grade.

Q424 Mr. Carswell: But you would support it in principle?
Jim Knight: Even my view is above my pay grade.

Q425 Mr. Carswell: You do not have a view? You do not have a view on Gordon Brown's White Paper about the governance of Britain?
Jim Knight: I think that you understand what I am saying.

Q426 Mr. Carswell: No, I do not understand. You do not have a view? Does Gordon know?
Jim Knight: I do not have a view that I am going to share with you now.
Ralph Tabberer: I was taught in my briefing not to answer rhetorical questions.

Q427 Annette Brooke: As I recollect, when we talked to the examination bodies, they did not really see the need for a development agency, because they could take on its work. They could give you advice for free—they would not need two chief executives. Have you considered that?
Jim Knight: Yes, it has passed through the mind, but not for long. We are making the change because of the perceived conflict of interest in the QCA between its regulation and its development functions. Replacing that conflict of interest with a different conflict of interest, by giving the development of qualifications to the same people who make money out of them, did not seem sensible.

Q428 Annette Brooke: I might not applaud a competitive model, but I thought that you might.
Jim Knight: There is a very competitive qualifications market out there, which responds to the qualifications that are designed in structural and policy form by the QCA. We do not have a problem with competition, but I do not want the conflict that your interesting notion would produce.

Q429 Chairman: One very last thing, Minister. When did you last consult teachers directly on how they feel about the quality of the testing and assessment system? I do not mean through the unions, I mean through either the General Teaching Council or, more importantly, by the direct polling of teachers about their experiences and happiness, and what suggestions they would make to improve the system. When was the last time you did that?
Jim Knight: You said "not through unions", but I frequently talk to teachers' representatives about the matter. I do not know when we last carried out any kind of proper research or survey of the work force on the issue, but we would have to carry out parallel research with parents, because we have a system that is designed for the consumers of the product as well as the producers.

Q430 Chairman: That ducks the question. When was the last time you consulted? If you have not done it recently, why do you not do so now?
Jim Knight: I tried not to duck it by saying that I did not know.

Q431 Chairman: Is it a good idea? Could you do even a sample survey?
Jim Knight: It might be a good idea to do both. Obviously, we carry out various types of research, but we have a budget on which there are many demands. As always, I will listen carefully.

Q432 Chairman: This is a key part of Government policy. Surely the people who deliver the policy up and down the land should be consulted on how they feel about the regulatory and testing and assessment frameworks. Surely it is key to know what they think.
Jim Knight: It is key, which is why I have frequent discussions with teachers' representatives about it.

Q433 Chairman: We all know about vested interests. We set up the General Teaching Council to cut away from that, but you evaded talking about it.
Jim Knight: I meet with the GTC and hold discussions with it—I shall speak at its parliamentary reception fairly soon, and I look forward to seeing members of the Committee there.

Q434 Chairman: Ralph, you were in charge of teacher training. Do you think that it is important to keep in touch? As we know, one of the key elements in delivering quality education is a highly motivated

and highly trained work force, so should it be consulted on the very important issue of testing and assessment?

Ralph Tabberer: I am sure that it should be consulted—your argument is very clear. We are not short of teachers' views on the matter, whether they are presented directly by representatives, at school visits, or through the GTC. We are shorter on responses from parents, and we might take a look at the range of their views. It is enormously important to look at the public information on what is happening in the school system. I still believe—you have heard me defend it today—that our system is particularly transparent because of what we can show parents regarding what works and what does not work in our schools and in the system at large. We should not give that up too quickly.

Q435 Chairman: It is not that anyone on the Committee particularly disagrees with that, Ralph, but I am calling on someone to find those things out scientifically. I am not simply talking about chatting to the National Union of Teachers or the NASUWT. Their representatives failed to appear before the Committee to talk about this important subject, which is an indication of how important they consider testing and assessment. You should find out scientifically, by going directly to teachers and evaluating their opinion. We would be happy if you spoke to parents at the same time, but you should get a clear resonance of what is going on out there.

Jim Knight: In the same way that I heard the message that we might want to think about research on CVA and what information parents find useful, I hear your message that we could usefully do some proper, quantitative research with the teaching work force on testing and the other things that we have talked about today, and I shall certainly take that back to the Department.

Chairman: Thank you very much. This has been a long session. We have learned a lot, and we have enjoyed it—I hope you did, too.

Further memorandum submitted by Jim Knight MP, Minister of State for Schools and Learners

Following my appearance before the Committee on 18 February, I am writing to provide further details relating to the evidence I gave. I hope the Committee will find this additional information useful in making its report.

TEACHERS' AND PARENTS' ATTITUDES TO ASSESSMENT

The Committee expressed concern that they were not aware of any recent research into teachers' and parents' attitudes to assessment.

We have recently consulted widely on the possible new direction of assessment policy. Teachers were amongst those consulted about our *Making Good Progress* proposals. Responses to the consultation indicated general support for the idea of taking a test when ready, although there was a lack of understanding at that stage about how the single level tests would operate.

We also have an opportunity to learn more about teachers', pupils' and parents' views through the *Making Good Progress* pilot. As part of their independent evaluation of that, PriceWaterhouseCoopers are conducting surveys with all these groups, as well as interviews with teachers and pupil focus groups in a small sample of the pilot schools.

We are also considering asking for parents views on our proposals for future assessment policy, including the use of Single Level Tests, as part of our forthcoming consultations on the Children's Plan.

Furthermore, as part of a series of ongoing research, Ipsos MORI was commissioned by the QCA to carry out a sixth wave of research to assess teachers', parents', students' and the general public's perceptions of the A level and GCSE exam system shortly after the 2007 examinations session. This follows five "waves" of a quantitative survey of perceptions towards A levels between 2003 and 2006 (and, since 2004, perceptions towards GCSEs). The report, *GCSEs and A level: the experiences of teachers, students, parents and the general public. A Research Study Conducted for the Qualifications and Curriculum Authority (QCA)* was published in last month. Some of the key findings were as follows:

- Teachers are most likely to agree that most A level students get the grades their performance deserves, followed by the students themselves. 63% of parents agree with this statement.
- Teachers are less likely to express concerns about the A level system this year than previously—over a third (35%) claim they have no concerns.
- Teachers' confidence in the A level system continues to increase (76% now agree that they have confidence in the system).
- Just under three in five (59%) A level teachers say they are confident in the accuracy and quality of the marking of A level papers.
- Teachers' confidence in the GCSE system has increased since 2006 (76% agree they have confidence, up from 66% in 2006).

— Two-thirds of GCSE teachers (67%) have confidence in the accuracy and quality of marking of GCSE papers.

— The same proportion of teachers as the previous two years (77%) agree that the exam system needs to be independently regulated and controlled by a body other than the Department for Children, Schools and Families. In line with last year, three-quarters (76%) of teachers feel that the QCA is effective in this role.

— Parents' concerns are similar to those of teachers, with problems about how exams are graded and marked coming out top, followed by how time consuming the work is for students, and the number of exams that students face. Concerns relating to coursework appear to be less of an issue this year (possibly due to the removal of coursework from some GCSE subjects).

Performance Tables and how useful they are to Parents

I have considered further the points that members made about contextual value added (CVA), and its transparency to parents. The prime purpose of the Tables is public accountability—one element of which is choice of school. This links to the presentation of information for use by parents, and it should be considered within the wider context, for example, by looking at parental reactions to notions of progression, as I suggested in my answer to Annette Brooke (Q355). I will also be looking at we already know about parents' views on school accountability, and will then consider whether more research is needed, or whether we should indeed address questions of the presentation of CVA or other aspects of school data.

Appointments to the New Independent Regulatory Body for Qualifications

The Committee expressed interest in the scrutiny of appointments to the new independent regulatory body for qualifications and examinations. Subject to Parliament passing the necessary legislation, the Chair of the Office of the Qualifications and Examinations Regulator will be a Crown appointment. Prior to legislation, an interim regulator will be established as a committee of the QCA, to carry out the regulatory functions of the QCA as they are currently set out in legislation. The Chair of the interim regulator is being recruited under Office of the Commissioner for Public Appointments rules, and will become the first Chair of the Office of the Qualifications and Examinations Regulator.

The establishing legislation will reflect official advice on the appropriate processes to be followed for public appointments to the Office of the Qualifications and Examinations Regulator. On 23 January 2008, the Minister for the Cabinet Office sent to the Liaison Committee a list of existing public appointments which the Government propose should be subject to the pre-appointment scrutiny by their relevant Select Committee, and placed a copy in the Library of the House.

Ed Balls wrote to you on 17 December, when the consultation document *Confidence in Standards* was published, and invited the Committee to consider how best it could monitor and review the work of the new statutory regulator and the interim regulator.

I hope this information helps the Committee in its Inquiry and I look forward to seeing the report.

March 2008

Written evidence

Memorandum submitted by Association of Colleges (AoC)

Introduction

AoC (the Association of Colleges) is the representative body for colleges of further education, including general FE colleges, sixth form colleges and specialist colleges in England, Wales (through our association with fforum) and Northern Ireland (through our association with ANIC). AoC was established in 1996 by the colleges themselves to provide a voice for further education at national and regional levels. Some 98% of the 415 plus general FE colleges, sixth form colleges and specialist colleges in the three countries are in membership. These colleges are the largest providers of post-16 general and vocational education and training in the UK. They serve over four million of the six million learners participating in post-statutory education and training, offering lifelong learning opportunities for school leavers and adults over a vast range of academic and vocational qualifications. Levels of study range from the basic skills needed to remedy disadvantage, through to professional qualifications and higher education degrees.

The key role played by the sector and its 250,000 staff in raising the level of skills and competitiveness of the nation's workforce make colleges central to the Government's national and regional agenda for economic prosperity and social inclusion. AoC services to member college corporations include information, professional development and support in all aspects of institutional management, governance, curriculum development, quality, employment, business development and funding. AoC also works in close partnership with the Government and all other key national and regional agencies to assist policy development, continuously to improve quality and to secure the best possible provision for post-16 education and training.

A summary of the key points included in this submission are:

1. The assessment framework should mirror the move towards a modular approach to the curriculum and therefore allow for the assessment of bite-sized chunks of learning which builds on learners' existing achievements.
2. A flexible approach should be taken to forms of assessment to ensure that they can be accessed by all learners: this is particularly important at a time when there is a greater emphasis on summative assessments.
3. This learner centred approach will be facilitated by ensuring that lecturers have a role in devising appropriate assessment tools.
4. Whatever form of assessment is used it is key that the burden and/or nature of assessment does not become more important than the learning taking place.
5. As the platform of awarding bodies widens there should be differentiation in the rigour of the regulatory framework and monitoring procedure based upon the quality and maturity of the provider.
6. The impact of Key Skills tests on raising attainment levels is open to question. Consideration should be given to broadening the range of such tests if they are to continue to allow for a broader picture of young peoples' skills and abilities.
7. League tables do not provide a true reflection of the work of the FE sector and the emphasis on them is in danger of shifting the focus within the school sector to league results rather than individual need.
8. There is no recognition given in league tables to the 48% of young people who do achieve Level 2: the inclusion of Level 1 results would provide a much needed incentive for schools to invest resources and expertise in this significant group of learners.
9. The current league table reporting mechanisms based on individual institutions will not meet the needs of the 14–19 reforms where partnerships and shared ownership of students will be the key.
10. The current testing system at 16 is, in our view, a disincentive to progression. We advocate the development of an experiential learning route post 14 with an appropriate assessment and testing framework.
11. The potential negative impact upon teaching and learning and on the cohesive view of subjects created by modular assessments could be mitigated by the wider use of such initiatives as the Level 3 extension project.
12. There can be no justification for universities imposing another hurdle for young people to overcome through the use of entry tests. If the rationale for these tests is dissatisfaction with the standards of A-level then this is the problem which should be addressed.

Testing and Assessment

AoC is pleased to provide evidence to the Committee on the key issue of testing and assessment: an issue which has been the subject of much recent debate.

1. General Issues

1.1 AoC would want to reiterate its consistently stated view of the key principles which should, in our view, underpin any system of testing and assessment. These are that assessments should:

- be able to be accessed by all learners regardless of their learning styles
- build on learners' existing achievements
- allow for the assessment of bite- sized chunks of learning
- allow for formative charting of progress

1.2 The assessment framework should mirror the move towards a modular approach to the curriculum which responds effectively to the needs of both learners and employers. This incremental approach whereby learners would be able to build up a bank of achievements would not be impeded by the current restrictions of age and level but would allow the learner to develop skills and knowledge against agreed standards in a way which would maximise individual potential and maintain motivation.

1.3 AoC believes that the Qualifications and Credit Framework currently being piloted is an example of best practice. Whilst welcoming the commitment to use QCF principles in the design of the Foundation Learning Tier AoC believes that if the decision had been taken to encompass all qualifications within this framework it would have resulted in a genuinely flexible assessment and testing system which would have allowed all learners to achieve their full potential.

2. Testing and Assessment

2.1 *Why do we have a centrally run system of testing and assessment?*

2.1.1 The current central system has facilitated the development of a clear and transparent system of qualifications with structured levels and pathways. This provides all stakeholders with confidence in the qualification and assessment system.

2.2 *What other systems of assessment are in place both internationally and across the UK?*

2.2.1 Internationally there are many examples of central testing and accreditation. The European Credit Transfer System (ECTS) is a credit based system for HE which has been successfully tested across Europe.

2.3 *Does a focus on national testing and assessment reduce the scope for creativity in the curriculum?*

2.3.1 Whilst a national framework does not of necessity reduce the scope for creativity, the degree of restriction within that framework can significantly limit the possibility of delivering curriculum in a differentiated and imaginative way. Testing reduces creativity if it is inappropriate: for example a written test of mechanical skills or musical understanding diverts the curriculum towards that skill and away from mechanics or music. The system of testing should allow individuals to be tested in ways that are appropriate for them: it should therefore provide for a range of assessment methods. These assessments should be able to credit steps of attainment so that individuals gain a sense of achievement and motivation and their progress can be effectively charted. National assessment should not drive the learning. Colleges tell us that young people are now more interested in the assessment criteria than they are in the learning material. If this is the case, then too great an emphasis is being placed on the test and insufficient on the learning experience.

2.4 *Who is the QCA accountable to and is this accountability effective?*

2.4.1 QCA as a regulating body accountable to the DfES plays an essential role in ensuring the integrity and credibility of the examination system. This regulatory function has the dual stands of establishing the regulatory framework and of monitoring its implementation.

2.4.2 AoC believes that the time may now be appropriated to consider building in greater flexibility and differentiation into QCA's monitoring processes so that a "lighter touch" is adopted with those colleges with a strong track record of operating testing and assessments within a rigorous internal quality framework. This would reflect the commitment of moving towards greater self regulation within the college system and recognise the maturity of many colleges.

2.4.3 The impact of technology on the testing and assessment mechanisms is a further driver to reviewing the necessity of maintaining the monitoring system in its current form. With the growing introduction of on line testing Awarding Bodies are now able to identify and address any examples of aberrant marking immediately. This should allow awarding bodies to accurately assess risk and differentiate between centres. Our aim here is to reduce the spiralling cost of examinations and the bureaucracy associated with them.

2.4.4 The Leitch report advocates a more responsive approach whereby employers are able to devise and accredit qualifications. AoC recognises that greater flexibility in the qualification offer is key to the achievement of this demand led approach, however this wider platform of awarding bodies will in the medium term require a strong regulatory framework and robust monitoring procedures.This, therefore supports the proposal for a differentiated monitoring role played dependent upon the nature or quality of the provider.

2.5 *What role should exam boards have in testing and assessment?*

2.5.1 There is a self evident role for exam boards in producing examining tools and ensuring the standardisation of tests and assessments. AoC does, however, question the current trend towards greater reliance on more traditional, paper- based summative tests. We understand that the rationale behind this move is to ensure rigour and to avoid plagiarism but find it regrettable as it will inevitably penalise those learners who are unable to evidence their true ability/knowledge through this form of assessment. The focus on paper- based testing may not always be appropriate or, indeed, valid for what is being tested and therefore may not serve either employers' or selectors' needs. AoC has strongly supported the emphasis on meeting individual learner needs and welcomes the current focus on providing a framework for personalising both the school and FE sector. There is a real danger that this personalisation will be undermined by the use of a single form of assessment.

2.5.2 AoC has consistently lobbied for lecturers to be given more responsibility in devising assessment tools which are appropriate for their learners. This would not only ensure a closer fit to learner need but also, by placing greater responsibility and trust in lecturers, would professionalize the workforce. A more confident workforce taking a more active role in the development of testing/assessment methods would in turn have a positive impact on teaching and learning. However, we recognise that this would need to be supported both by training and rigorous quality controls especially for new providers.

3. National Key Stage Tests

3.1 *How effective are current Key Stage tests?*

3.1.1 As the recipients of students who have undertaken Key Stage tests, colleges are concerned about the degree of remedial work they have to undertake to ensure that students' skills are at the necessary level; this must create some doubts about how far the tests have helped to raise attainment levels. This leads us to question the use that is currently being made of the outcomes of the Key Stage tests.

3.1.2 It would seem sensible that if young people are going to be tested regularly that the tests should allow for the assessment of a broader picture of their skills and abilities to be developed, for example what are their preferred learning styles or assessment styles? These outcomes could be used to inform the teaching and learning strategies used and could be passed on to, for example, the Careers staff to allow for more targeted advice and guidance to be given. This would be in keeping with the development of the personalisation agenda and would also provide the profile necessary to support the progression of young people to the most appropriate mix of programme within the Specialised Diplomas.

3.2 *Are league tables based on test results an accurate reflection of how well schools are performing?*

3.2.1 The league tables currently lead to schools placing all their efforts on ensuring that as many of their pupils as possible obtain at least 5 GCSEs at Grade C or above. It is probable that the outcomes of league table results, rather than individual need, take precedence; that drilling young people on the skills needed to pass tests rather that enhancing their ability to learn and apply that learning has moulded classroom practice. The consequence of this is that there is a focus of additional time and resources on those borderline Grade C/D pupils: many of whom subsequently succeed. However, whilst these young people have the necessary grades to progress to level 3 courses they often do not have the necessary concepts or intellectual rigour to deal with the demands of the next level course. This raising of false expectations resulting in a sense of inadequacy may well account for the high drop out rate at 17.

3.2.2 Both the current league tables and the measurements proposed in the new proposals on "Progression Measures" only report at Level 2: there is therefore no recognition given to pupil achievement at Level 1. Given that in 2005–06 41.5% of pupils left school without having achieved five GCSEs at Grade C or above (and that this group of young people form a significant pool of untapped talent who, because their needs are not addressed, frequently account for a great deal of disruptive behaviour and/or fall into

the NEET group), AoC strongly believes that Level 1 achievement should be included in league tables. This would serve to highlight the need for high quality provision at this essential stage and provide incentives to schools to focus resources and expertise on this level of work.

3.2.3 The development of the Foundation Learning Tier would, in our view, provide an excellent framework for that significant tranche of young people who are not ready to complete a full Level 2 qualification at the age of 16.

4. Testing and Assessment at 16 and after

4.1 Although there is no specific request for comment on Performance Measures in FE, AoC considers that there are issues linked to this which should be drawn to the Committee's attention.

4.2 AoC has consistently objected to the basis of league tables in that, of their nature, they favour those schools that are in a position to recruit more able pupils. We recognise that the development of value-added measures mitigate against this bias to a certain extent but the focus on academic examination results continues to undervalue the work done by the FE sector which has traditionally recruited young people with lower qualifications and/or those who prefer the vocational to the academic route.

4.3 The implementation of the Government's 14–19 reform policy raises another fundamental issue about league tables which AoC believes should be addressed. The current tables measure performance in individual institutions: however, the 14–19 reforms are based upon partnerships both across schools and across the school/FE sectors so that frequently students will be "shared" by two or more institutions. The current league table reporting mechanism will not promote the partnership working which is at the heart of the new 14–19 curriculum and consequently will not encourage the development of provision which meets the needs of young learners. AoC recommends that thought to be given to disbanding the current individual performance base in favour of developing a form of local target setting against which the local performance of all partners can be measured.

4.4 *Is the testing and assessment in "summative" tests fit for purpose?*

4.4.1 As indicated above AoC believes that any testing regime should allow for all learners to demonstrate their skills and knowledge. Summative testing has a role in any testing and assessment framework in that it provides the vehicle to judge the learners' overall understanding and knowledge. However, the nature of summative testing at 16 is not in our view "fit for purpose".

4.4.2 AoC has always argued that the current form of Level 2 testing at 16 only serves to create a barrier in that it does not give many young people the opportunity to demonstrate what they are good at. This artificial hurdle at the age of 16 is, in our view, a disincentive to progression and a significant contributory factor to drop out both pre and post 16.

4.4.3 In order to address this issue we would advocate the development of an experiential learning route post 14 with an appropriate assessment and testing framework. We continue to advocate an incremental learning model in which young people can mix levels of learning in different subjects according to their ability. The potential this would create of achieving at a higher level in areas of strength and interest would be a powerful motivating influence and would increase the number of young people who continue learning beyond the age of 16.

4.4.4 AoC welcomes the introduction of the extended project at Level 3 as it will provide a more enabling way for young people to demonstrate their skills and attributes than the traditional testing methods. We believe that the introduction of such innovations should be extended to develop a richer and more differentiated range of 14–19 testing and assessment tools.

4.5 *Are the changes to GCSE coursework due to come into effect in 2008 reasonable? What alternative forms of assessment might be used?*

4.5.1 We consider that GCSE provision falls within the schools' remit. However, on the principle of the changes being introduced, we would stress that colleges' whole ethos is to develop the individual's learning skills and ability to take responsibility for and ownership of research for projects: this runs counter to the planned changes to the GSE testing methodology. AoC recognises the concerns over plagiarism which lies behind these changes but would draw the Committee's attention to the alternative approach being taken to this issue by the Joint Council for Qualifications who, in a pilot project with colleges, are encouraging an approach that places the responsibility on the young person to utilise a software package to check for plagiarism. This is, in our view, a more appropriate manner of dealing with the concerns of intellectual property and improper use of the Internet as a research tool. It anticipates and prepares young people to recognise and acknowledge their sources and to develop referencing skills that will be required at higher levels of study.

4.6 *What are the benefits of exams and coursework? How should they work together? What should the balance be between then?*

4.6.1 AoC has always asserted that, ideally, assessment should form a natural part of learning. One of the key attributes of good teachers is the ability to embed the ongoing checking of learning into their teaching practice thus being able to chart individual progress and provide additional or alternative learning opportunities as appropriate; a model used naturally in Work Based Learning. We recognise that this is the ideal and appreciate the Government's desire to retain rigour within assessment processes; however, we believe that currently the balance is swinging too far towards summative assessment and that consideration should now be given to redressing this balance.

4.7 *Will the ways in which the new 14–19 diplomas are to be assessed impact on other qualifications such as GCSE?*

4.7.1 The 14–19 diplomas will be assessed using a range of internal and external assessment methods; detail of these have not yet been finalised. AoC would wish the same principles of a rigorous framework which gives flexibility for teachers to devise assessment tools suitable to their learners' needs to be adopted. We would want the opportunity to comment upon the detail of the Diploma assessment and testing methodology as the detail is developed.

4.8 *Is holding formative summative tests at ages 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?*

4.8.1 There is some concern in the sector that this pattern of summative tests can result in too much time being dedicated to preparation for examinations rather than giving students a real understanding of their subjects. Summative tests at 17 appear to be the main cause for concern and a review of the nature of assessment at this stage may be appropriate.

4.9 *To what extent is frequent, modular assessment altering both the scope of teaching and the style of learning?*

4.9.1 Modular assessment has facilitated the introduction of bite-sized learning so that students are able to focus on a specific component of their course and gain accreditation for it prior to moving on to another component, thus increasing motivation. There is, however, a down side to this modular approach in that if overused it can lead to a fragmentation of the curriculum whereby students do not glean an overview of the subject or of natural links within a subject. The introduction of such initiatives as the extended project at Level 3 will serve to reduce this risk of fragmentation and we would advocate the wider use of this model.

4.10 *How does the national assessment system interact with university entrance? What does it mean for a national system of testing and assessment that universities are setting entrance tests as individual institutions?*

4.10.1 AoC sees an intrinsic illogicality in the fact that there is a structured qualifications system which does not allow for automatic progression to Higher Education. We are disappointed that universities remain elitist and unaccountable bodies who are able to use additional entrance tests to further their own selection agendas. If the rationale behind these entrance tests is that universities are not satisfied with the standards of A levels then the perceived problem ie A level should be tackled rather than imposing yet another hurdle for young people to overcome. The elitism in the Higher Education system is exacerbated by the use of tests for which the privileged will be coached whilst others will be denied access. This cannot be fair nor is it in line with Government policy to increase access to degree level qualifications.

May 2007

Joint memorandum submitted by Paul Black, Emeritus Professor of Education King's College London, John Gardner, Professor of Education Queen's University Belfast, and Dylan Wiliam, Professor and Deputy Director, Institute of Education, London

Introduction

This submission highlights the limited reliability of national tests and testing systems and argues that the misuse of such assessments would be greatly reduced if test developing agencies and awarding bodies were required to inform the public fully on reliability matters. It is to be stressed that the limited reliability is systemic and inevitable, and does not imply any lack of competence or professionalism on the part of the test developers and awarding bodies involved.

The basis for our argument is that results of assessments, whether by tests or other forms of assessment, are imperfect, ie they are subject to measurement error. The argument may be summarised by the following six points:

1. That some degree of error in such results is inescapable.
2. That the best information currently available indicates that up to 30% of candidates in any public examination in the UK will receive the wrong level or grade.
3. That sound information on the errors in examination results is not available from any UK test developers or awarding bodies for any public examination results, and we know of no plan or intention to obtain or publish such information.
4. That fully informed decisions using examination results cannot be made unless those making the decisions are aware of the chances of error in them.
5. That there are also important issues concerning the validity of assessment results, which interact with the issues of reliability, and would broaden the scope, but not affect the nature, of our arguments.
6. That policy makers, pupils, parents and the public need to have information about such errors, to understand the reasons for them, and to consider their implications for the decisions which are based upon, or are affected by, assessment results.

These six points are explained below in detail.

1. *Error is Inescapable*

Errors in test results may arise from many sources. We are not concerned here with errors in administration and in marking which we assume to be minimal. The problem we address arises because any test measure will be in error because it is bound to be based on a limited sample of any candidate's attainment.

At the end of (say) five years of study for GCSE in a subject, a candidate could attempt to answer many questions, or other tasks, different from one another in form, in difficulty, and in the particular topics and skills which they explore: if the candidate could attempt all of these, and could do each on more than one occasion to guard against day to day variations in performance, the average score from all of these attempts would represent her/his "true score". In practice this is impossible.

So with any test that is affordable, in terms of its costs of time and money, the score is based on a limited sample of the candidate's possible attainments, a sample that is limited in the types of demand, in the topics and skills explored, and in the test occasions. This limitation does not arise from any lack of expertise or care from those doing either the setting, or the marking, or the administration.

Thus the measured score differs from the "true score" because it is only a limited sample of the possible evidence. The errors involved are random errors, in that some candidates' scores may be higher than their true score, others lower, some may bear a very small error, some a very large one. For any particular candidate, there is no way of knowing what this "sampling" error is, because the true score cannot be known.

2. *How Big is the Sampling Error?*

If a candidate's marks varied hardly at all from one question to another, we can be confident that increasing the length of the test would make very little difference: the sampling error is small. If the marks varied wildly from one question to another, we can be confident that the sampling error is large—a far longer test is needed. By analysing the range of such variations from one question to another, over many candidates, it is possible to make a statistical estimate of the errors involved. If the candidates are tested on more than one occasion, then the differences in their scores between the two occasions gives further information about the likelihood of error. We submit, with this document, a copy of a chapter from a recent book which reviews publications in which the results of this type of analysis are reported. The main conclusions may be summarised as follows:

(i) In such public examinations as Key Stage tests, 11+ selection tests GCSE and A-level, the sampling error is such that between 20% and 30% of candidates will earn a result which will differ by at least one Level or Grade from their "true" score.

(ii) There is fairly close agreement, that the errors can be this large, between analyses of standardised tests used in California, the 11+ selection tests used in Northern Ireland, and tests at Key Stages 2 and 3 in England.

(iii) That a common measure of so-called "reliability" of a test can be used to indicate the error in the raw marks, but that the effect of such error on wrong assignments of levels or grades will depend on how close, in terms of marks, the boundaries between these levels or grades happen to be.

(iv) That one way in which this sampling error can be reduced is to narrow the range of the question types, topics and skills involved in a test—but then the result will give misleading information in that it will give users a very limited estimate of the candidates' attainments.

(v) That another way is to increase the testing time so that the sample can be larger; unfortunately, the reliability of testing only increases very slowly with the test length eg as far as can be estimated, to reduce the proportion of pupils wrongly classified in a Key Stage 2 test to within 10% would require 30 hours of testing.

(vi) That there are no other ways of composing short tests that could reduce these errors. It would be possible in principle to produce more reliable results if the information that teachers have about their pupils could be assembled and used, because teachers can collect evidence of performance on many different occasions, and in several tasks which can cover a range in content and in types of demand.

(vii) That these findings are concerned with the errors in the results of individual pupils. The mean of a group of observed scores will be closer to the mean true score of the group simply because many of the sources of random variation in the performance of individuals will average out.

3. *Is enough known about error in our Public Examinations?*

The short answer to this question is "no". One of us (PJB) wrote to the chief executive of the QCA, in January 2005, enquiring whether there were any well researched results concerning the reliability of those tests for which the QCA had responsibility. The reply was that "there is little research into this aspect of the examining process", and drew attention only to the use of borderline reviews and to the reviews arising from the appeals system. We cannot see how these procedures can be of defensible scope if the range of the probable error is not known, and the evidence suggests that if it were known the volume of reviews needed would be insupportable.

Of the information quoted in section 2 above, one source, that for the 11+ selection tests in Northern Ireland, is based on a full analysis of data from test papers. The estimates for Key Stage tests and for A-level are based either on measures provided by others or on well established data from comparable examination systems.

4. *Do we need Measures of the error in our Examinations?*

The answer to this question ought not to be in doubt. It is profoundly unsatisfactory that public examination data in this country do not conform to the requirements which have been set out as Standards for Educational and Psychological Tests in a joint publication from the American Educational Research Association, the American Psychological Association, and the American National Council on Measurement in Education. In the USA, tests that do not conform to these requirements, are, to all intents and purposes, indefensible (at least in law).

However, there is a more immediate reason why the answer to this question must be "yes": the point is that decisions are made on the basis of these measures, and such decisions will be ill-judged if those making them assume that the measures are without error.

Examples of decisions which might be made differently if the probability of error were taken into account are as follows:

(i) The general argument is that where people know results are unreliable, they may seek alternative sources of evidence for confirmation. Where they regard the test as flawless, they are more likely to rely entirely on them.

(ii) Public policy at present is based on a premise that tests results are reliable, teachers' assessments are not, so the best combination will give little or no weight to teachers' own assessments. Policy ought to be based on using the optimum combination of the two. Of course, to do this we would need serious development work and well researched data, to enhance and measure the reliability of teachers' own assessments. Some development work of this sort has been done in the past, and some is being developed now: it deserves greater priority because use of teachers' own assessments is the only approach available to overcome the serious limitations to accuracy which we present in section 2. We draw attention to the policy of the Australian State of Queensland, where for over 25 years state test certificates have been based entirely on teachers' own assessments, albeit within a rigorous system of inter-school collaboration to ensure authenticity and comparability of these assessments.

(iii) The recent proposals, in the paper entitled *Making Good Progress*, to introduce single level tests available every six months, are a good example of the limitations that follow from ignoring the errors in test measures. The proposals pay no attention to the effects of measurement error. For example, since the proposals allow pupils to make several repeated attempts, and given the random errors in each attempt, any pupil with a true score which is only a few marks below a given level is bound to succeed eventually, so in time standards will appear to have risen even if the "true" scores have not risen at all.

(iv) The majority of class teachers and school managements seem to be unaware of the limitations of test results. Our experience of working with schools is that many, either from lack of trust in their own judgments, or because parents pay more attention to test paper scores rather than to the

teachers' own knowledge of the pupils, will rely heavily, even exclusively, on the results of short formal tests, often using "off-the-shelf" test papers taken from previous public examinations. The results of such tests are then used for setting and streaming in later years, and/or to advise pupils to drop some subjects. It is very likely that for a significant proportion of pupils, such decisions and advice may be ill-advised and even unjust.

(v) Both pupils themselves, their parents, and those using test results for selection or recruitment, ought to have information about the probability of error in these results.

5. *Validity and Reliability*

We have, in this paper, limited our arguments almost entirely to the issue of reliability. It is possible to have a very reliable tests result which is based only on performance on (say) tests of memory under formal examination conditions. Since few work place situations require only such ability, the results do not tell employers what they need to know—the results may be reliable, but they are invalid. The two requirements, of validity and of reliability, are inter-related. One example of this interaction was pointed out above: narrowing the range of topics and skills addressed in a test can improve reliability, but make the test less valid as it is a more narrow measure. Another example is that use of pupils' performance on extended projects, involving library research, or practical investigations in science, can enhance validity, and might also improve reliability in basing the measure on a wider range of activities extending over a longer time period. Whilst consideration of validity will broaden the range of the arguments, the recommendations in 6 below would still apply, but would have to be expanded in scope.

6. *So what should be done?*

Those responsible for our public examinations should:

(i) review all available evidence about their reliability;

(ii) set up a continuing process of research studies to enhance, and keep up to date, evidence of the best possible quality;

(iii) set up discussions to determine the optimum policies for obtaining assessment data in the light of evidence about errors in the various sources and methods available; such discussions including fresh consideration of the potential for far greater use of assessments made by teachers, individually and by and between schools; and

(iv) develop a programme to ensure that all who use or depend on assessment results are well informed about the inevitable errors in these results.

Concluding Comment

We stress that the above is not an argument against the use of formal tests. It is an argument that they should be used with understanding of their limitations, an understanding which would both inform their appropriate role in an overall policy for assessment, and which would ensure that those using the results may do so with well-informed judgement.

The issue that we address here cuts across and affects consideration of almost all of the issues in which the Committee has expressed interest. In particular it is relevant to.

— the accountability of the QCA;

— whether Key Stage tests adequately reflect the performance of children and schools;

— the role of teachers in assessment;

— whether and how the system of national tests should be changed;

— whether testing and assessment in "summative" tests (for example, GCSE, AS, A2) is fit for purpose; and

— the appropriateness of changes proposed in GCSE coursework.

May 2007

Memorandum submitted by Cambridge Assessment

Cambridge Assessment is Europe's largest assessment agency and plays a leading role in researching, developing and delivering assessment across the globe. It is a department of the University of Cambridge and a not-for-profit organisation with a turnover of around £175 million. The Group employs around 1,400 people and contracts some 15,000 examiners each year.

Cambridge Assessment's portfolio of activities includes world-leading research, ground-breaking new developments and career enhancement for assessment professionals. Public examinations and tests are delivered around the globe through our three highly respected examining bodies.

The assessment providers in the Group include:

University of Cambridge English for Speakers of Other Languages (Cambridge ESOL)

Tests and qualifications from Cambridge ESOL are taken by over 1.75 million people, in 135 countries. Cambridge ESOL's Teaching Awards provide a route into the English Language Teaching profession for new teachers and first class career development opportunities for experienced teachers. Cambridge ESOL works with a number of governments in the field of language and immigration.

University of Cambridge International Examinations (CIE)

CIE is the world's largest provider of international qualifications for 14–19 year-olds. CIE qualifications are available in over 150 countries. CIE works directly with a number of governments to provide qualifications, training and system renewal.

OCR

The Oxford Cambridge and RSA awarding body (OCR) provides general academic qualifications and vocational qualifications for learners of all ages through 13,000 schools, colleges and other institutions. It is one of the three main awarding bodies for school qualifications in England.

ARD

The Assessment Research and Development division (ARD) supports development and evaluation work across the Cambridge Assessment group and administers a range of admissions tests for entry to Higher Education. The ARD includes the Psychometrics Centre, a provider and developer of psychometric tests.

A View of the Scope of the Enquiry

1. At Cambridge Assessment we recognize that it is vital not to approach assessment on a piecemeal basis. The education system is exactly that: a system. Later experiences of learners are conditioned by earlier ones; different elements of the system may be experienced by learners as contrasting and contradictory; discontinuities between elements in the system (eg transition from primary to secondary education) may be very challenging to learners.

2. Whilst understanding the system as a system is important, we believe that the current focus on 14–19 developments (particularly the development of the Diplomas and post-Tomlinson strategy) can all too readily take attention away from the serious problems which are present in 5–14 national assessment.

3. Our evidence tends to focus on assessment issues. This is central to our organisation's functions and expertise. However, we are most anxious to ensure that assessment serves key functions in terms of supporting effective learning (formative functions) and progression (summative functions). Both should be supported by effective assessment.

4. We welcome the framing of the Committee's terms of reference for this Inquiry, which make it clear that it intends to treat these two areas as substantially discrete. Cambridge Assessment's qualifications deliverers (exam boards), OCR and University of Cambridge International Examinations, have tendered evidence separately to this submission. They have looked chiefly at 14–19 qualifications.

5. This particular submission falls into two sections: Firstly Cambridge Assessment's views on the national assessment framework (for children aged 5–14). These are informed by, but not necessarily limited to, the work which we carried out through out 2006 in partnership with the Institute for Public Policy Research (IPPR) and the substantial expertise in the Group of those who have worked on national tests.

6. The second section is on University Entrance Tests. Cambridge Assessment has been involved in the development of these for nearly a decade and uses a research base that stretches back even further. At first their scope was limited to Cambridge University but over the last four years it has grown to include many other institutions. That they are administered under Cambridge Assessment's auspices (as opposed to those of one of our exam boards) is a reflection of their roots within our research faculty and the non statutory nature of the tests themselves.

Section 1

1. National Assessment Arrangements

7. In this section we have sought to outline the problems that have built up around the national assessment arrangements. We have then gone on to discuss the changes proposed in our work with the IPPR. We also then discuss the problems that appear to be inherent in the 'Making Progress' model that the Government is committed to trialling. Our conclusion is that there is a window of opportunity before us at

the present time, just one of the reasons that the Committee's Inquiry is so timely, and that the Government should not close it with the dangerous haste that it seems bent on. There are a range of options and to pursue only one is a serious mistake.

8. We have included two Annexes:

— an overview of the evidence on national assessment dealing with questions ranging from "teaching to the test" to "measurement error"; and

— a brief discussion of why the sometimes mooted return of the APU might not deliver all the objectives desired of it.

2. Diagnosis of the Challenge—Critique and Revision of National Assessment Arrangements

9. It is important to note that Cambridge Assessment is highly supportive of the principle of a National Curriculum and related national assessment. The concept of "entitlement" at the heart of the National Curriculum has been vital to raising achievement overall; raising the attainment of specific groups (eg females in respect of maths and science); and ensuring breadth and balance. We recognise that enormous effort has been put in, by officials and developers, to improving the tests in successive years. We support the general sentiment of the Rose Review—that the system has some strong characteristics—but it is clear that deep structural problems have built up over time.

10. Whilst being concerned over these problems, Cambridge Assessment is committed to the key functions supported by national assessment: provision of information for formative and diagnostic purposes to pupils, teachers and parents; information on national standards, and accountability at school level. We return to these key functions in more detail below. However, Cambridge Assessment is critical of the way in which national assessment has been progressively and successively elaborated into a system which appears to be yielding too many serious and systemic problems.

Accumulating problems in National Assessment—a vessel full to bursting point?

11. There are two particularly significant problems in the highly sensitive area of technical development of national assessment arrangements. Firstly, previous statements by agencies, departments and Government have exaggerated the technical rigour of national assessment. Thus any attempts to more accurately describe its technical character run the risk of undermining both the departments and ministers; " . . . if you're saying this now, how it is that you said that, two years ago . . .". This prevents rational debate of problems and scientifically-founded development of arrangements.
Secondly, as each critique has become public, the tendency is to breathe a sigh of relief as the press storm abates; each report is literally or metaphorically placed in a locked cupboard and forgotten.

12. In contrast, we have attempted here to take all relevant evidence and integrate it; synthesising it in such a way that underlying problems and tendencies can accurately be appraised—with the intention of ensuring effective evaluation and refinement of systems.

14. Put simply, if a minister asks a sensible question: " . . . are attainment standards in English going up or down and by how much? . . ." there is no means of delivering a valid and sound response to that question using current arrangements. This is a serious problem for policy formation and system management. It is not a position which obtains in systems which use independent light sampling methods such as the US NAEP (National Assessment of Educational Progress).

Functions

15. Current national curriculum assessment arrangements within England have attracted increasing criticism in respect of the extent to which they are carrying too many purposes (Brooks R & Tough S; Bell J et al; Daugherty R et al). Since 1988 a substantial set of overt and tacit functions have found themselves added. The original purposes specified in the TGAT Report (Task Group on Assessment and Testing) comprised:

1. formative (diagnostic for pupils; diagnostic for teachers);
2. summative (feedback for pupils and parents);
3. evaluative (providing information at LEA and school level); and
4. informative (providing information on educational standards at system level).

16. The following have been added, as increasingly elaborated uses of the flow of detailed data from national assessment:

— school accountability;

— departmental accountability;

— apportionment of funds;

— inspection patterns and actions;

— upwards pressure on standards/target setting;

— structuring of educational markets and school choice;

— emphasis of specific curriculum elements and approaches;

— detailed tracking of individual attainment, strengths and weaknesses; and

— quantification of progress.

17. Unsurprisingly, many educationalists have expressed the view that the current tests carry too many functions and that the underlying management processes are too elaborated. To carry this broad range of functions, the system of assessing every child at the end of each Key Stage is dependent on maintaining test standards over time in a way which is in fact not practical.

18. If you want to measure change, don't change the measure. But the nation does—and should—change/update the National Curriculum regularly. Whenever there is change (and sometimes radical overhaul) the maintenance of test standards becomes a particularly aggressive problem. It does, of course, remain a constant problem in areas such as English Literature when one could be pretesting a test on Macbeth which will be taken in 2008 but the pupils are currently studying As You Like it when they sit the pretest. There are remedies to some of the problems this creates—namely switch to different sampling processes; announcing radical recalibration, or switch to low stakes sampling of children's performance, using a NAEP or a modernized APU-style model (Assessment of Performance Unit—see Annexe 2).

19. Attempting to use national assessment to measure trends over time has produced some of the most intense tensions amongst the set of functions now attached to national testing. Stability in the instruments is one of the strongest recommendations emerging from projects designed to monitor standards over time. Running counter to this, QCA and the DfES have—in line with commitments to high quality educational provision, the standards agenda and responses from review and evaluation processes—sought to optimize the National Curriculum by successive revision of content, increasing the "accessibility of tests", and ensuring tight linkage of the tests to specific curriculum content.

20. These are laudable aims—and the emphasis on the diagnostic function of the data from tests has been increasing in recent innovations in testing arrangements However, pursuit of these aims has led to repeated revision rather than stability in the tests. The Massey Report suggested that if maintenance of standards over time remained a key operational aim, then stability in the test content was imperative (Massey A et al). In the face of these tensions, a light sampling survey method would enable de-coupling of national assessment from a requirement to deliver robust information on national educational standards. This would enable testing to reflect curriculum change with precision, to optimize the learning-focussed functions of testing, and enable constant innovation in the form of tests to optimize accessibility.

21. It is therefore clear that the current functions of national testing arrangements are in acute and chronic tension. Using the pragmatic argument that "every policy should have a policy instrument" we conclude that national arrangements should indeed support school accountability and improvement, report to parents and monitor national standards but that a change of arrangements is required to achieve this. A range of approaches are necessary to deliver these functions and we outline some viable options below.

3. Alternative Approaches to National Assessment (KS1, KS2, KS3)

Objectives

22. There is a need to conceptualise a number of possible models for consideration in an attempt to address the problems of "multipurpose testing". It is vital to note that we present here three alternatives. We do this to show that there are credible alternatives for delivering on the key objectives of national assessment—it is simply not the case that there is only one way of moving forward.

23. We believe the aims should be to:

— reduce the assessment burden on schools;

— provide formative assessment for teaching and learning;

— provide information for school accountability; and

— provide information on national standards.

24. In order to secure widespread support within the education community (including parents) a firm re-statement of educational purpose (values) and a commitment to high degrees of validity is essential. It is not enough to initiate changes merely because of concerns about the defects of existing arrangements. We do not here outline values and validity in detail, but recognise that this is a vital precondition of designing revised arrangements, putting them in place, and monitoring their operation. It is important that a full discussion of these matters precedes any executive decision regarding revised arrangements.

Alternative models for national assessment

Model 1: Validity in monitoring plus accountability to school level

25. The aim of this approach is to collect data using a national monitoring survey and to use this data for monitoring standards over time as well as for moderation of teacher assessment. This would enable school performance to be measured for accountability purposes and would involve a special kind of criterion referencing known as domain referencing.

26. Question banks would be created based on the curriculum with each measure focusing on a defined domain. A sample of questions would be taken from the bank and divided into lots of small testlets (smaller than the current KS tests). These would then be randomly allocated to each candidate in a school. Every question is therefore attempted by thousands of candidates so the summary statistics are very accurate and there are summary statistics on a large sample of questions. This means that for a particular year it is known, for example, that on average candidates can obtain 50% of the marks in domain Y.

27. The following year it might be found that they obtain 55% of the marks in that domain. This therefore measures the change and no judgement about relative year on year test difficulty is required. Neither is there a need for a complex statistical model for analysing the data, although modelling would be required to calculate the standard errors of the statistics reported. However, with the correct design they would be superfluous because they would be negligible. It would be possible to use a preliminary survey to link domains to existing levels and the issue of changing items over time could be solved by chaining and making comparisons based on common items between years. Although each testlet would be an unreliable measure in itself, it would be possible to assign levels to marks using a statistical method once an overall analysis had been carried out. The average of the testlet scores would be a good measure of a school's performance given that there are sufficient candidates in the school. The appropriate number of candidates would need to be investigated.

28. The survey data could also be used to moderate teacher assessment by asking the teacher to rank order the candidates and to assign a level to each of them. Teacher assessment levels would then be compared with testlet levels and the differences calculated. It would not be expected that the differences should be zero, but rather that the need for moderation should be determined by whether the differences cancel out or not. Work would need to be done to establish the levels of tolerance and the rules for applying this process would need to be agreed. The school could have the option of accepting the statistical moderation or going through a more formal moderation process.

29. There would be a number of potential advantages related to this model. Validity would be increased as there would be greater curriculum coverage. The data would be more appropriate for the investigation of standards over time. The test development process would be less expensive as items could be re-used through an item bank, including past items from national curriculum tests. There would also be fewer problems with security related to "whole tests". No awarding meetings would be needed as the outcomes would be automatic and not judgemental. Since candidates would not be able to prepare for a specific paper the negative wash-back and narrowing of the curriculum would be eliminated (ie the potential elimination of "teaching to the test"). There would also be less pressure on the individual student since the tests would be low stakes.

30. Given that there are enough students in a school, the differences in question difficulty and pupil question interaction would average out to zero leaving only the mean of the pupil effects. From the data it would be possible to generate a range of reports eg equipercentiles and domain profiles. Reporting of domain profiles would address an issue raised by Tymms (2004) that "the official results deal with whole areas of the curriculum but the data suggests that standards have changed differently in different sub-areas".

31. Work would need to be done to overcome a number of potential disadvantages of the model. Transparency and perception would be important and stakeholders would need to be able to understand the model sufficiently to have confidence in the outcomes. This would be a particularly sensitive issue as students could be expected to take tests that prove to be too difficult or too easy for them. Some stratification of the tests according to difficulty and ability would alleviate this problem. There is an assumption that teachers can rank order students (Lamming D) and this would need to be explored. Applying the model to English would need further thought in order to accommodate the variations in task type and skills assessed that arise in that subject area.

32. Eventually the model would offer the possibility of reducing the assessment burden but the burden would be comparatively greater for the primary phase. Although security problems could be alleviated by using item banking, the impact of item re-use would need to be considered. Having items in the public domain would be a novel situation for almost any other important test in the UK (except the driving test).

33. Discussion and research would be needed in a number of areas:

— values and validity;

— scale and scope eg number and age of candidates, regularity and timing of tests;

— formal development of the statistics model;

— simulation of data (based on APU science data initially);

— stratification of tests / students; and

— pilots and trials of any proposed system.

Model 2: Validity in monitoring plus a switch to "school-improvement inspection"

34. Whilst the processes for equating standards over time have been enhanced since the production of the Massey Report, there remain significant issues relating to:

— teacher confidence in test outcomes;

— evidence of negative wash-back into learning approaches;

— over-interpretation of data at pupil group level; inferences of improvement or deterioration of performance not being robust due to small group size;

— ambiguity in policy regarding borderlining;

— no provision to implement Massey recommendations regarding keeping tests stable for five years and then "recalibrating" national standards; and

— publishing error figures for national tests.

35. In the face of these problems, it is attractive to adopt a low-stakes, matrix-based, light sampling survey of schools and pupils in order to offer intelligence to Government on underlying educational standards. With a matrix model underpinning the sampling frame, far wider coverage of the curriculum can be offered than with current national testing arrangements.

36. However, if used as a replacement for national testing of every child at the end of KS1, 2 and 3, then key functions of the existing system would not be delivered:

— data reporting, to parents, progress for every child at the end of each key stage; and

— school accountability measures.

37. In a system with a light sampling model for monitoring national standards, the first of these functions could be delivered through (i) moderated teacher assessment, combined with (ii) internal testing, or tests provided by external agencies and/or grouped schools arrangements. The DfES prototype work on assessment for learning could form national guidelines for (i) the overall purpose and framework for school assessment, and (ii) model processes. This framework of assessment policy would be central to the inspection framework used in school inspection.

38. The intention would be to give sensitive feedback to learners and parents, with the prime function of highlighting to parents how best to support their child's learning. Moderated teacher assessment has been proven to facilitate staff development and effective pedagogic practice. Arrangements could operate on a local or regional level, allowing transfer of practice from school to school.

39. The second of these functions could be delivered through a change in the Ofsted inspection model. A new framework would be required since the current framework is heavily dependent on national test data, with all the attendant problems of the error in the data and instability of standards over time. Inspection could operate through a new balance of regional/area inspection services and national inspection—inspection teams operating on a regional/area basis could be designated as "school improvement teams". To avoid competition between national and regional inspection, national inspections would be joint activities led by the national inspection service.

40. These revised arrangements would lead to increased frequency of inspection (including short-notice inspection) for individual schools and increased emphasis on advice and support to schools in respect of development and curriculum innovation. Inspection would continue to focus on creating high expectations, meeting learner needs, and ensuring progression and development.

Model 3: Adaptive, on-demand testing using IT- based tests

41. In 2002, Bennett outlined a new world of adaptive, on-demand tests which could be delivered through machines. He suggests that "the incorporation of technology into assessment is inevitable because, as technology becomes intertwined with what and how students learn, the means we use to document achievement must keep pace". Bennett (2001) identifies a challenge, "to figure out how to design and deliver embedded assessment that provides instructional support and that globally summarises learning accomplishment". He is optimistic that "as we move assessment closer to instruction, we should eventually be able to adapt to the interests of the learner and to the particular strengths and weaknesses evident at any particular juncture . . .". This is aligned to the commitments of Government to encourage rates of progression based on individual attainment and pace of learning rather than age-related testing.

42. In the Government's five year strategy for education and children's services (DfES, 2004) principles for reform included "personalisation and choice as well as flexibility and independence". The White Paper on 14–19 Education and Skills (2005) stated, "Our intention is to create an education system tailored to the needs of the individual pupil, in which young people are stretched to achieve, are more able to take qualifications as soon as they are ready, rather than at fixed times . . ." and "to provide a tailored programme for each young person and intensive personal guidance and support". These intentions are equally important in the context of national testing systems.

43. The process relies on item-banking, combining items in individual test sessions to feed to students a set of questions appropriate to their stage of learning and to their individual level of attainment. Frequent low-stakes assessments could allow coverage of the curriculum over a school year. Partial repetition in tests, whilst they are "homing in" on an appropriate testing level, would be useful as a means of checking the extent to which pupils have really mastered and retained knowledge and understanding.

44. Pupils would be awarded a level at the end of each key stage based on performance on groups of questions to which a level has been assigned. More advantageously, levels could be awarded in the middle of the key stage as in the revised Welsh national assessment arrangements.

45. Since tests are individualised, adaptivity helps with security, with manageability, and with reducing the "stakes", moving away from large groups of students taking a test on a single occasion. Cloned items further help security. This is where an item on a topic can include different number values on a set of variables, allowing the same basic question to be systematically changed on different test administrations, thus preventing memorisation of responses. A simple example of cloning is where a calculation using ratio can use a 3:2 ratio in one item version and 5:3 ratio in another. The calibration of the bank would be crucial with item parameters carefully set and research to ensure that cloning does not lead to significant variations in item difficulty.

46. Reporting on national standards for policy purposes could be delivered through periodic reporting of groups of cognate items. As pupils nationally take the tests and when a critical nationally representative sample on a test is reached, this would be lodged as the national report of standards in a given area. This would involve grouping key items in the bank eg on understanding 2D representation of 3D objects and accumulating pupils' performance data on an annual basis (or more or less frequently, as deemed appropriate) and reporting on the basis of key elements of maths, English etc.

47. This "cognate grouping" approach would tend to reduce the stakes of national assessment, thus gauging more accurately underlying national standards of attainment. This would alleviate the problem identified by Tymms (2004) that "the test data are used in a very high-stakes fashion and the pressure created makes it hard to interpret that data. Teaching test technique must surely have contributed to some of the rise, as must teaching to the test".

48. Data could be linked to other cognate groupings, eg those who are good at X are also good at Y and poor on Z. Also, performance could be linked across subjects.

49. There are issues of reductivism in this model as there could be a danger to validity and curriculum coverage as a result of moving to test forms which are "bankable", work on-screen and are machine-markable. Using the Cambridge taxonomy of assessment items is one means of monitoring intended and unintended drift. It is certainly not the case that these testing technologies can only utilise the most simple multiple-choice (mc) items. MC items are used as part of high-level professional assessment eg in the medical and finance arenas, where well-designed items can be used for assessing how learners integrate knowledge to solve complex problems.

50. However, it is certainly true that, at the current stage of development, this type of approach to delivering assessment cannot handle the full range of items which are currently used in national testing and national qualifications. The limitation on the range of item types means that this form of testing is best used as a component in a national assessment model, and not the sole vehicle for all functions in the system.

51. School accountability could be delivered through this system using either (i) a school accumulation model, where the school automatically accumulates performance data from the adaptive tests in a school data record which is submitted automatically when the sample level reaches an appropriate level in each or all key subject areas, or (ii) the school improvement model outlined in model 2 above.

52. There are significant problems of capacity and readiness in schools, as evidenced through the problems being encountered by the KS3 ICT test project which has successively failed to meet take-up targets. It remains to be seen whether these can be swiftly overcome or are structural problems eg schools adopting very different IT network solutions and arranging IT in inflexible ways. However, it is very important to note that current arrangements remain based on "test sessions" of large groups of pupils, rather than true on-demand, adaptive tests. These arrangements could relieve greatly the pressures on infrastructure in schools, since sessions would be arranged for individuals or small groups on a "when ready" basis.

53. There are technical issues of validity and comparability to be considered. The facility of a test is more than the sum of the facility on the individual items which make up each test. However, this is an area of intense technical development in the assessment community, with new understanding and theorisations of assessment emerging rapidly.

54. There are issues of pedagogy. Can schools and teachers actually manage a process where children progress at different rates based on on-demand testing? How do learners and teachers judge when a child is ready? Will the model lead to higher expectations for all students, or self-fulfilling patterns of poor performance amongst some student groups? These—and many more important questions—indicate that the assessment model should be tied to appropriate learning and management strategies, and is thus not neutral technology, independent of learning.

Overall

55. Each of the models addresses the difficulties of multipurpose testing. However, each model also presents challenges to be considered and overcome. The Statistics Commission (2005) commented that "there is no real alternative at present to using statutory tests for setting targets for aggregate standards". The task is to find such an alternative. The real challenge is to provide school accountability data without contaminating the process of gathering data on national standards and individual student performance. All three models have their advantages and could lead to increased validity and reliability in national assessment arrangements and—crucially—the flow of reliable information on underlying educational standards; something which is seriously compromised in current arrangements.

4. New Progress Tests—Serious Technical Problems

56. As a possible line of development for new arrangements, the DfES recently has announced pilots of new test arrangements, to be trialled in 10 authorities. Cambridge Assessment has reviewed the proposals and, along with many others in the assessment community, consider that the design is seriously flawed. The deficiencies are significant enough to compromise the new model's capacity to deliver on the key functions of national assessment; ie information on attainment standards at system level; feedback to parents, pupils and teacher; and provision of school accountability.

57. Cambridge Assessment's response to the DfES consultation document on the progress tests covered the subject in some detail and we reproduce it below for the Select Committee:

i We welcome the developing debate on the function and utility of national assessment arrangements. We applaud the focus on development of arrangements which best support the wide range of learning and assessment needs amongst those in compulsory schooling.

ii As specialists in assessment, we have focused our comments on the technical issues associated with the proposals on testing. However, it is vital to note that Cambridge Assessment considers fitness for purpose and a beneficial linkage between learning and assessment to be at the heart of sound assessment practice.

iii We consider effective piloting, with adequate ethical safeguards for participants, to be essential to design and implementation of high quality assessment arrangements. It is essential that evaluation method, time-frames, and steering and reporting arrangements all enable the outcomes of piloting to be fed into operational systems. There is inadequate detail in the document to determine whether appropriate arrangements are in place.

iv We remain concerned over conflicting public statements regarding the possible status of the new tests (TES 30 March), which make it very unclear as to whether existing testing arrangements will co-exist alongside new arrangements, or whether one will be replaced by the other. This level of confusion is not helpful.

v We see three functions as being essential to national assessment arrangements:

Intelligence on national standards—for the policy process.

Information on individual pupil performance—for the learner, for parents, for teachers.

Data on school performance—for accountability arrangements.

We do not feel that the new model will meet these as effectively as other possible models. We would welcome discussions on alternatives.

vi We believe that, by themselves, the new test arrangements will not provide robust information on underlying standards in the education system. With entry to single-level tests dependent on teachers' decisions, teachers in different institutions and at different times are likely to deploy different approaches to entry. This is likely to be very volatile, and effects are unlikely to always cancel out. This is likely to contaminate the national data in a very new ways, compared with existing testing arrangements. There are no obvious remedies to this problem within the proposed arrangements, either in the form of guidance or regulation.

vii Teachers are likely to come under peculiar pressures, from institutions wishing to optimise performance-table position, from parents of individual children etc. This is an entirely different scenario to the "all to be tested and then a level emerges" character of current arrangements. Tiering invokes a similar, though not as here all-pervasive, effect.

viii Although advanced as "on-demand" testing, the regime is not an "on-demand" regime, and it is misleading to promote it as such. It provides one extra test session per year.

ix The frequency of testing is likely to increase the extent to which testing dominates teaching time. This is not a problem where the majority of washback effects from testing are demonstrably beneficial; we believe that other features of the tests mean that washback effects are likely to be detrimental. It is not clear what kind of differentiation in teaching will flow back from the tests. Ofsted and other research shows differentiation to be one of the least developed areas of teaching practices. We are concerned that the "grade D" problem (neglect of this those not capable of getting a C and those who will certainly gain a C) will emerge in a very complex form in the new arrangements.

x The tests may become MORE high stakes for learners. Labelling such as " . . . you're doing Level 2 for the third time! . . ." may emerge and be very pernicious. Jean Rudduck's work shows such labelling to be endemic and problematic.

xi We are unclear regarding the impact on those learners who fail a test by a small margin—they will wait six months to be re-tested. Do teachers judge that they should "lose six months of teaching time" to get them up to the required level or just carry on with no special support. If special support is given, what is the child not doing which they previously would have done? This is a key issue with groups such as less able boys—they will need to take time out of things which they are good at and which can bolster their "learning identities". Those who are a "near miss" will need to know—the document does not make clear whether learners will just "get a level back"; will get a mark; or an item-performance breakdown.

xii Testing arrangements are likely to become much more complex—security issues, mistakes (such as wrong test for a child) etc are likely to gain in significance.

xiii The length of the tests may be an improvement over existing tests, but further investigative work must be done to establish whether this is indeed the case. 45-minute tests may, or may sample more from each subject domain at an appropriate level, compared with existing national tests. This is an empirical question which needs to be examined. Lower sampling would reduce the reliability of the tests. Compounding this, the issue of pass marks must be addressed—compensation within the tests raises not only reliability questions but also washback effects into formative assessment. People who pass may still need to address key areas of learning in a key stage, if compensation and pass marks combine disadvantageously. The length of the tests and the need to cover the domain will tend to drive tests to a limited set of item types, raising validity issues. This in turn affects standards maintenance—if items are clustered around a text, if the text is changed (remembering test frequency is increased 100%) then all the items are no longer usable. This represents a dramatic escalation of burden in test development. Constructing and maintaining the bank of items will be very demanding.

xiv If a high pass mark is set (and the facility of items tuned to this) there will be little evidence of what a child cannot do. Optimising the formative feedback element—including feedback for high attainers—in the face of demand for high domain coverage, reasonable facility, and accessibility (recognisable stimulus material etc) will be very demanding for test designers. Level-setting procedures are not clear. The regime requires a very fast turnaround in results—not least to set in place and deliver learning for a "re-take" in the next test session (as well as keeping up with the pace of progression through the National Curriculum content). This implies objective tests. However, some difficult factors then combine. The entry will be a volatile mix of takers and re-takers.

xv While calibration data will exist for the items, random error will increase due to the volatility of entry, feeding into problems in the reliability of the item data in the bank. Put crudely, with no awarding processes (as present in existing national tests) there will be a loss of control over the overall test data—and thus reliability and standards over time will become increasingly problematic. As one possible solution, we recommend the development of parallel tests rather than successively different tests. Pre-tests and anchor tests become absolutely vital—and the purpose and function of these must be explained clearly to the public and the teaching profession. More information on this can be provided.

xvi Having the same tests for different key stages (as stated by officials) is problematic. There is different content in different stages (see English in particular). QCA has undertaken previous work on "does a Level 4 mean something different in different key stages"—the conclusion was that it did.

xvii The 10-hour training/learning sessions are likely to be narrowly devoted to the tests. This may communicate strong messages in the system regarding the importance of drilling and "surface learning"—exactly the opposite of what the DfES is supporting in other policy documents. Although superficially in line with "personalisation", It may instil dysfunctional learning styles.

xviii We applaud the sensitivity of the analysis emerging from the DfES in respect of the different populations of learners who are failing to attain target levels. We also support the original Standards Unit's commitment to a culture of high expectations, combined with high support. However, this level of sensitivity of analysis is not reflected in the blanket expectation that every child should improve by two levels.

xix We do not support "payment by results" approaches—in almost any form these have successively been found wanting. Undue pressure is exerted on tests and test administration—maladministration issues escalate.

xx In the face of the considerable challenge of developing a system which meets the demanding criteria which we associate with the operation of robust national assessment, we would welcome an opportunity to contribute to further discussions on the shape of enhanced national arrangements.

5. The Way Forward for National Assessment

58. What is needed is a new look at options—and both the technical and political space for manoeuvre. Cambridge Assessment has not only attempted to assemble the evidence but have produced a "3 option" paper which outlines possible approaches to confront the very real problems outlined above. We commend a thoroughgoing review of the evidence. Not a—"single person review" like "Dearing" or "Tomlinson", but a more managed appraisal of options and a sober analysis of the benefits and deficits of alternatives. For this, we believe that a set of clear criteria should be used to drive the next phase of development:

— technically-robust arrangements should be developed;

— the arrangements should be consistent with stated functions;

— insights from trialling should fed into fully operational arrangements;

— unintended consequences are identified and remedied;

— full support from all levels of the system is secured in respect of revised arrangements;

— a number of models should be explored at the same time, in carefully designed programmes—in other words there should be parallel rather than serial development, trialling and evaluation; and

— appropriate ethical safeguards and experimental protocols should be put in place during development and trialling.

59. It is, of course, vital to consider not only the form of revised arrangements which better deliver the purposes of national assessment but also to consider the methods and time frame for development arrangements, as well as the means of securing genuine societal and system support.

60. The last two elements listed above are critical to this: currently, there are no plans for trialling more than one revised model for national testing. However, a cursory glance in the education research field shows that there is a range of contrasting approaches to delivering the key functions of national testing, many of which may well be presented to this Inquiry . . . It therefore would seem important to trial more than one model rather than "put all eggs in one basket" or take forward only modifications of existing arrangements.

61. It is unclear whether adequate safeguards have been put in place to protect learners exposed to revised national assessment arrangements. Cambridge Assessment recommends—in line with the standards being developed by the Government's Social Research Unit—that new protocols should be developed, as a matter of urgency for the trialling of revised arrangements.

National Assessment—Annex 1

An Overview of the Evidence

1. *Measurement error and the problems with overlaying levels onto marks*

This does not refer to human error or mistakes in the administration of tests but to the issue of intrinsic measurement error. Contemporary standards in the US lead to the expectation that error estimates are printed alongside individual results: such as " . . . this person has 3,592 (on a scale going to 5,000 marks) and the error on this test occasion means that their true score lies between 3602 and 3582 . . ..". Is this too difficult for people to handle (ie interpret)? In the current climate of increasing statistical literacy in schools, it should not be. Indeed, results could be presented in many innovative ways which better convey where the "true score" of someone lies.

Error data are not provided for national tests in England, and both the Statistics Commission and commentators (eg Wiliam, Newton, Oates, Tymms) have raised questions as to why this international best practice is not adopted.

Of course, error can be reduced by changing the assessment processes—which most often results in a dramatic increase in costs. Note "reduce" not "remove"—the latter is unfeasible in mass systems. For example, double marking might be adopted and would increase the technical robustness of the assessments. However, this is impractical in respect of timeframes; it is already difficult to maintain existing numbers of markers, etc. Error can be reduced by increased expenditure but is escalating cost appropriate in the current public sector policy climate?

One key point to bear in mind is that one must avoid a situation where the error is significantly less than the performance gains which one is expecting from the system, and from schools—and indeed from teachers within the schools. Unfortunately, 1–2% improvement lies within the bounds of error—get the level thresholds wrong by two marks either way (and see the section on KS3 Science below) and the results of 16,000 pupils (ie just over 2%). could be moved.

Measurement error becomes highly significant when national curriculum levels (or any other grade scale) is overlaid onto the scores. If the error is as above, but a cut score for a crucial level is 48 (out of 120 total available marks) then getting 47 (error range 45–49) would not qualify that person for the higher level, even though the error means that their true score could easily be above the level threshold. In some cases the tests are not long enough to provide information to justify choosing cut-scores between adjacent marks even though the difference between adjacent marks can have a significant effect on the percentages of the cohort achieving particular levels. There are problems with misclassification of levels applied. Wiliam reports that "it is likely that the proportion of students awarded a level higher or lower than they should be because of the unreliability of the tests is at least 30% at Key Stage 2 and may be as high as 40% at Key Stage 3".

Criterion referencing fails to work well since question difficulty is not solely determined by curriculum content. It can also be affected by "process difficulty" and/or "question or stimulus difficulty", (Pollitt et al). It is also difficult to allocate curriculum to levels since questions testing the same content can cover a wide range of difficulty.

It is believed that error could be communicated meaningfully to schools, children, parents and the press, and would enhance both intelligence to ministers and the educational use of the data from national tests.

The current practice of overlaying levels onto the scores brings serious problems and it is clear that the use of levels should be reviewed. One key issue: consider the following:

Level 5 upper boundary

Child C

Child B

Level 4 upper boundary

Child A

Level 3 upper boundary

Both Child B and Child C are Level 5. But in fact Child A and B are closer in performance, despite A being level 4 and B being Level 5. Further, if Child A progresses to the position of Child B over a period of learning, they have increased by one level. However, if Child B progresses to the same position as Child C, they have progressed further than Child A over the same time, but they do not move up a level. Introducing sub-levels has helped in some ways (4a, 4b etc) but the essential problem remains.

2. *QCA practice in test development*

Pretesting items is extremely helpful; it enables the performance characteristics of each item to be established (particularly how relatively hard or easy the item is). This is vital when going into the summer levels setting exercise—it is known what is being dealt with in setting the mark thresholds at each level. But subject officers and others involved in the management of the tests have had a history of continuing to change items after the second pretest, which compromises the data available to the level setting process, and thus impacts on maintaining standards over time. In addition, the "pretest effect" also remains in evidence—learners are not necessarily as motivated when taking "non-live" tests; they may not be adequately prepared for the specific content of the tests; etc. This places a limit on the pre-test as an infallible predictor of the performance of the live test.

3. *Borderlining*

The decision was taken early in national assessment to re-mark all candidates who fall near to a level threshold. QCA publish the mark range which qualifies children to a re-mark. However, the procedure has been applied only to those below the threshold and who might move up, and not to those just above, who might move down. This has had a very distorting effect on the distributions. Although done in the name of fairness, the practice is seriously flawed. For years, arguments around changing the procedure or removing borderlining completely foundered on the fact that this would effect a major (downward) shift in the numbers gaining each level, and therefore could not be sanctioned politically. A poorly-designed and distorting practice therefore continued. Current practice is unjustifiable and would not be sanctioned in other areas of public awarding (eg GCSE and A/AS).

It has now been agreed between QCA and the DfES that borderlining will be removed in 2008, when the marking contract passes from Pearson's to the American ETS organization. At this point, a recalibration of standards could be effected to mask the effect of correction and this standard could be carried forward, or a clear declaration could be made on how removal of borderlining affects the "fairness" of the test and has resulted in a change in the numbers attaining a given level. First identified by Quinlan and Scharaskin

in 1999, this issue has been a long-running systemic problem. Again, it is a change in practice (alongside changes in the accessibility of the tests, in inclusion of mental arithmetic etc) which compromises the ability of the tests to track change in attainment standards over time.

4. *Fluctuations in Science at KS3*

At Levels 6 and above, standards of attainment have moved up and down in an implausible fashion:

2005	**37**	(% of children gaining Levels 6 and 7)
2004	**35**	
2003	**40**	
2002	**34**	
2001	**33**	

The movement over the three year period 2002–04 has involved a 6% increase followed by a 5% decrease—a movement of 11% over two years. This is implausible, and points to problems in the tests and level setting, and not to a real change in underlying standards or in the cohort taking the tests. Significantly, when interviewed on causes, officials and officers gave very different explanations for the effect—in other words, the true cause has not been established with precision.

5. *The Massey Report and Tymms' analysis*

The Massey report used highly robust method to triangulate national tests 1996–2001 and yields solid evidence that attainment standards have risen over that period, but not to the extent in all subjects and all key stages that has been argued by DfES and ministers. Tymms' less robust method and research synthesis suggests broadly the same. Massey made a series of recommendations, some of which have been adopted by QCA, such as equating a number of years' tests and not just the preceding year. However, the absence of a consistent triangulation method and the failure to adopt the Massey recommendation that standards should be held for five years and then publicly recalibrated has not been adopted.

6. *Ofsted's over-dependence on national test outcomes*

The new Ofsted inspection regime is far more dependent on the use of national assessment data than previously. This delivers putative economies since Ofsted feels it can better identify problematic and successful schools, and can use the data to target areas of schools—eg weak maths departments, or poor science etc. The revised regime is broadly welcomed by schools, and has a sound emphasis on each school delivering on its stated policies. But the regime fails to acknowledge the weaknesses of the data which lie at the heart of the pre-inspection reports, and which guides Ofsted on the performance of schools. The greatly increased structural dependence on data which is far less accurate than is implied is problematic. The new regime delivers some valuable functions—but the misapprehension of the real technical rigour of the assessment data is a very serious flaw in arrangements.

7. *Assessment overload accusations whilst using many other non-statutory tests*

This is an interesting phenomenon—the optional tests are liked, the statutory tests are frequently disliked (QCA). KS2 score are mistrusted (ATL). The use of "commercial" CAT tests and CEM's tests (MIDYIS etc) is widespread. CAT scores are trusted by teachers because the results are more stable over time in comparison with national curriculum tests; this reflects the different purpose of the respective instruments. Children say "I did SATs today" when they do a statutory Key Stage test. They also frequently say that when they have taken a CAT test. There is widespread misunderstanding of the purpose of the range of tests which are used. QCA was lobbied over a five-year period to produce guidance on the function of different tests—not least to clarify the exact purpose of national testing. However, no such guidance has been produced. As a result of this, the arguments regarding "over-testing" are extremely confused, and adversely muddy the waters in respect of policy.

8. *Is the timing right?*

Changing the timing of the tests would require a change in primary legislation. However, it is an enhancement of testing which should be considered very seriously. In the final report of the Assessment Review Group in Wales, Daugherty (2004) recommends that "serious consideration should be given to changing the timing of Key Stage 3 statutory assessment so that it is completed no later than the middle of the second term of Year 9". The Group believed the current timing to be unhelpful in relation to a process that could, in principle, inform, and that, "one source of information that would be of use potentially to pupils and their parents is not available until after the choice of pathway for Year 10 and beyond has been made". There are also implications for the potential use of Key Stage 1 and 2 data for transition between phases. "School ownership"—taking the outcomes very seriously in managing learning—would be likely to increase in this re-scheduling of the tests.

9. *The reliability of teacher assessment*

Particularly in the high stakes context of performance tables, we feel that relying on teacher assessment, as currently operated, is not a robust option. Work in 2000 by QCA Research Team showed a completely unstable relationship between TA and test scores over time at school level. This is compelling evidence against an over-dependence on teacher assessment. There are means of delivering moderated teacher assessment for reporting to parents, and bolstering accountability not by testing but by regional inspection based on high expectations and school improvement models (see recommendations below). National standards in underlying attainment could be delivered through a light sampling model (with matrix sampling to cover all key content of the national curriculum). This would enable a valid answer to the ministerial question " . . . nationally, what's happening to standards in English?".

10 *Teaching to the test*

The recent lobbying by Baroness Professor Susan Greenfield and eminent colleagues is merely the most recent critique of the problems of teaching to the test. The "Texas Test Effect" (Wiliam, Oates) is well known but poorly presented to Government. Bill Boyle (CFAS) is the latest empirical study of the adverse effects of teaching to the test and its almost universal domination of educational purposes in the English school system. It is a very serious issue, and it may be one significant factor (not the sole one) lying behind the "plateau effect" associated with the majority of innovations such as the Primary Literacy and Numeracy Strategies. In other words—a succession of well-intended and seemingly robust initiatives repeatedly run out of steam.

National Assessment—Annex 2

The Assessment of Performance Unit—should it be re-instated?

The origins and demise of the APU

1. The inability of current arrangements to provide a robust flow of policy intelligence on trends in pupil attainment has emerged as a serious problem. The causes are multifaceted, and include:

— instability in standards within the testing system (Massey, Oates, Stats Commission);

— acute classification error affecting assignment of pupils to levels (Wiliam, Tymms); and

— teaching to the test/"Texas Test Effect" (Wiliam).

2. Growing awareness of this issue has prompted increasing calls for " . . . a return to the APU . . ." (the Assessment of Performance Unit)—a separate, "low stakes", light-sampling survey for the purpose of reliable detection of patterns of pupil attainment, and of trends in attainment over time. But there are dangers in an unreflective attempt to re-instate arrangements which actually fell short of their aims. The APU processes were innovative and progressive. They mirrored the fore-running US NAEP (National Assessment of Educational Progress) and pre-dated the arrangements now in place in New Zealand and Scotland. Running surveys from 1978–88, politicians and civil servants saw it being redundant in the face of the data on each and every child of age 7,11 and 14 which would be yielded from National Curriculum assessment processes. The APU was hardly problem-free. Significant issues emerged in respect of:

— under-developed sampling frames;

— tensions between subject-level and component-level analysis and reporting;

— differing measurement models at different times in different subjects;

— lack of stability in item forms;

— escalating sampling burden;

— difficulty in developing items for the highest attaining pupils;

— the nature of reporting arrangements;

— replacement strategy in respect of dated items;

— ambiguity in purpose re "process skills" as a principal focus versus curriculum content;

— research/monitoring tensions; and

— compressed development schedules resulting from pressure from Government.

3. There was acute pressure on the APU to deliver. Rather than recognize that the function of the APU was essential for high-level policy processes and would need to persist (NEAP has been in place in the US since 1969), the intense pressure led to poor refinement of the technical processes which underpinned the operation of the APU, and high turnover in the staff of the different subject teams (Gipps and Goldstein 1983). Crucially, the compressed piloting phases had a particularly adverse impact; there was no means of undertaking secure evaluation of initial survey work and feeding in "lessons learned":

> " . . . the mathematics Group, in particular, felt that they were continually being rushed: their requests for a delay in the monitoring programme were rejected; their desire for three pilot surveys was realized as only one; they experienced a high turnover of staff and a resulting shortage of personnel. The constant rush meant that there was no time for identifying and remedying problems identified in the first year of testing.

4. In fact, all three teams suffered from a rapid turnover of staff, put down to the constant pressure of work combined with a lack of opportunity was no time to 'side track' into interesting research issues . . ." (Newton P 2005 p14)

5. This is not a trivial failure of a minor survey instrument. The APU was founded in 1974 after publication of a DES White Paper (*Educational Disadvantage and the Needs of Immigrants*). It was the result of a protracted strategic development process, which led from the DES-funded Working Group on the Measurement of Educational Attainment (commissioned in 1970) and the NFER's DES-funded development work on Tests of Attainment in Mathematics in Schools. If it had successfully attained its objectives, it would have relieved National Curriculum testing of the burden of attempting to measure standards over time—a purpose which has produced some of the most intense tensions amongst the set of functions now attached to national testing. Stability in the instruments is one of the strongest recommendations emerging from projects designed to monitor standards over time. In sharp tension with this, QCA and the State has—in line with commitments to high quality educational provision; the standards agenda; and responses from review and evaluation processes—sought to optimize the National Curriculum by successive revision of content; increasing the "accessibility of tests"; and ensuring tight linkage of the tests to specific curriculum content. These are laudable aims—and the emphasis on the diagnostic function of the data from tests has been increasing in recent innovations in testing arrangements. But pursuit of these aims has led to repeated revision rather than stability in the tests.

6. The Massey Report suggested that if maintenance of standards over time remained a key operational aim, then stability in the test content was imperative. In the face of these tensions, retaining an APU-style light sampling survey method would enable de-coupling of national assessment from a requirement to deliver robust information on national educational standards, and enable testing to reflect curriculum change with precision, to optimize the learning-focussed functions of testing, and enable constant innovation in the form of tests (eg to optimize accessibility).

7. Thus, the deficits and closure of the APU were, and remain, very serious issues in the operation and structure of national assessment arrangements. Temporal discontinuity played a key role in the methodological and technical problems experienced by the APU developers. As outlined above, rushing the development phases had a variety of effects, but the most serious of these was the failure to establish with precision a clear set of baseline data, accompanied by stable tests with known performance data; " . . . an effective national monitoring system cannot be brought 'on stream' in just a couple of years . . ." (Newton P, 2005).

8. Our conclusion is not "bring back the APU", but develop a new light sampling, matrix-based model using the knowledge from systems used in other nations and insights from the problems of the APU. Models 1 and 2 which we outline as alternatives in the main body of this evidence rely on the development of new versions of the APU rather than simple re-instatement.

Section 2

1. Higher Education Admissions Tests

Determining role and function

1. Since the publication, in September 2001, of the Schwartz report (Fair Admissions to higher education: recommendations for good practice), the issue of the role and function of admissions tests has been a controversial area. Cambridge Assessment has been cautious in its approach to this field. We have based our development programme on carefully-considered criteria. We believe that dedicated admissions tests should:

- — produce information which does not duplicate information from other assessments and qualifications;
- — make a unique and useful contribution to the information available to those making admissions decisions; and
- — predict students' capacity to do well in, and benefit from higher education.

2. Since the Cambridge Assessment Group includes the OCR awarding body, we are also heavily involved in refining A levels in the light of the "stretch and challenge" agenda—working to include A* grades in A levels, inclusion of more challenging questions, and furnishing unit and UMS scores (Uniform Mark Scheme scores –a mechanism for equating scores from different modules/units of achievement) as a means of helping universities in the admissions process.

3. We recognize that HE institutions have clear interests in identifying, with reasonable precision and economy, those students who are most likely to benefit from specific courses, are likely to do well, and who are unlikely to drop out of the course. We also recognize that there is a strong impetus behind the "widening participation" agenda.

4. Even with the proposed refinements in A level and the move to post- qualification applications (PQA), our extensive development work and consultation with HE institutions has identified a continuing need for dedicated assessment instruments which facilitate effective discrimination between high attaining students and are also able to identify those students who possess potential, but who have attained lower qualification grades for a number of reasons.

5. We are very concerned not to contribute to any unnecessary proliferation of tests and so have been careful only to develop tests where they make a unique and robust contribution to the admissions process, enhance the admissions process, and do not replicate information from any other source. To these ends, we have developed the BMAT for medical and veterinary admissions. We have developed the TSA (Thinking Skills Assessment), which is being used for admissions to some subjects in Cambridge and Oxford and is being considered by a range of other institutions. The TSA items (questions) also form part of the uniTEST which were developed in conjunction with ACER (Australian Council for Educational Research). UniTEST is being trialled with a range of institutions, both "selecting" universities and "recruiting" universities.

6. This test is designed to help specifically with the widening participation agenda. Preliminary data suggests that this test is useful in helping identify students who are capable of enrolling on courses at more prestigious universities than the ones for which they have applied as well as those who should consider HE despite low qualification results.

7. The TSA should be seen more as a test resource rather than a specific test: TSA items are held in an "item bank", and this is used to generate tests for different institutions. Although TSA items were originally developed for admissions processes in Cambridge where discrimination between very high attaining students is problematic and A level outcomes inadequate as a basis for admissions decisions, Cambridge Assessment research team is developing an "adaptive TSA". This utilizes the latest measurement models and test management algorithms to create tests which are useful with a very broad range of abilities.

8. The validation data for the TSA items is building into a large body of evidence and the tests are yielding correlations which suggest that they are both valid and useful in admissions—and do not replicate information from GCSE and AS/A2 qualifications. In other words, they are a useful addition to information from these qualifications and allow more discriminating decisions to be made than when using information from those qualifications alone. In addition, they yield information which is more reliable than the decisions which are made through interviews and will provide a stable measure over the period that there are major changes to AS and A2 qualifications.

The American SAT

9. Cambridge Assessment supports the principles which are being promoted by the Sutton Trust and the Government in respect of widening participation. However, we have undertaken evaluation work which suggests that the promotion of the American SAT test as a general admissions test for the UK is ill-founded. The five-year SAT trial in the UK is part-funded by Government (£800,000), the College Board (the test developers) contributing £400,000 and with the Sutton Trust and NFER each contributing £200,000.

10. The literature on the SAT trial in the UK states that the SAT1 is an "aptitude" test. It also makes two strong claims that are contested:

> " . . . Other selection tests are used by universities in the United Kingdom, but none of these is as well constructed or established as the SAT©.
>
> In summary, a review of existing research indicates that the SAT© (or similar reasoning-type aptitude test) adds some predictive power to school / examination grades, but the extent of its value in this respect varies across studies. In the UK, it has been shown that the SAT© is an appropriate test to use and that it is modestly associated with A-level grades whilst assessing a different construct. No recent study of the predictive power of SAT© results for university outcomes has been undertaken in the UK, and this proposal aims to provide such information . . ."
>
> Source: (http://www.nfer.org.uk/research-areas/pims-data/outlines/update-for-students-taking-part-in-this-research/a-validity-study-background.cfm)

11. The claim that "none of these is as well constructed or established as the SAT©" fails to recognise that Cambridge Assessment has assembled comprehensive data on specific tests amongst its suite of admissions tests and ensures that validity is at the heart of the instruments. These are certainly not as old as the SAT but it is entirely inappropriate to conflate quality of construction and duration of use.

12. More importantly, the analysis below suggests that the claim that the SAT1 is a curriculum-independent "aptitude" test is deeply flawed. This is not the first time that this claim has been contested (Jencks, C. and Crouse, J; Wolf A and Bakker S), but it is that first time that such a critique has been based on an empirical study of content.

13. It is important to note that the SAT is under serious criticism in the US (Cruz R; New York Times) and also, despite many UK-commentators' assumptions, the SAT1 is not the sole, or pre-eminent, test used as part of US HE admissions (Wolf A and Bakker S). The SAT2 is increasingly used—this is an avowedly curriculum-based test. Similarly, there has been a substantial increase in the use of the Advanced Placement Scheme—subject-based courses and tests which improve students' grounding in specific subjects, and are broadly equivalent to English Advanced Level subject-specific qualifications.

14. It is also important to note that (i) the US does not have standard national examinations—in the absence of national GCSE-type qualifications, a curriculum-linked test such as the SAT1 is a sensible instrument to have in the US, to guarantee that learners have certain fundamental skills and knowledge—but GCSE fulfils this purpose in England; (ii) the USA has a four-year degree structure, with a "levelling" general curriculum for the first year; and (iii) the SAT1 scores are used alongside college grades, personal references, SAT2 scores and Advanced Placement outcomes:

> " . . . One of the misunderstood features of college selection in America is that SATs are only one component, with high school grades and other 'portfolio' evidence playing a major role. The evidence is that high school grades are a slightly better predictor of college achievement than SAT scores, particularly for females and minority students. Combining both provides the best, though still limited, prediction of success . . ."
>
> (Stobart G)

Curriculum mapping—does the SAT mirror current arrangements?

15. In the light of research comment on the SAT and emerging serious criticisms of the instrument in its home context, Cambridge Assessment commissioned a curriculum mapping of the SAT in 2006 comparing it with content in the National Curriculum (and, by extension, GCSE) and the uniTEST.

16. It is surprising that such a curriculum content mapping has not been completed previously. Prior studies (McDonald *et al*) have focused on comparison of outcomes data from the SAT and qualifications (eg A level) in order to infer whether the SAT is measuring something similar or different to those qualifications. But the failure to undertake a comparison of the SAT with the content of the English National Curriculum is a serious oversight. The comparison is highly revealing.

17. The study consisted of a comparison of published SAT assessment criteria, items included in SAT1 sample papers, the National Curriculum programmes of study, and items within the uniTEST. The SAT assessment criteria and National Curriculum programmes of study were checked for analogous content. The National Curriculum reference of any seemingly relevant content was then noted and checked against appropriate SAT1 specimen items. The full analysis was then verified by researchers outside the admissions team, who were fully acquainted with the content of the National Curriculum and GCSEs designed to assess National Curriculum content. The researchers endorsed the analysis completed by the admissions test developers.

The outcomes of the curriculum mapping study

18. The full results are shown in Higher Education admissions tests Annex 3. Column 1 shows the sections and item content of the SAT1. Column 2 gives the reference number of the related National Curriculum content. For example, MA3 2i refers to the statement:

> Mathematics Key Stage 4 foundation
>
> Ma3 Shape, space and measures
>
> Geometrical reasoning 2
>
> Properties of circles

recall the definition of a circle and the meaning of related terms, including centre, radius, chord, diameter, circumference, tangent, arc, sector, and segment; understand that inscribed regular polygons can be constructed by equal division of a circle.

19. Column 3 in Annex 3 shows the relation between the content of the SAT1, the relevant components of the National Curriculum and the Cambridge/ACER uniTEST admissions test.

20. The analysis indicates that

- — the SAT1 content is largely pitched at GCSE-level curriculum content in English and Maths, and replicates GCSE assessment of that content; and
- — the item types and item content in the SAT1 are very similar to that of GCSEs.

It is therefore not clear exactly what the SAT1 is contributing to assessment information already generated by the national examinations system in England.

21. Previous appraisals of the SAT1 have been based on correlations between GCSE, A level and SAT1 outcomes. This has shown less than perfect correlation, which has been interpreted as indicating that the SAT1 assesses something different to GCSE and A level. But GCSE and A level are based on compensation—particularly at lower grades, the same grade can be obtained by two candidates with different profiles of performance. The inferences from the data were previously made in the absence of a curriculum mapping. The mapping suggests that discrepancies between SAT1 and GCSE/A level outcomes may be the result of the candidates not succeeding at certain areas in these exams, nonetheless gaining a reasonable grade—but this being re-assessed by SAT1 and thus their performance found wanting.

22. The existence of such comprehensive overlap suggests that the SAT either presents an unnecessary replication of GCSE assessment or an indication of the problems of compensation in the established grading arrangements for GCSE.

23. Identical analysis of uniTEST, currently being piloted and developed by Cambridge Assessment and ACER, suggests that uniTEST does not replicate GCSE assessment to the same extent as the SAT1 but focuses on the underlying thinking skills rather than on formal curriculum content. There is some overlap in the areas of verbal reasoning, problem solving, and quantitative and formal reasoning. There are, however, substantial areas of content which are not covered in the National Curriculum statements of attainment nor in the SAT1. These are in critical reasoning and socio-cultural understanding. This suggests that uniTEST is not replicating GCSE and does offer unique measurement. Preliminary data from the pilot suggest that uniTEST is detecting learners who might aspire to universities of higher ranking that the ones to which they have actually applied.

June 2007

Higher education admissions tests—Annex 3

SAT	*National Curriculum*	*uniTEST*
Writing Section		
Short Essay (1)		
organise and express ideas clearly	En3 1d-o	
develop and support the main idea	En3 1d-o	
use appropriate word choice and sentence structure	En3 1b, c, e, f	
Multiple Choice (49)		
improve sentences and paragraphs	En3 2a, b, c; 7a-e	
identify errors	En3 2a, b, c; 7a-e	
Critical Reading Section (67)		**Verbal and Plausible Reasoning** (30)
Sentence Completion		
knowledge of meaning of words	En1 5; 6e; En2 1a	
understanding of how parts of sentences fit together	En3 7a-e	
Passage Based Reading		
vocabulary in context	En2 1a, g, 3a, b	
literal comprehension	En2 1a, h, 3a, b	
extended reasoning	En2 1a-d, g, h, i	interpretation socio-cultural understanding
		Critical Reasoning (30) decision making argument analysis
Mathematics Section (54)		**Quantitative and Formal Reasoning** (30)
Number and Operation		
arithmetic word problems	Ma2 4a	problem solving
properties of integers	Ma2 2a-f	
rational numbers	Ma2 2 c-d	
logical reasoning	Ma2 1a-b	problem solving
sets	KS2 (NNS)	
counting techniques	KS2 Ma2 2a	
sequences and series	Ma2 1j, 6a	
elementary number theory	Ma2 2a	
Algebra and Functions		
substitution and simplifying algebraic expressions	Ma2 5	
properties of exponents	Ma2 6f	
algebraic word problems	Ma2 5 h	problem solving
solution of equations and inequalities	Ma2 5e, f, i-m	
quadratic equations	Ma2 5k	
rational and radical equations	Ma2 5b, f, k	
equations of lines	Ma2 6b-h	
absolute value	–	
direct and inverse variation	Ma2 5h	
concepts of algebraic functions	Ma2 5	
newly defined symbols based on commonly used ops.	Ma2 5a	

SAT	*National Curriculum*	*uniTEST*
Geometry and Measurement		
area and perimeter of a polygon	Ma3 2i (KS2 Ma3 2e)	
area and circumference of a circle	Ma2 5g, Ma3 2h	
volume of a box, cube and cylinder	Ma3 2i	
Pythagorean theorem and special properties of triangles	Ma3 2b, f	
properties of parallel and perpendicular lines	Ma2 6c; Ma3 2a	
coordinate geometry	Ma3 3e	
geometric visualisation	Ma3 2i	
slope	Ma2 6c	
similarity	Ma3 2g	
transformations	Ma3 2g, 3b-d	
Data Analysis, Statistics and Probability		
data interpretation	Ma4 1a 5	dealing with information
statistics (mean, median and mode)	KS3 Ma4 4b, e, f	
probability	Ma4 4b, c, d, g, h	

NB Key Stage 4 National Curriculum English and Mathematics (higher) references are used, unless indicated.

— The SAT and uniTEST were developed independently and, whilst there are some areas of similarity, there are many areas of difference. One of the clearest differences between the SAT and uniTEST is that much of the SAT assesses knowledge of formal elements of Mathematics and English—something that is intentionally avoided in uniTEST. The focus of uniTEST is primarily on the assessment of the reasoning skills that underpin higher education studies. This reflects the different contexts for which the two tests were designed: the SAT is used in the absence of a National Qualifications Framework; uniTEST would be used in addition to the National Qualifications Framework.
— The vast majority of the material covered by the SAT is included in the English and Mathematics programmes of study of the National Curriculum. The level of difficulty is roughly equivalent to Level 2 of the National Qualifications Framework (eg higher tier GCSE). It might be expected that performance on the SAT could be affected by a candidate's educational background, and by extensive preparation. Since uniTEST relies less on candidates' knowledge it is expected that, beyond a limited amount of familiarisation with the test, performance will be less susceptible to "coaching".
— Where there is overlap in the types of skills assessed by uniTEST and the SAT, the questions in uniTEST tend to be of a higher order of difficulty and complexity.

Memorandum submitted by the Chartered Institute of Educational Assessors (CIEA)[1]

1. The CIEA welcomes the opportunity to submit evidence to the select committee on Education and Skills and also welcomes moves by government towards more personalised learning suited to the needs of the individual and not the testing and assessment system but believes that more needs to be done in supporting good assessment practice in schools, colleges and the workplace.

2. The Chartered Institute of Educational Assessors (CIEA) is an independent charity set up with the support of DfES and QCA to improve the quality of assessment by providing continuing professional development to teachers, lecturers and workplace assessors as well as the external assessment community.

3. In England there exists a well-established external testing system used as a prime means of validating, verifying and reporting on students' attainments.

4. Although the cost of this system is large at £610 million (source PWC February 2005), it represents 1% of the total education budget of government and has created a system which attracts the continued confidence of the general public.

5. However, a system of external testing alone is not ideal and government's recent policy initiatives in progress checks and diplomas have made some move towards addressing an imbalance between external testing and internal judgements made by those closest to the students, ie the teachers, in line with other European countries.

6. The Institute welcomes these moves towards creating a balance between external testing and internal assessment of student's attainments. which could result, given the right professional support, in a teaching community more adept at different practices in assessment alongside rigorous standards moderated by the awarding bodies.

7. Other European countries have already made this move towards a balance between internal and external assessment. According to the Programme for International Assessment of Students (PISA), other countries, such as Finland, achieve good standards in education while relying more heavily on internal assessment, only undertaking external assessment in a student's progression at the point of transfer from secondary to higher education. If it is the UK Government's stated intention to continue to support young people with education and training up to the age of eighteen then the CIEA believes that a reliance on an external testing system need not be as great as is currently the case and other models of assessment such as those adopted by other European Countries could be investigated further.

[1] The Institute of Educational Assessors received its Royal Charter on 2 April 2008.

8. So a move towards a balance of assessment practice between the internal judgements of the teacher alongside external moderation from organisations such as the awarding bodies should be welcomed.

9. In England, according to Ofsted, one of the areas of concerns regarding school performance is that of assessment which remains one of the more variable practices within centres. The work of the National Strategies has helped to address these issues but the CIEA thinks more work needs to be done in supporting teacher's professionalism in this vital area of education, that of teaching, learning and assessment. The CIEA believes that a move towards balancing internal teacher judgements alongside rigorous standards in external assessment will help to address some of these concerns if teacher judgements can be supported through a framework that allows for the development of their own skills and capabilities in this vital area of education.

10. In support of this work to help teachers make valid and reliable assessment of their students' capabilities, The CIEA is working with the Training and Development Agency (TDA) to support initial teacher training in assessment. At present PGCE courses only allow for a limited time on assessment issues, sometimes as little as 60 minutes over a nine month course, so the CIEA has developed a toolkit for newly qualified teachers and students called *First Steps in Assessment* that seeks to detail the practice, process and techniques of assessment needed by every teacher entering the profession for the first time.

11. Once in the classroom, the use of standardised assessment instruments are not always the best solution in dealing with the individual capabilities of every student. Initiatives such as Assessment for Learning and a move to more personalised learning, in line with government policy, are going a long way to addressing these issues. However these initiatives need to be effectively supported by professional development opportunities for teachers if they are to be more meaningful and sustainable. This will help teachers make consistently valid and reliable judgements resulting in better quality of assessment in schools throughout the country.

12. Indeed a move to personalisation underpinned by regular assessments, both formative and summative (the latter informed by tests and tasks drawn from a national bank and moderated externally by the awarding bodies), over the course of a period of study, might provide a more comprehensive picture of the capabilities of students across a wide range of core competencies within a given subject domain. These assessments would provide a more rounded view of the achievements of students and enable learning programmes, personal to the student, to be developed while maintaining national standards through light touch external moderation.

13. Over the course of the coming months, CIEA will be outlining a role, purpose and scope for individuals based in schools, colleges and other places of learning to take the lead on assessment issues, ie Chartered Educational Assessors. This would help to address the problems of variability and inconsistency in assessment practices, which Ofsted has identified. The role of these centre-based specialists, supported by a programme of professional development from the Institute, would be to provide assessment tools, support, mentoring and coaching for colleagues. These individuals would also be responsible for quality assuring assessment processes and standards within schools so that a common approach to good assessment practice can be developed between subject specialists and across different subjects. They would ensure that the data derived from assessment is used to feed back into and improve the teaching and learning process rather than merely for reporting or tracking purposes.

14. The Institute has also developed a Professional Framework which could underpin an accredited programme of training and qualifications, delivered by CIEA's educational stakeholders, which would allow individual teachers to demonstrate their assessment knowledge and good assessment practices. The Framework outlines the role, competencies and behaviours needed by those undertaking assessment tasks in the classroom as well as externally via the awarding bodies. This tool is available online at www.ioea.org.uk. The Framework provides a structured approach to the process of assessment and the continuing professional development of those involved in assessment tasks so that, once assessment requirements have been identified, teachers can adopt good practice. The Framework will underpin the role of Chartered Educational Assessors in delivering better assessment practice.

15. All these developments are pertinent to the development of national curriculum and other assessments. For example, trust in the assessment system and, in particular, teachers' ability to make good assessments is essential to the changes, which the DfES' proposals on *Making Good Progress* envisage. The provision of progress tests for students is only one part of the full picture of delivering better educational standards. In order for progress tests to be meaningful we need to improve the system, processes and personal skills, which allow for good assessment practice in schools. This can only come about through the provision of professional support and development for teachers to allow them to become more expert at practising good assessment. Through structured training and qualifications as well as access to exemplification materials, tools and resources we can bring about better quality assessment in our schools on a more flexible and personalised basis in a consistent manner. Such an approach would effectively support more personalised learning that allows for the fulfilment of every student's potential in education. The CIEA hopes to contribute to these changes through its Professional Framework and Chartered Educational Assessor proposals.

16. *General Issues*

17. *Why do we have a centrally run system of testing and assessment?*

18. Originally we relied on external testing to identify those students who would progress to higher education though university entrance examinations. Under the auspices of the Department and its predecessor bodies, a uniform system emerged loosely regulated from the centres but with the setting of syllabuses (specifications) and examinations delegated to independent examination boards. Over the years, as more students stayed in education and took examinations and as competition for places and jobs intensified, the demand for greater comparability across examinations grew and equal access to curriculum and qualifications became the norm. The introduction of National Criteria for GCSE and a National Curriculum and associated assessment arrangements in the 1980's resulted in the centralised system we now have. We now use external testing and examinations as the prime validation of a student's achievements and as a means of measuring the performance of schools and colleges. Unlike other leading European countries, we provide limited effective support to teachers in assessment. The Assessment for Learning programme is making inroads here, although this does not necessarily better equip teachers with improved summative assessment skills and expertise.

19. *What other systems of assessment are in place both internationally and across the UK?*

20. In the UK, vocational examinations rely on assessment throughout the course and in the workplace. These assessments are valued by the community and by employers, with none of the concerns attached to the validity of teacher assessments in schools manifesting themselves in this context. Indeed, students who achieve good results in traditional tests and examinations do not always go on to become active contributors to society. As we are all aware, employers often criticise the skills of young people in terms of basic numeracy and communications skills.

21. The OECD's Programme for International Assessment of Students (PISA) has revealed that countries where teacher assessment is at the hub of the assessment system generally out perform countries more reliant on the external testing of students, such as in the UK. This need not be the case in future if we can develop better assessment practice within schools, colleges and the workplace to match current assessment standards provided by the external awarding bodies.

22. In both the 2000 and 2003 PISA surveys Finland, Korea and Japan performed consistently well across the range of measures covered in the surveys. In each of these countries compulsory school age student assessment is largely the responsibility of teachers rather than a reliance on external testing.

23. Finland's linear comprehensive education system, where students remain at the same schools from age 7–16, relies purely on teacher assessment to determine student achievement. All teachers must achieve a Masters degree prior to being allowed to teach. Schools are self-regulating (no external inspections) but must adhere to a national curriculum and national standards in assessment regulated by the National Board of Education.

24. Korea uses national scholastic achievement tests alongside continual teacher assessment based on a national curriculum. In the national tests two subjects are tested on one occasion each year with a sample of between 0.5 and 1% of the total student population in Years 6, 9 and 10, ie at ages 12, 15 and 18. The results of school assessment are, however, generally not made public. Test scores are not supposed to be made available to students or parents. The test results are primarily used to monitor school standards and to monitor student progress.

25. Japan's compulsory schooling between age 6 and 15 does not use compulsory testing. Certification of completion at both the end of elementary and junior high school phases is made on the basis of internal teacher assessment following structured national guidance based on a compulsory national curriculum.

26. Therefore a move towards balancing assessments between internal teacher judgements and externally moderated standards is welcomed by the CIEA in light of government's recently announced initiatives in Diplomas and progress tests.

27. *Does a focus on national testing and assessment reduce the scope for creativity in the curriculum?*

28. There is a danger in any system that learners will follow "guidance" diligently and that teachers will be fearful of deviating from it and "teach to the test". Creativity and learning may then become secondary to the need to perform well on the big occasion, whether it be SATS, GCSE or A-level, on the results of which the school/college will be judged. There is nothing new here: as long ago as 1889 this danger was recognised, as the following quotation from *The Sacrifice of Education to Examination, letters from "all sorts and conditions of men*, (ed Auberon Herbert) shows:

i. The evil done by examinations to the teachers, or rather to those who should be teaching, but whose energies are largely absorbed in examining, might be diminished if the present excessive number of examinations were reduced, if their minuteness and detail were lessened, and if a wider range of tests were permitted, and less excess weight attached to the power of covering paper within narrow limits of time . . . . . My own experience has shown me that (examinations) have the

> most widely diverse effects, both on the teacher and taught, according to the principles upon which they are arranged. Where a minute specification of topics, and still worse, of books, is adopted, they are mischievous to both. They fetter the teacher in arranging his course of education; and they lead the taught to aim at cramming and to be impatient of any instruction not obviously resolvable into tips.

29. QCA's recent curriculum revisions, intended to allow for more creativity in the application of the curriculum are welcome. However, all teachers do not yet have the necessary skills to be more flexible, more personalised and more adaptive in the use of new technology, new teaching techniques and new forms of assessment.

30. The recent curriculum review needs to be supported by structured Continuing Professional Development (CPD) programmes, as outlined above using the Institute's Professional Framework, for teachers to be able to make the most use of the greater flexibility that the curriculum allows.

31. *Who is the QCA accountable to and is this accountability effective?*

32. It would be inappropriate for the CIEA to comment, as this is a matter for QCA.

33. *What role should exam boards have in testing and assessment?*

34. Traditionally examination boards (or awarding bodies) have set as well as applied the national standard to their syllabuses, examinations and grading. More recently the regulator has taken on the responsibility for determining the national standard and monitoring its application. Awarding bodies continue to play a major role in the application of that standard through their specifications and the consistent marking to the standard of candidates' work in all subjects. Where teacher judgements are concerned the role of the examination board is to ensure that these, like the marks of examiners of externally assessed components, are consistent with the national standard—ie not to replace teacher judgements regarding the capabilities of their students but to moderate the marks given against the national benchmark.

35. Teachers are best placed to identify the strengths and weaknesses of individual learners but are not always given the right support, skills or training to be able to be effective in this role. There is a role for awarding bodies in both the moderation of teacher judgements and the local training of teachers within local networks of suitably qualified assessors supported through a structured programme of Continuing Professional Development (CPD) aimed at developing the skills of teachers in good assessment practice. The qualified assessors—the most experienced of whom would aspire to Chartered status—would be accredited to national standards by an independent organisation such as the CIEA, They would be capable of making consistent judgements regarding student progress and achievements over time. Such a system would provide assurances that teacher judgements provide assessments to given national standards.

36. Work on assessing the reliability of teacher assessments in comparison with national external testing has been developed by QCA in collaboration with DfES and the National Strategies and conducted by QCA via the Assessing Pupil Performance programme in English and maths which has been adopted for national rollout by the DfES. This work can be built upon by the introduction of Chartered Educational Assessors, as outlined above.

37. *National Key Stage Tests*

38. The current situation

39. *How effective are the current Key Stage tests?*

40. The current Key Stage tests have around fourteen objectives for their assessments. These include testing the individual student, testing the teacher of a group of students, assessing national standards of achievement for a range of students, assessing the performance of a group of teachers in a single school or across a number of schools, and also testing individual school's achievements in a single year and over time. In reality assessment instruments can only have a small number of objectives to be reliable, valid and fit for purpose.

41. Key Stage tests also provide a snapshot in time of an individual student's performance and, as intimated above, are not a valid and reliable indicator of the overall skills and capabilities of an individual student.

42. Rather than relying on a single snapshot in time of a student in a limited range of subjects, it would be preferable to assess their skills, knowledge and experience over a greater period of time in order to arrive at an effective assessment of that student's capabilities. This would lend itself to a balance between teacher-based assessment, moderated by professionals such as Chartered Educational Assessors, and supported by light touch moderation from the awarding bodies.

43. *Do they adequately reflect levels of performance of children and schools, and changes in performance over time?*

44. Key Stage tests reflect achievements of schools over time to a certain extent. In terms of individual children, standardised tests can never accurately reflect individual students' capabilities without being tailored to a certain extent for accessibility issues. More personalised assessments, which the Government has announced as its direction of travel, will achieve greater validity over time but there is a tension between validity and reliability where single tests will give more reliable but less valid results than more personalised tests.

45. As different tests are run each year then the nature of a series of standardised levels is somewhat misleading. To get an accurate picture of consistent standards over time the same test would have to be run with similar groups of students each year, but this is impracticable.

46. From research carried out by the Institute of Education it has been suggested that as many as 30% of all students achieve an incorrect grade or mark for their work using standardised test scores. There is nothing new in such findings: the first studies on reliability of marking in the late 19th century came up with similar findings. Consequently, tests are not a reliable indicator of a student's overall capability, although as the test score error is consistent over time from series to series, it brings with it a degree of consistency.

47. *Do they provide assessment for learning (enabling teachers to concentrate on areas of a pupil's performance that needs improvement)?*

48. Data from Key Stage tests can provide an indication of the likely areas students will need to develop in order to extend their capabilities but attention needs to be given to the nature of the questions selected which may not reflect accurately the knowledge attained by an individual student as the questions posed have been pre-selected by external examiners.

49. A weak link between testing and assessment for learning is the individual capability of the teacher to be able to accurately take data from tests and apply it to personalised learning plans, combined with personalised and more flexible assessment methods, for a group of students. Teachers struggle to validate this type of learning consistently and so more structured CPD programmes of support are needed for teachers to be able to become more effective in the development of personalised learning and assessment plans.

50. Additionally, can summative outcomes from tests be used on a daily basis in lessons to develop the formative processes of students? Professor Paul Black argues that it is unlikely and that Assessment for Learning is process driven not summative outcome driven. In a recent article in the CIEA's *Make the Grade* magazine, he said:

51. A frequent misunderstanding is that any assessment by teachers, and in particular the use of a weekly test to produce a record of marks, constitutes formative assessment. It does not. Unless some learning action follows from the outcomes, such practice is merely frequent summative assessment: the key feature—interaction through feedback is missing. Another misunderstanding is the belief that this is about the coursework assessment that forms part of some GCSEs; such assessment cannot aid learning unless there is active feedback to improve pupils' work as it develops . . . The research showing that a diet of marks does not improve learning, and that comments can do so only if pupils are not distracted by marks.

52. Key stage tests data does not readily transfer to aid assessment in the foundation subjects; this is a further argument for the introduction of highly skilled Chartered Educational Assessors to develop assessment practices in both formative and summative capacities for these subjects.

53. *Does testing help to improve levels of attainment?*

54. Testing helps to improve levels of attainment in tests but due to the nature of teaching to the test, many students have a grasp of a limited range of knowledge and skills which may not meet the needs of employers for more rounded students who are able to apply critical thinking skills from one area to another area.

55. The points made in response to the previous question also apply here.

56. *Are they effective in holding schools accountable for their performance?*

57. National Curriculum Tests are accepted by parents and the general public as giving a broad indication of the achievements of both individual children and individual schools at a moment in time and form part of the accountability framework of education to parents and the general public.

58. However, the tests are not perfect due to the reasons mentioned above. In light of this the CIEA believes that government policy initiatives around progress tests are set to address some of these issues however at present the objectives for which they are held are too numerous and there are concerns over validity and reliability.

59. *How effective are performance measures such as value-added scores for schools?*

60. Contextual value added scores are seen as a more reliable indicator of success than league tables based on raw data. Schools use this data to help in tracking performance along with other data, such as the Fischer Family Trust data and CATs (or their equivalent).

61. As an indicator of performance measures the value added data alone is not as reliable as when combined with other data sources but it is preferable to raw data.

62. *Are league tables based on test results an accurate reflection of how well schools are performing?*

63. Performance tables are only a small slice of the overall picture of a school or student's achievements. Other indicators such as the development of a student's attitude and motivation to learn, and to provide more holistic teaching experiences all provide a better quality measurement of a student's experience are missing from a simplistic testing instrument.

64. *To what extent is there "teaching to the test"?*

65. The answers provided to earlier questions are pertinent here. While little empirical evidence exists, anecdotal evidence suggests that teaching to the test is widespread among teachers and schools because the main focus of government and public attention is test results as a means of indicating education standards but the two are not the same. A further pertinent observation from *The Sacrifice of Education to Examination, letters from "all sorts and conditions of men* (1889) may be of interest to the Committee.

66. Everything is now sacrificed to the whim of the examiner, who may be a clever man, but who evidently writes his questions with the one aim of showing his own amount of learning. But the worst feature of the case is that all interest is taken out of the studies. A teacher must not now awaken an enthusiasm that will send a student to ransack a library on the loved subject,—because it is not prescribed by the examiner! We are becoming year by year narrower and shallower, more shut into one rut, more confined to a few subjects.

67. As that quotation from a different age would suggest, testing and examinations do not necessarily result in the provision of a rounded education to individuals who are capable of making an effective contribution to society. Rather, we may be churning out individuals who can pass tests and who can achieve good results to a given, known test, but who cannot necessarily apply their knowledge and skills to other situations, hence the concern from employers about skill levels among young people.

68. Instead we need a more rounded indicator of the capabilities of students' performance and the CIEA's view is that this needs to come from a prolonged assessment of an individual carried out in their locality over the course of their study by a suitably trained and qualified educational assessor, but still subject to moderation by the awarding bodies to ensure that national standards of achievement are maintained and education continues to attract the confidence of parents and the general public.

69. *How much of a factor is "hot-housing" in the fall-off in pupil performance from Year 6 to Year 7?*

70. There is a known and well-documented phenomenon which highlights the decline in attainment from Year 6 to Year 7.

71. This may be caused by the hothouse effect, namely learners studying a limited curriculum in the final term of Primary education in preparation for the National Curriculum Tests. On arriving in a new institution, they are inclined not to work as efficiently or effectively since there is no immediate terminal public examination, a further example of the undesirable backwash effect of external examinations on the curriculum.

72. Fall-off in performance may also be caused by learners switching from a regime where they are taught by the same person for all subjects, to a regime where they are taught by specialist teachers using specialist equipment in discrete physical locations for each curriculum area.

73. It may be caused by the need for individuals to reorganise their own social structures and hierarchies, having left one institution where each was the oldest within the hierarchy, to one where each is the youngest within the hierarchy. They also have to renegotiate their relationships with their peers, many of whom they have not met before. Each individual also has to learn the rights, responsibilities and rules within the new institution and develop a working relationship with others in the new institution.

74. The decline in attainment and progress made by some Year 7 learners is probably caused by a mixture of all of these factors.

75. *Does the importance given to test results mean that teaching generally is narrowly focused?*

76. Again, earlier comments are pertinent to this question.

77. Possibly. It may be narrowly focused on achieving a test result rather than producing high levels of educational standards among our young people so in real terms our competitiveness is not the highest among European countries as measured by PISA, and can be improved.

78. *What role does assessment by teachers have in teaching and learning?*

79. Earlier observations on the impact of external assessment on teaching and learning are relevant to this question. It seems to CIEA that, were teachers to develop their skills in educational assessment, the impact on the learning programmes would be beneficial. Teaching and learning only gives half of the picture. Without any educational assessment, teachers would not know what to teach next or if the teaching has been effective. In order to make the best use of the data provided by assessment, teachers need good support mechanisms, such as CIEA's Professional Framework.

80. Instead of being involved in teaching and learning, therefore, teachers need to be involved in teaching, learning and assessment. Like other countries which adopt a balance between internal and external assessment on students up to the age of 18, when decisions are made about career choices or higher education, teachers need to engage in assessment to a greater degree.

81. Better and more sharply focussed assessment by the teacher in the classroom would benefit the taxpayer who pays around £610 million to support the current external tests system as well as improving teaching and learning.

82. The Future

83. *Should the system of national tests be changed?*

84. We need to develop national tests over time in line with the needs of students and parents. More localised assessment supported by a rigorous programme of structured CPD, providing the appropriate skills to teachers to allow them to mark to national standards and to allow for both personalised learning and a more flexible response to the needs of individual students. This would still need to be supported by a system of national external moderation to ensure that assessments continued to attract the confidence of the general public in educational standards over time.

85. The CIEA Professional Framework and the Chartered Educational Assessor, described earlier in this response, could be the tools to provide such teacher judgements to national standards.

86. *If so, should the tests be modified or abolished?*

87. The Secretary of State has suggested that there should be a move to more personalised assessment to measure how a pupil's level of attainment has improved over time. Pilot areas to test proposals have just been announced. The CIEA supports this move.

88. *Would the introduction of this kind of assessment make it possible to make an overall judgment on a school's performance?*

89. Although the proposals are to be welcomed as an indicator of intent, they are too rigid and inflexible as they seek to address issues of performance by rolling out more external testing of the key stage type which are the CIEA's underlying causes of concern about educational standards in this country.

90. *Would it be possible to make meaningful comparisons between different schools?*

91. For the reasons given above, simply rolling out more external testing may not address the issues behind improving our educational standards.

92. *What effect would testing at different times have on pupils and schools?*

93. It is hard to predict if we don't know what we are testing or how we are testing it. Clearly the organisational implications of greater personalisation would need to be considered. Those in schools and colleges are best placed to comment on this aspect of change.

94. *Would it create pressure on schools to push pupils to take tests earlier? If Key Stage tests remain, what should they be seeking to measure? If, for example, performance at Level 4 is the average level of attainment for an eleven-year-old, what proportion of children is it reasonable to expect to achieve at or above that level?*

95. If students were given the right amount of support by suitably qualified teachers in educational assessment then they would be able to take tests when they, and not the system, are ready. This would help to ease the burden on an already stretched external testing system. However, if students' testing could be undertaken in the classroom and externally validated through light touch moderation by the awarding bodies and more localised assessment support, then there is no reason why this move to balance internal and external testing should not work as is the case in other European countries.

96. Over time more personalised learning development plans could be introduced supported by personalised assessment plans that would allow each student to develop to his or her own potential in a supportive system.

97. A number of schools do this already and there is a case for on-demand testing that is being increasingly advocated by leading assessment experts, such as Professors David Hargreaves, Paul Black and Dylan Wiliam.

98. *How are the different levels of performance expected at each age decided on? Is there broad agreement that the levels are appropriate and meaningful?*

99. These indicators are set by QCA following consultation with educational and awarding body stakeholders who are better placed than CIEA to comment on their appropriateness

100. Testing and Assessment at 16 and after

101. *Is the testing and assessment in "summative" tests (for example, GCSE, AS, A2) fit for purpose?*

102. Like the national curriculum tests, the CIEA considers that these public examinations may not be ideal but they are accepted by parents and the general public as accountable measures of attainment.

103. There is, however, a very real question as to whether we need a national system of qualifications at age 16 when, increasingly, students are staying in education and training. Indeed a recent government policy was announced in the autumn of 2006 indicating that all children will be supported by access to education and training opportunities until the age of 18, effectively raising the school leaving age by two years.

104. The question of the appropriateness of a public examination at 16 was raised in the 1940s by the Norwood Report and when GCE O and A-levels were introduced in 1951 the expectation was that those students who progressed to A-level (at that time very few) would by-pass O-level. The current talk of an 18 year old leaving age would point to a 16 year old "leaving examination" being redundant and could be replaced by a system of moderated and standardised teacher assessment, with all the training, constraints and use of national standards as mentioned earlier in this response. Again, the concept of Chartered Educational Assessors would be an important feature of such a system.

105. An externally provided qualification like A level and others (the Diploma, for example) would remain relevant and necessary to 18 year olds as a summative statement of their achievements in the education system and as an entry into Higher Education and employment. Whether it would be right to rely on these assessments as a means of holding schools and colleges accountable to the public and government is a question that needs debate.

106. Additionally tests that are truly valid are tailored to the needs of individual students in assessing what needs to be assessed in the right manner. However, tests that are reliable across different groups of students are naturally standardised around a norm and are hence could be considered not truly valid tests.

107. *Are the changes to GCSE coursework due to come into effect in 2009 reasonable? What alternative forms of assessment might be used?*

108. Course work has had various interpretations over its many years of existence—specified work carried out at home or in the field, as is often the case now in examinations, or an assessment of the work carried out by students as part of the course. Highlighting the latter enables a coherent picture of the students' attainments in the course to emerge and to provide a rounded picture of what has been achieved. Work carried out as part of the course should be work undertaken in the classroom and validated by the teacher over the course of the academic year which can, therefore, be a true reflection of the learning that is taking place. It also overcomes the problems of plagiarism. Accredited specialist assessors in a school, such as Chartered Educational Assessors, could be a means of ensuring good quality school/college-based assessments which command public confidence.

109. *What are the benefits of exams and coursework? How should they work together? What should the balance between them be?*

110. Earlier answers are pertinent to this question.

111. Formal exams can be a useful measure of knowledge gained and retained and can help the learners, to some extent, use their skills developed over the course of study. However, in some subjects like history, learners will not have the opportunity to use acquired skills, eg research skills, in a formal exam. Assessing the work of the course enables those wider skills to be formally acknowledged in the profile which is built up of a student's capabilities and attainment.

112. On the other hand there is a constant risk of plagiarism if coursework is uncontrolled and learners' time may be disproportionately used for the relatively low weightings for coursework in some subjects, although others have significantly more weighting of marks to coursework.

113. The current review of coursework and the introduction of controlled tasks is therefore important and the CIEA would appreciate an input into this.

114. *Will the ways in which the new 14–19 Diplomas are to be assessed impact on other qualifications, such as GCSE?*

115. There is always a danger in introducing a new qualification alongside existing, respected qualifications. Steps have to be taken to ensure that the new product gains respect and is wanted by both the learner and users of qualifications such as HE and employers. In the case of the new Diploma there is evidence of a real intention to make the new qualification work and of cooperation across educationalists, employers and awarding bodies. The CIEA is anxious to support the new qualification and, in particular the element within the Diplomas which relies on "local" assessment—ie assessment of applied learning. We envisage this being undertaken by a Chartered Educational Assessor, authenticated by the CIEA and we are pleased to have been involved in discussions with QCA, NAA and SSAT. The Chartered Educational Assessor could quality assure the assessment regimes across consortia, at both the local school or college level and across the consortium as a whole. This quality assurance will underwrite the accuracy of the assessments across the consortium and validate the assessment outputs.

116. The assessment of the new Diplomas will be different in kind from assessments that have gone before them. The new diplomas will rely on a mixture of end of unit tests, managed by the diploma awarding bodies—similar in type to the current GCSE or GCE. In addition, however, they will have an element of local assessment that will focus on the application of skills, understanding and knowledge.

117. This new approach could have implications for GCSE and for the moderation of course work, since it will allow those individuals accredited as a Chartered Educational Assessors to moderate the assessment of a school and leave the awarding body with only a light touch sample of moderation to be carried out. It could, therefore, demonstrate that such an approach might be relevant in other contexts like GCSE, although there would continue to be a need for awarding bodies to take steps to assure national standards. The CIEA has no purchase or ownership of those standards; our role is in relation to the support of teachers and the continuing professional development which they need.

118. *Is holding formal summative tests at ages 16, 17 and 18 imposing too great a burden on students? If so, what changes should be made?*

119. The answer given above to testing after 16 is relevant here. In brief, if government policy is to ensure training and education for all individuals up to the age of 18, then the need for national formalised external testing is reduced below this age level.

120. Instead we should aim for national cohort sampling to get national standards of achievement and opt for more localised teacher-based assessment supported by light touch moderation from the awarding bodies up to the age of 18 with external testing being retained at age 18 when career choices are made to continue studying at higher education or enter the workplace.

121. *To what extent is frequent, modular assessment altering both the scope of teaching and the style of teaching?*

122. Modular assessment has had major benefits in enabling students for whom a single end-of-course assessment would be too great a hurdle to attain the standard of a qualification such as A-level. However, there needs to be a balance struck in any modular course between the coherence of the whole and its fragmentation into shorter learning chunks. Too many modules can, in this respect, be as detrimental as a single end of course examination was to the attainment of the cohort as a whole.

123. Allowing teachers more flexibility to deliver and assess via modular courses has merits, such as not teaching to a final end-of-course test (although more atomised testing can have the same effect), and demerits, such as the possibility of not applying consistent and rigorous processes and standards. The awarding bodies have a major role to play here ensuring consistency and rigour from module to module as well as across the whole subject.

124. Modular assessment can allow for a wider scope of teaching and learning styles to be accommodated provided that the assessment instrument applied to a particular module is sensitive to the objectives of the module—a practical orientation, for example, requires a different form of assessment than does concentration on factual knowledge. Modular learning might be appropriate to individualised learning pathways.

125. *How does the national assessment system interact with university entrance?*

126. University entrance is traditionally based upon achievement at "A" Level. Over recent years alternative demonstrations of reaching the standard required have been accepted—for example vocational qualifications, the IB and access courses. Widening participation is taken very seriously by Higher Education.

127. However, according to some universities the number of candidates acquiring higher grades suggests that traditional qualifications provide insufficient discrimination to enable the best students to be identified—hence the call for an A*grade. A number of selective universities have based their entry requirements on unit grades, that is the scores and grades achieved, not in the subject as a whole, but rather in the scores attained in each of the unit tests, thereby giving a fuller picture of attainment.

128. This equates to a six-fold increase in the amount of data available to an admissions tutor. Many universities are unlikely to have the staffing to interpret or collect such a wealth of data.

129. Again, it is easy to forget that concern over the quality of students entering university is not a new phenomenon. In 1960 when a mere 5% of the student population entered Universities, the Northern Universities' Joint Matriculation Board observed in its Annual Report.

130. Among freshmen in general the level of ability to write English is disappointingly low. The suitability of the present GCE Examination in English Language at the Ordinary level is not here being criticised so far as it concerns the 16-year-old candidate for whom it was designed, although opinion about this aspect of the examination is not wholly favourable. It seems to be generally agreed however that the degree of ability to express oneself which might be accepted from the 16-year-old candidate is not sufficient at university entry, that too often apparently such facility as may be present at 16 is not encouraged to develop *pari passu* with the development which goes on in the other aspects of the Sixth form curriculum. It may well be that if all the students were sufficiently "literate" at entry, some of them might lapse into comparative "illiteracy" while at the university unless care were taken to ensure that further development is actively encouraged and fostered within the university itself. That is a matter for the university authorities themselves; the duty of the Board is to ensure that at entry those who have been examined by the Board as potential university students have gone further than what is now accepted as O-level English Language." (AQA Archive, 1960)

131. *What does it mean for a national system of testing and assessment that universities are setting entrance tests as individual institutions?*

132. Clearly this is all about the ability of universities to select with confidence the best students for their courses. Were they to set their own tests, they could undermine public confidence in test outcomes produced by the awarding bodies. It would not be in the best interest of students if they were faced with a battery of individual entry examinations. In reality only a very few—admittedly the most prestigious—universities would take this step and it is questionable whether even they would have the resources needed to do so.

133. More worryingly is the stand which universities may take on the value of the new Diplomas. Their acceptance for university entrance is critical to their value and to public confidence.

134. Teacher-based judgements on the abilities of students within the summative A-level system would enable a more rounded picture of the students to be provided to users of qualifications, including the universities. Those judgements would need to be supported by CPD in order to command public trust. As stated earlier, other countries like Finland already do this.

135. Concluding Remarks

136. The CIEA would be pleased to elaborate on any part of this submission in a written form or in person, if called to give evidence to the Committee. As an Institute we are committed to improving the standard of assessment in schools, colleges and the workplace, to supporting those involved in assessments through CPD and to increasing public confidence in assessment by means of Chartered status. We are well placed to support new initiatives such as progress tests and the new Diplomas and are ready to work with educational and employer stakeholders to ensure that these and other initiatives improve the quality of learning and its outcomes for the benefit of students and the nation.

May 2007

Memorandum submitted by Institute for Public Policy Research (IPPR)

Executive Summary

S.1 Pupils, teachers and parents all need assessment for different reasons. Pupils need assessment to assist them in their learning. Teachers need assessment so they can understand their pupils' needs and consequently adapt their teaching. Parents need assessment so that they can understand how their children are developing and how well their school is performing. Head teachers, governors, local authorities and central government all need assessment to hold schools accountable for the outcomes they achieve.

S.2 It is appropriate to have a national system for assessment. These objectives apply for pupils, parents and schools in a similar way across the country. It is also important that the system is widely understood and trusted by parents and teachers especially, and this would be more difficult under system which varied substantially from place to place.

S.3 The current English system achieves all of the objectives to some extent. However, it suffers from some significant shortcomings. The current national tests do not provide highly reliable or valid measures at the level of the individual pupil. National tests are used in performance tables and can trigger Ofsted inspections. They thus exert strong incentives on school leaders and teachers, and this can have unintended outcomes. The current assessment and accountability framework can impact on valuable elements of assessment such as assessment for learning. This can happen through narrow and shallow learning, questions-spotting and risk-averse teaching.

S.4 Any system will need to achieve a balance in respect of all the objectives. We believe a reformed system can better meet the objectives in the round. We believe that it is possible to change the assessment system so that it facilitates better teaching and learning, provides a better measure of child attainment, and maintains a high level of school and national accountability. In summary, our proposals are:

a. Every child should be assessed throughout each key stage by their teachers.

b. Every child should sit a small number of national tests at the end of each key stage, but not in every area of every subject. The results of these monitoring tests should be used to measure overall school performance, but not the individual pupil's attainment.

c. The school's performance should be used to moderate the teacher assessments, producing detailed, nationally-comparable data for each pupil.

S.5 We do not claim to have described in detail the perfect model, but rather we have presented here the broad features of such a system. It would represent a revolution for English education, but could potentially meet each of our objectives better than the existing system of assessment:

a. It should be more reliable, because teacher assessments could be based on much more information than can be captured in a single round of tests.

b. It should be more valid at the national level as there would be data on a wider range of subject areas.

c. It should have greater validity at the pupil level, because teacher assessments could more successfully measure the different aspects of each pupil's progress.

d. It should thus provide a better measure of pupil attainment, beyond pen and paper tests.

e. The monitoring tests should maintain accountability at the school level, and should provide a better measure of national progress on standards.

f. It should facilitate assessment for learning.

g. Teachers would have more time to focus on the development of other important skills such as noncognitive skills.

A. *Objectives for the assessment system*

A.1 Pupils, teachers and parents all need assessment for different reasons. Pupils need assessment to assist them in their learning. Teachers need assessment so they can understand their pupils' needs and consequently adapt their teaching. Parents need assessment so that they can understand how their children are developing and how well their school is performing. Head teachers, governors, local authorities and central government all need assessment to hold schools accountable for the outcomes they achieve.

A.2 More formally, the assessment system as a whole should achieve the following aims:

a. It should be valid and reliable. The assessment system should measure what it claims to measure—generally, the degree to which a pupil has learned and understood a subject or skill. In addition, it should produce reliable results: the measurement should be reasonably accurate and not suffer from a lot of random variation between schools or over time.[2]

b. It should periodically identify and record the level of achievement that individual pupils have reached. Such certification is probably the most traditional function of the assessment system, and is often called "summative" assessment.

c. It should allow parents and others to hold learning institutions accountable for their performance. This is probably the aspect of assessment that has most increased in importance over the last decade with the arrival of school performance tables and a system of national targets for key stage results. This function has previously been described as evaluative assessment, where it is the education service rather than the pupil that is being evaluated.

d. It should facilitate the learning of the pupil. Assessment should help identify the gaps and weaknesses in an individual's understanding, as well as the strengths that can be built upon, so as to help inform the teaching and support subsequently received by the pupil. Such assessment is often now called formative assessment, or assessment for learning.

e. It should be clearly understood and enjoy public trust and confidence, particularly among parents and teachers.

f. It should enable the monitoring of national standards over time.

A.3 It is important to note that these are objectives for the assessment system as a whole, not for each component part. One of the key arguments we make later in this paper is that different forms of assessment are better suited to achieving different objectives, and that our current emphasis on one form of assessment—the national curriculum tests at the end of the key stages—is having undesirable side-effects.

A.4 How much weight should we accord each of these objectives? To start with, validity and reliability are a prerequisite of achieving the others. Whatever use assessment is put to, it must measure with reasonable accuracy what we want it to measure. As we will see later in this paper, this is more demanding than it sounds. Trust and confidence, meanwhile, should be the outcome of a system of assessment that is widely seen to meet the other criteria effectively.

A.5 When and why do we need summative assessment? During the course of schooling, parents want to know that their children are making sufficient progress, and often pupils themselves do too. Summative assessment also helps pupils and parents choose options such as GCSE subjects or a specialist secondary school. However, the major occasion for such assessment occurs at the end of secondary school, when GCSEs and/or GNVQs (General National Vocational Qualifications) become the passport to further education and employment. This objective is thus important, but much more important at the end of schooling than before this point.

A.6 Since the 1970s schools and teachers have become much more accountable for what they do and for the outcomes achieved by their students. This is absolutely right and appropriate. Accountability means that parents and the public can understand what is happening in each school (and in the schools system as a whole); that schools take responsibility when things go wrong (and also when things go well); and that schools are responsive to pressure for change and improvement. All of this requires that good-quality information on school performance is available, and assessment is one of the critical means of providing this. If we are going to recommend changes to the assessment system, we need to be very careful that they do not weaken school accountability.

A.7 Perhaps the most important role of assessment is to help pupils learn. Assessment for learning is sometimes described as "the new pedagogy", but it has been a live issue in education theory for some time.

[2] For a detailed discussion of reliability and validity see Wiliam (2000a, 2000b).

A.8 The most effective schools now practise a culture of continuous teacher-led assessment, appraisal and adjustment of teaching practices to personalise learning for all their pupils. It seems clear that assessment that does not assist the learning of the child is of very limited value, and in many ways the certification of achievement and the accountability role of assessment are only important because of their links to this.

A.9 The public debate about assessment in schools often seems self-contradictory. Those who care about equality sometimes call for an increased focus on low-attaining pupils at the same time as complaining about the burden of assessment—presumably over concern about the means of identifying the relevant pupils in the first place. Meanwhile, if national test results go up, some will take this as evidence that the tests are getting easier, while failure to make progress towards the relevant target will also be criticised.

A.10 As we will go on to explain, these apparent contradictions can be addressed through changing the nature of the assessment system. We are not simply thinking in terms of 'more' or "less" assessment, but of changing the nature of assessment mechanisms and the shifting balance between them.

B. *How does the assessment system currently work?*

B.1 The foundations of the current system of assessment were brought into force by the Education Act of 1988. National testing at the ages of seven, 11, 14 and 16 accompanied the introduction of the National Curriculum, which for the first time specified programmes of study and attainment objectives for all pupils attending maintained schools. There were many benefits to the introduction of the National Curriculum, in particular improved coherence across hitherto uncoordinated geographical areas and different phases of schooling. Training for teachers also accompanied its roll-out. The system of a National Curriculum, with national tests and target levels of attainment at the end of each key stage, is still in place today.

B.2 The initial roll-out of national assessment was accompanied by a move to capture the potential of assessment as a pedagogical tool, in line with the intentions of the independent Task Group on Assessment and Testing (TGAT) that was set up to advise the Government of the day on these matters. Yet the Conservative government's emphasis on using parental choice of school as an incentive for schools to improve their performance, and the accompanying stress on results as a very public form of school and teacher accountability, led the pendulum to swing away from formative uses of assessment in England over the 1990s.[3]

B.3 The model that became dominant is sometimes described as a "high-stakes" assessment system. The stakes were high for schools first because their results were published for parents and the public to see, and second because poor results would attract the attention of the schools inspectorate, Ofsted. One important concern is that such high stakes systems may give schools perverse incentives to do things that are not in the best interests of their pupils, an issue we investigate below. However, we do want to give schools strong incentives of the right kind to help their pupils do well, and we certainly want to be able to identify underperforming schools so that we can intervene where necessary. To some degree there will thus always be a "high stakes" aspect to an assessment system that holds schools accountable.

B.4 Labour governments since 1997 have broadly endorsed, developed and placed more weight on the system of national tests taken in key subjects by all children at the end of Key Stages 1 to 3. Great emphasis has been placed on increasing the proportion of pupils who achieve the target levels in these tests, and on minimising the number of schools where unacceptably low proportions of pupils do so. This is the standards agenda; we discuss the outcomes in terms of attainment in another paper (Brooks and Tough 2006), where we emphasise the need for standards to continue to improve, and for the attainment gaps between different groups to close.

B.5 A sustained faith by the Government in the possibility of a quasi-market in school places, in which parental choice is meant to drive up standards, has meant a continued emphasis on school performance tables ("league tables"), including their extension to Key Stage 3. New, "value added" measures of attainment that focus on pupil progress have been introduced to give a truer picture of school performance than that provided by raw results, and from 2007 "contextual value added" data will take into account the individual and family characteristics of pupils to further improve the measure of school performance. Results are now used to inform school decisions about performance-related pay, to inform Ofsted decisions about whether schools should be given light or heavy touch inspections and, combined with targets, to inform judgments about the efficacy of educational initiatives such as the Primary Strategies.

[3] In 1988 the Government accepted proposals for national assessment put forward by the Task Group on Assessment and Testing that included assessment at seven, 11 and 14, driven by an aim to support formative practice. It concluded that there should be a framework of clear criteria, or criterion referencing, spanning age ranges, so that teachers could make best use of assessment to inform learning. Under the TGAT's proposals, teacher assessment was combined with external tests and standards were to be maintained by teachers comparing results with the results of the national tests and with the judgments of other teachers. It emphasised that league tables should not be published. For a historical account of how these proposals were gradually abandoned, see Daugherty (1995).

B.6 The consequence of using pupil assessment to judge teachers and institutions in this way has been that external testing has decisively eclipsed teacher assessment for all phases beyond Key Stage 1, where the mode for national assessment was reverted to placing a much greater emphasis on teacher judgments with only a Teacher Assessment level being reported nationally from September 2004[4]. Pupils still sit tests at Key Stage 1, but these are used to corroborate and inform the teacher's judgment alongside references to the pupil's work over the year. There is, however, considerably more flexibility in terms of which tasks/ tests they can use and when (see www.teachernet.gov.uk/educationoverview/briefing/news/ks1assessment/).

B.7 Yet at the same time the Government has recognised the need to both make better use of teaching professionals, and to deliver more personalised teaching, through assessment for learning. This can be thought of as operating at two new and distinct levels.

B.8 The first level is a more nuanced use of data in school and system management. Individual, pupil-level national assessment data, initially collated for the purposes of measuring the value added by schools (and hence holding schools accountable), has been built on and assimilated into sophisticated data banks that provide a highly versatile tool for a sensitive and contextualised version of performance management—or what former Schools Minister David Miliband has termed "intelligent accountability" (Miliband 2003). Cross-system data, and school-level data, aggregated from individual performance measures can be used for monitoring performance issues, evaluating progress, diagnosing difficulties, setting appropriate targets based on a full understanding of what might be achievable, and deciding where to channel effort. The schools system has become incredibly data-rich.

B.9 On one hand this has facilitated top-down performance management by central government. However, the data has also been made accessible at the local level, transforming it into a potentially highly valuable diagnostic tool. Through the allowing of local access to contextualised data via Ofsted's Performance and Assessment (PANDA) Reports, the Department for Education and Skills (DfES)'s Pupil Achievement Tracker software (PAT) and its successor RAISEonline (Reporting and Analysis for Improvement through School self-Evaluation), there is now potential for local authorities and school leaders to use the data to set appropriate attainment targets for their pupils, to assess their progress against that of similar pupils elsewhere, and to compare their own test results against good national comparators. The level of detail available goes down to being able to compare different groups of pupils' success with different types of questions. Assessment has become a powerful tool for supporting informed and rigorous self-management and we should be careful not to lose this valuable information in any reform of the system.

B.10 The second level is the positive promotion of formative assessment as one of the most effective pedagogical approaches. Although originally underplayed by the Government in the wake of TGAT, research has convincingly shown that formative assessment is one of the most effective pedagogical approaches for improving attainment, differentiating teaching and nurturing vital skills such as self-regulation (Black and Wiliam 1998, 1998a). As a result, the idea of using frequent, interactive teacher- and pupil-led assessment is being widely endorsed and is now viewed by many as the ultimate tool for personalised learning.

B.11 DfES and associated governmental agencies such as Ofsted have produced a wealth of advice and materials for teachers specifically aimed at increasing understanding of effective formative assessment practices, for example adopting the Assessment Reform Group's "10 Principles for Assessment for Learning" and adapting it into a poster for schools. Meanwhile, the Qualifications and Curriculum Authority (QCA) has been asked to review the Key Stage 3 curriculum, specifically with a view to giving advice on strengthening teacher assessment in foundation subjects, and to developing a bank of tasks relating to attainment targets that can form the basis for formative assessment (although external assessment remains a non-negotiable staple for end-of-phase assessment in core subjects).

B.12 While proponents of formative assessment have traditionally placed emphasis on comment-only marking (to avoid pupils focusing only on their marks), software drawing on pupil-level data creates the possibility for teachers to use performance data as a starting point for formulating rigorous and appropriate targets for individuals that take into account realistic projections of how similar learners have progressed. Individual question analysis can be used to explore pupil strengths and weaknesses in particular modes of thinking as well as areas of study.

B.13 In summary, the current system continues to place enormous weight on national tests at the end of the key stages, while showing signs of a newer emphasis on assessment for learning. The big question is: do the various elements of the assessment system fit together in a way that successfully achieves our objectives? How does the current system match up to our objectives of validity and reliability, providing appropriate measures of achievement, ensuring accountability, facilitating learning, and achieving public understanding, confidence and trust? This is explored in more detail below.

[4] The first annual statistics reporting teacher assessment only therefore come from the 2005 data.

C. *Does the current system meet our objectives?*

C.1 Validity, reliability and recording achievement

Unfortunately, it turns out that the existing Key Stage tests are not very reliable at correctly classifying the level of attainment of individual students. The fundamental reason for this is that they rely on a series of tests that can only cover either a small area of the curriculum in detail, or a broad area very lightly. Work undertaken by Dylan Wiliam has estimated, with generous assumptions about the reliability of the Key Stage tests, that around 32% of Key Stage 2 results and around 43% of Key Stage 3 results are misclassified by at least one level. For a technical discussion of this work see Wiliam (2000b, Black and Wiliam 2006).

C.2 One apparently obvious solution for improving the validity of the tests would be to make them longer and thus cover a broader range of material. Unfortunately, the accuracy of the test only creeps up very slowly as its duration is increased, so that increasing the test by a reasonable amount of time will only slightly reduce the numbers of pupils being misclassified. To ensure that Key Stage 2 tests classify no more than 10% of pupils incorrectly, the tests would have to be over 30 hours long for each subject (Black and Wiliam 2006). This is not a problem specific to the current design of the Key Stage tests; rather it is an inherent problem with examinations where every entrant has to answer the same set of questions to test a reasonably extensive subject, so it applies equally to many qualifications.

C.3 A more fundamental problem emerges when we look for evidence of the validity of the Key Stage tests. These tests are focused on the core subjects of literacy, numeracy and science, on the basis that they represent core skills that are vital to every young person's future development. We thus need to be confident that they are providing a valid measure of pupils' true abilities with respect to these core skills.

C.4 In order to consider evidence for or against this hypothesis we would need to examine results from the national tests with a different, independent measure that reasonably claimed to be testing the same skills. If Key Stage test results mirror the independent measures, for example if both improve over the same time period, then this is some corroborating evidence that both are valid. However, if Key Stage test results are going in one direction while the independent measures are going in the other, then this is evidence of a problem with at least one of them.

C.5 The question of national standards recurs every time results are published for key stage assessments, GCSEs and A Levels. In summary, we believe that there has been real progress in each of the three core subjects, but less than is indicated by the key stage results. We do not think that the tests have become systematically easier;[5] rather, we believe that teaching and learning has focused more and more narrowly on achieving test results.

C.6 Professor Peter Tymms pulled together much of the available evidence on standards in primary schools in his paper *Are standards rising in English primary schools?* (Tymms 2004). This is the area that has seen the greatest increase in measured national standards, but these improvements have not been sustained at secondary school as the same cohorts of pupils take their Key Stage 3 tests. Using 11 independent measures of English (reading) and mathematics in primary schools over the period 1995–2003, Tymms finds that during the period 1995–2000 the dramatic increase in the national measure (that is, the Key Stage 2 results) appear to be overstated. The proportion attaining the target level in English at Key Stage 2 rose from 49–75% over the period 1995–2000, and this equates to two standardised points a year (this is the same in mathematics). The data from the six independent sources and the corresponding Key Stage 3 (matched to the relevant cohort of pupils) results do not corroborate such a striking rise, showing only an average rise of 0.77 points per year for reading and 1.54 for mathematics (Tymms 2004).

C.7 The evidence on standards over time is complex. Massey *et al* (2003) investigated whether the difficulty of the tests has changed over time. They found that at least half the national gains in English at Key Stage 2 from 1999–2002 were attributable to changes in standards of the tests. A smaller study by Green *et al* (2003) found that there had been improvements in writing over the period 1995–2002 and that marking had remained consistent between these years (Green *et al* 2003). While it is possible that there has been some inconsistency in test difficulty, it seems unlikely this is a major or systematic factor.

C.8 Teacher assessments have been collated alongside the Key Stage test results since 1995. Despite the Government officially viewing teacher assessment as "an essential part of the national curriculum assessment and reporting arrangements" (TeacherNet 2006: 3), interest in teacher assessments and consequently their inclusion in "league tables" has declined.

[5] There is, however, evidence that standards in some subjects have varied over time. For example, Massey et al (2003) found variation in standards in Key Stage 2 English between 1996 and 1999–2000, Key Stage 2 science between 1996 and 2001 and Key Stage 3 mathematics between 1996 and 2001.

C.9 Table 1 compares the test results from Key Stages 1, 2 and 3 with the teacher assessments. Key Stage 1 teacher assessment figures closely matched the test results, which is not surprising given the close relationship between the two forms of assessment at this key stage (Richards 2005). However, with later phases teacher assessments provide an interesting contrast to the key stage tests. Key Stage 2 results show an increase in 20 percentage points from 1995–2006 in the proportion obtaining a Level 4 or higher in English compared to 30 points in the external assessment (see Table 1). Mathematics demonstrates a similar trend. In essence, teacher assessments of pupil performance started slightly higher but have improved more slowly than the national tests. Key Stage 3 teacher assessments replicate the same pattern with tests reporting more improvement than the teacher assessments.

C.10 Table 1 Pupils achieving the target level (see Note 2) in Key Stage tests and teacher assessment

		Reading		*Writing*		*Mathematics*	
		Test	*Teacher assessment*	*Test*	*Teacher assessment*	*Test*	*Teacher assessment*
Key Stage 1	1995	78%	79%	80%	77%	79%	78%
	2004	85%	85%	82%	83%	90%	89%
	Point difference	***7***	***6***	***2***	***6***	***11***	***11***
		English				*Mathematics*	
		Test	*Teacher assessment*			*Test*	*Teacher assessment*
Key Stage 2	1995	49%	57%			45%	54%
	2006	79%	77%			76%	78%
	Point difference	***30***	***20***			***31***	***24***
		English				*Mathematics*	
		Test	*Teacher assessment*			*Test*	*Teacher assessment*
Key Stage 3	1995	55%	63%			58%	62%
	2005	74%	71%			74%	75%
	Point difference	***19***	***8***			***16***	***13***

Notes:
1. The 2006 data is provisional
2. The target or "expected" level at Key Stage 1 is Level 2, at Key Stage 2 is Level 4 and at Key Stage 3 is Level 5.
3. The data include all eligible pupils in maintained schools and in independent schools that opted to take part in the National Curriculum assessments.
4. For Key Stage 1 2004 figures are used, as the assessment method changed for the 2005 assessments and therefore the figures from 2005 onwards are not directly comparable to those prior to that year. For more details on the new arrangements see Shorrocks-Taylor *et al* (2004).
5. For Key Stage 3 2005 figures are used, as the 2006 figures based on teacher assessments have not yet been published (as at 11 December 2006) due to inconsistencies in the data.

Source: Teacher assessment data (DfES unpublished note 2006), test data see DfES (2006)

C.11 An international survey looking at mathematics and science attainment, "Trends in International Mathematics and Science Study" (TIMSS), shows some evidence from Year 5 pupils to support raising levels of overall achievement between 1995 and 2003 in mathematics and science. Question items common to both rounds of tests allow comparisons to be made between the two years 1995 and 2003 and in both subjects the percentage answering these questions correctly increased (by 9 percentage points in mathematics and by 4 in science) (Ruddock *et al* 2004). Although the two are not directly comparable, improvements in TIMSS are thus much less impressive than the measured improvements in Key Stage test results.

C.12 The Statistics Commission considered these issues in 2005 and concluded that:

C.13 "'The Commission believes that it has been established that (a) the improvement in Key Stage 2 test scores between 1995 and 2000 substantially overstates the improvement in standards in English primary schools over that period, but (b) there was nevertheless some rise in standards." (Statistics Commission 2005: 4)

C.14 Looking at the secondary phase, the percentages of pupils attaining the benchmark at Key Stage 3 and Key Stage 4 have continued to rise although progress on international attainment measures has stalled. Evidence from TIMSS for Key Stage 3 (Year 9) does not show any significant change in performance between 1995 and 2003 (Ruddock *et al* 2004). Analysis of the international study PISA (Programme for

International Student Assessment)[6] shows that for a given score at Key Stage 3 or Key Stage 4, pupils attained on average a higher PISA score in 2000 than in 2003 (Micklewright and Schnepf 2006).[7] One possible explanation for this is that the standards measured by PISA have changed between 2000 and 2003. Another is that the Key Stage 3 and Key Stage 4 test scores are not consistent over the period. Our preferred explanation is that improvements in the key stage results do not accurately mirror improvements in underlying pupil attainment, and that some of the improvement is due to more narrowly focused teaching.

C.15 Does the current system of assessment test the necessary range of skills and abilities? While the core subjects of English, mathematics and science are extremely important, there is growing evidence that young people need to develop a much wider range of skills such as the ability to work with others, to listen and present effectively, to reflect critically, to stay "on task" for extended periods of time and to act responsibly. These have sometimes been described as "soft skills" (non-cognitive skills), but recent work by ippr indicates that these are essential skills for life, that they have been growing in importance, and that there is now a significant social class gap in their attainment (Margo *et al* 2006). Some studies show that noncognitive skills (personal and social skills and personality attributes) are as important as cognitive abilities (such as reading and mathematics ability at age 10) in determining earnings later in life, and analysis of the 1958 and 1970 cohorts indicates that non-cognitive skills became significantly more important over this period (Blanden *et al* 2006).

C.16 While teachers often try to develop these skills in their pupils, it is not at all clear that they are effectively specified in the curriculum and assessed by the current system of national tests. The danger is that the current assessment system thus fails to consider some crucial aspects of a young person's development. We should try to create space for this in a reformed system, but we want to do so in a way that maintains school accountability.

C.17 Assessment for accountability

Schools are now held much more strongly accountable for the outcomes achieved by their pupils, and their attainment at the end of the key stages in particular. One of the mechanisms for this is parental choice of school, and we discuss this further in another paper (Brooks and Tough 2006). In addition, the results of national tests are a critical input into Ofsted inspections, and a bad inspection may result in a school being issued a notice to improve, or risk being placed in special measures. Entering special measures means that a school loses its autonomy and represents a severe criticism of the leadership of the school. Failure to emerge from special measures rapidly enough can result in a school being closed entirely. School leaders thus face very clear incentives to ensure that their results do not prejudice their inspection results.

C.18 It is quite right that there should be a robust inspection mechanism to provide schools with powerful incentives to improve, and especially to ensure that no school falls below a minimum acceptable standard. However, if test results are to play an important role in such a powerful incentive mechanism, it is all the more important that they are robust, valid, and do not negatively impact on other desirable aspects of the learning environment. This particular issue—that preparation for tests might be crowding out other desirable activities in schools—is deal with in the next section of this paper.

C.19 Test results are more reliable at the school level than for individuals, because in a large group of pupils individual misclassifications tend to cancel each other out. However, the problem of validity is equally acute at the school level as it is for individual pupils. Schools are held accountable for their test results. The evidence set out in the previous section of this paper suggests that test results are overstating real improvements in young people's abilities. The danger is thus that we are holding schools accountable for the wrong thing. Another danger is that because non-cognitive skills are not recognised by the assessment system, schools have little incentive to focus on their development.

C.20 It is important not to overstate these arguments. Ofsted inspections do take into consideration a wide range of factors in addition to test results. Even if there is a serious question about the validity of the tests, a school that is achieving poor test results, given its intake, is unlikely to be successfully developing the skills those tests are meant to be measuring. However, it is certainly the case that schools do have strong incentives to focus on the results of the tests we currently have. If there are problems with the validity of those assessments, there will be a problem with accountability. What is needed is not less accountability, but more valid and reliable tests.

[6] PISA looks at attainment in reading, mathematics and science literacy across participating countries (mainly OECD countries) every three years. the analysis described here (Micklewright and Schnepf 2006) compares reading and science literacy in 2000 and 2003. Mathematics is not included as the content areas used for measuring mathematics ability were different between 2000 and 2003.

[7] Research by the National Foundation for Educational Research (NFER) for DfES compared the familiarity and appropriateness of PISA and TIMSS for English pupils who take Key Stage 3 tests and GCSE examinations. For Science the familiarity suggested that 40–50% of pupils would be familiar with the PISA and TIMSS tests, whereas for mathematics TIMSS had a higher familiarity rating of 65–85% compared to 50–70% for PISa (ruddock *et al* 2006). The focus of PISA is on literacy and this is reflected in the PISA tests, which require much more reading than TIMSS, Key Stage 3 or GCSE. The PISA tests are also more focused on applying knowledge and are more heavily contextualised (*ibid*).

C.21 Assessment for learning

How widespread is good practice in assessment for learning in England? Ofsted finds that the use of assessment for learning is only good or better in a minority of secondary schools and unsatisfactory in a quarter (Ofsted 2005a) and that it is the least successful element of teaching in primary schools (Ofsted 2005b). It concludes that schools often do not seem to have the capacity, training or urgency to process information properly for use within the classroom. Despite efforts by DfES to spread good practice, assessment for learning needs to be given a higher priority in both initial teacher training and continued professional development. Responding to the 2006 General Teaching Council's Annual Survey of Teachers[8], 42.6% of teachers reported that they "will need" professional development in Assessment for Learning over the next 12 months (Hutching *et al* 2006).

C.22 One factor that is often cited as a barrier to effective teaching is an over-prescriptive and over-crowded curriculum. Nearly one third of teachers feel there is "little" or "no" flexibility in the curriculum (Hutching *et al* 2006). In fact, the national curriculum is much less restrictive than is often claimed, and schools also have the right to apply to the Secretary of State for exemptions from curriculum requirements.

C.23 Does the current system of national tests act as a barrier to the adoption of assessment for learning? A key argument of opponents of the current assessment system is that too much teaching time is taken up with non-productive preparation for tests, and that this is squeezing out other more valuable activities. A first question to ask is whether it is possible that some forms of preparation for tests might represent a bad use of classroom time, or on the other hand whether everything that improves test results is useful.

C.24 In fact, there does seem to be a range of activities that are intuitively of low educational value that nonetheless might improve test results:

a. Narrow learning. Because all pupils take the same test, and because each test covers a relatively large subject area, it is possible to predict with reasonable accuracy what will be tested—the most prominent elements of the curriculum. This enables teachers to concentrate on those aspects of the curriculum that are most likely to come up in the tests.

b. Shallow learning. Because all pupils take the same kind of test, it is possible to predict with reasonable accuracy how each component of the curriculum will be tested. This enables teachers to focus on this approach.

c. Question spotting. This is essentially follows from the previous two problems.

d. Risk-averse teaching with low levels of innovation.

C.25 It is very difficult to be sure of the extent and impact of practices such as shallow and narrow teaching and learning, and even more difficult to prove a causal link between their prevalence and the nature of the assessment system. However, there is evidence that teachers' own assessments become less formative and more summative in response to high-stakes testing. The Primary Assessment, Curriculum and Experience (PACE) project, a longitudinal study that followed a cohort of primary school pupils for eight years starting before the introduction of national tests for seven-year-olds, found that after the introduction of the tests teachers' own classroom assessment became more summative (Pollard *et al* 2000).

C.26 It is also important to emphasise that we do not have to choose between doing assessment for learning and assessment for summative purposes. It is important to remember that the key objective of assessment for learning is improved child development. There should thus in theory be no tension between formative assessment and evaluative assessment, because the first should be a means to improvement in the second. To some extent what is needed is a jump from one equilibrium, in which teachers are too busy training their pupils to pass national tests to focus on assessment for learning, to a better one in which teachers make more use of assessment for learning and as a result their pupils perform better in national tests.

D. *Recommendations*

D.1 We believe that it is possible to change the assessment system so that it facilitates better teaching and learning, provides a better measure of child attainment, and maintains a high level of school and national accountability. In summary, our proposals are:

a. Every child should be assessed throughout each key stage by their teachers.

b. Every child should sit a small number of national tests at the end of each key stage, but not in every area of every subject. The results of these monitoring tests should be used to measure overall school performance, but not the individual pupil's attainment.

c. The school's performance should be used to moderate the teacher assessments, producing detailed, nationally-comparable data for each pupil.

[8] A random sample of 10,000 teachers was drawn from a sample pool of 430,722 eligible teachers registered with the General Teaching Council, that is, those who were in service in state schools in England in September 2005. In total 3,665 completed questionnaires were received, a response rate of 37%. The achieved sample was compared with the population in terms of key variables, and while there were minor differences between the sample and the population, these were small enough not to affect representativeness, so that generalisation from the sample to the population could be made with confidence.

D.2 Every child should be assessed throughout each key stage by their teachers

D.3 Short tests lead to unreliable results, and written tests can only assess certain kinds of ability. We should therefore use a wider range of assessment methods, for a broader range of skills, over a longer period. All this suggests that evidence of the level of a pupil's ability should be gathered over the length of their study rather than in an hour-long examination, which further implies a bigger role for teacher assessment.

D.4 It should be possible in large part to use formative assessment for summative purposes. The TGAT Report (Task Group on Assessment and Testing) concluded in 1987 that formative assessment could "meet all the needs of national assessment at ages before 16" (V.26). TGAT recommended that formative assessment should be the basis of national assessment at seven, 11 and 14 and that assessment should only be designed for summative purposes at the end of compulsory schooling when information is required for certification. For the seven, 11 and 14 age groups, key elements of their scheme included:

a. A significant role for teacher assessment. This would require teachers to rate pupils according to criteria set out for each level in the national curriculum.
b. Emphasis on assessing a full range of skills. A range of standardised, cross-curricular assessment tasks would ensure that those skills not easily measurable through written tests were nonetheless assessed.

D.5 Such a system would represent a major challenge to the teaching profession, and would be demanding to implement. In 1993, teachers boycotted standardised assessment tasks, complaining that the practical classroom observations and tasks and complex "criterion reference" marking were too burdensome. There were, for example, 17 attainment targets for science and 15 for maths on which teachers had to pass judgments for each child (Green 2006). If teacher assessment is to be a success then it will need to be designed and implemented in a way that has the support of teachers and is not excessively burdensome to operate, and it will require significant investment in professional training and development.

D.6 What can be offered to teachers in return for the challenges of a new assessment system is the opportunity for better, more appropriate and more effective exercise of their professional skills. The evaluation of the trial exploring a move to reporting-only teacher assessment at Key Stage 1 (which was rolled out nationally in 2004) reported that many teachers saw opportunities for reducing their workload as well as for developing their professionalism (Shorrocks-Taylor *et al* 2004). The report also found that parents generally preferred their child not to be tested but noted that parents still wanted information regarding schools' test performance (*ibid*). Research shows that 70% of head teachers (sample 375) and 76% of Year 2 teachers (sample 306) felt that the new assessment arrangements at Key Stage 1 had a positive effect on teaching and learning in the classroom (Reed and Lewis 2005).

D.7 Another major worry about teacher assessment is that it will be unreliable because teachers will not accurately grade their pupils. Part of this concern may be that there would be unintended bias in teacher assessment. The experience of shifting to teacher assessment at Key Stage 1 does not support this hypothesis—national results went down after its introduction in 2004. Part of the concern may also be that teachers will adjust their expectations and therefore their grades in the light of their school's or class's circumstances. All judgments are relative and therefore teacher judgments on individual pupils may be affected by the range of abilities in the group (Laming 2004). If this happened then grades would be inflated in schools with low average attainment, and depressed in schools with high average attainment. A system of monitoring testing and moderation would ensure this would not happen.

D.8 Every child should sit a small number of national tests at the end of each key stage, but not in every area of every subject. The results of these monitoring tests should be used to measure overall school performance, but not the individual pupil's attainment

D.9 National testing performs two vital functions. First, it provides an independent and external measure of school performance, and second it provides a measure of overall national progress. However, it may be possible to fulfil both of these functions and at the same time reduce the burden of such tests on individual pupils and on the schools system as a whole. The key is to stop using the national tests to measure individual pupil attainment.

D.10 For national tests to measure individual pupil performance on a consistent basis, each pupil must sit the same tests and must be tested in every subject. However, if the tests are being used only to measure school and overall national performance, it may be possible for each pupil to sit tests in just some of their subjects, and for different pupils to sit different tests in the same subject. The tests could thus in aggregate cover much more of each subject, and it would become much more difficult for teachers (and pupils) to predict their content. This would make it much harder to "teach to the test", and even if there would still be some scope to narrow the focus of learning due to the inherent limitations of timed tests, it would become necessary to teach to the curriculum more broadly.

D.11 These tests could be used in the same way as the current Key Stage tests to hold schools to account for their pupils' performance. Schools would thus still have a strong incentive to achieve good results, even though these would not directly determine individual pupils' scores. Shifting towards a system of monitoring tests would be independent of any decision about the publication of school performance information. Once school performance data is collected it can be collated and presented in table format, so even if DfES stopped

publishing performance tables it is likely that it would continue to be produced. Monitoring test data could be in the form of absolute performance, value added, or contextualised value-added measures. At the school level very much the same data would continue to be available as it is now. Accountability would if anything be improved because the results would become a more valid reflection of true pupil abilities across a wider curriculum. At the national level the test results would give a more valid and reliable picture of pupil attainment, because they would cover the curriculum in a much broader and deeper way than is possible where every student takes the same exam.

D.12 In the current system each pupil is tested on each subject area, namely English, mathematics and science. In the proposed system all that would be necessary is a monitoring test that gives a statistically reliable guide to the overall performance of the school. The required number of tests for each pupil would thus vary with the size of the school. For a large school this might significantly reduce the amount of national tests that each pupil would have to sit compared to the status quo. For a very small school it might not be possible to reduce the number very much. However, it is important to remember that even the current extensive system of tests does not provide a reliable guide to the quality of small schools, whose results can fluctuate significantly from one year to the next simply due to the small number of students being tested. Another concern might be that although schools have strong incentives to achieve good results, pupils do not, and they might therefore not put any effort into the tests. This may or may not be a significant issue: arguments can be made in theory either way, and more research and evaluation will be required in this area.

D.13 The school's performance should be used to moderate the teacher assessments, producing detailed, nationally comparable data for each pupil

D.14 The final piece of the puzzle is to use the monitoring tests to moderate teacher assessments. A system of unmoderated teacher assessment would be unlikely to command public confidence, as it would be possible for different teachers to be allocating different grades to similar pupils.

D.15 Part of the response to this problem should be to try to improve the quality of teacher assessment for both formative and evaluative purposes, both in initial teacher training and in continued professional development. This is likely to be an important part of any major shift in this direction. The evaluation of the Key Stage 1 trial found that "accuracy and teacher confidence in making Teacher Assessments is strongly affected by the quality of training and moderation" (Shorrocks-Taylor et al 2004: 4). Assessment for learning should certainly be given significantly more profile in initial teacher training. At present, formative assessment is not mentioned explicitly in the professional standards for teachers. The standards are currently under review and the draft revised standards for Qualified Teacher Status (the standards that need to be reached to become a qualified teacher) do include specific reference to being "informed of . . . the importance of formative assessment" as well as to "know how to use local and national statistical information to evaluate the effectiveness of their teaching, to monitor the progress of those they teach and to raise levels of attainment" (TDA 2006, Q11: 10).

D.16 It would also be possible to develop of a cadre of specialist teacher assessors, to encourage the exchange of teachers involved in assessment between different schools, or to develop professional external moderators to assist schools. One option to raise the status of teachers and their ability to undertake accurate assessment would be to have at least one qualified assessor in each school. This idea was first put forward by the Association of School and College Leaders (ASCL) (and then the Secondary Heads Association), which described a vision of a chartered examiner in every large department in secondary school.

D.17 There are many advantages to such a model. It could help to restore trust in teacher assessment, and it would also provide a route similar to the "excellent teacher" or "advanced skills teacher": a route of progress for experienced teachers who do not want to take the leadership route. There would be increased costs involved as these assessors would command a higher salary as well as more staff time dedicated to preparing for and moderating assessment. However, as with the costs associated with the increased training, these would be balanced with the reduced financial burden of the external examinations bill. Recent research commissioned by the QCA indicates the total cost of the examinations and testing system in England to have been £610 million in 2003–04 (QCA/PwC 2004). A more accurate reflection of the costs of National Curriculum Testing would be £346 million as this removes the direct costs of the three awarding bodies for post-14 qualifications (Edexcel, AQA and OCR)[9].

D.18 However, the best guarantee of comparable results and thus high levels of public confidence would be, in addition to any other measures, to use the monitoring test results to adjust the teacher-assessed pupil scores. We have not worked through the details of such a moderation scheme, and there would undoubtedly be complexities. However, the schematic picture is as follows: the monitoring tests independently indicate the distribution of attainment in the school in each of the core subjects. This information can then be used to scale the teacher assessments for individual pupils so that they fit into the known school-level distribution.
E.1

[9] Although this figure still includes the costs of administering GCSEs, A Levels, and so on, for exam centres.

E. *Conclusions*

E.1 A system of assessment such as the one outlined here would require extensive research and development, piloting and phased introduction alongside a major programme of teacher training and the creation of new systems of teacher assessment and national testing. We do not claim to have a model, but have presented here the broad features of such a system. It would represent a revolution for English education, but could potentially meet each of our objectives better than the existing system of assessment:

a. It should be more reliable, because teacher assessments could be based on much more information than can be captured in a single round of tests.
b. It should be more valid at the national level as there would be data on a wider range of subject areas.
c. It should have greater validity at the pupil level, because teacher assessments could more successfully measure the different aspects of each pupil's progress.
d. It should thus provide a better measure of pupil attainment, beyond pen and paper tests.
e. The monitoring tests should maintain accountability at the school level, and should provide a better measure of national progress on standards.
f. It should facilitate assessment for learning, both because teacher assessments of individual pupils could be built up from formative assessment results, and because it would make it much more difficult to "teach to the tests" and should thus promote teaching the whole curriculum in the most effective way.
g. Teachers would have more time to focus on the development of other important skills such as noncognitive skills.

F. *References*

Note: web references correct at December 2006

Black P and Wiliam D (1998) *Inside the Black Box: Raising Standards Through Classroom Assessment* London: King's College London

Black P and Wiliam D (1998a) "Assessment and classroom learning" *Assessment in Education* 5 (1) 7–74

Black P and Wiliam D (2006) "The Reliability of Assessments" in Gardner J (ed) (2006) *Assessment and Learning* London: Sage

Blanden J, Gregg P and Macmillan L (2006) *Accounting for Intergenerational Income Persistence: Non-Cognitive Skills, Ability and Education* London: Centre for the Economics of Education, London School of Economics, reported in Margo J and Dixon M with Pearce N and Reed H (2006) *Freedom's Orphan's: Raising youth in a changing world* London: Institute for Public Policy Research

Brooks R and Tough S (2006) *Pupil Attainment: Time for a three Rs guarantee* London: Institute for Public Policy Research. Available at: www.ippr.org/publicationsandreports

Brooks R and Tough S (2007) *School Admissions: Fair choice for parents and pupils* London: Institute for Public Policy Research. Available at: www.ippr.org/publicationsandreports/publication.asp?id = 546

Daugherty R (1995) *National Curriculum Assessment: A Review of Policy 1987–1994* London: Falmer Press

Department for Education and Skills (DfES) (2006a) "National Curriculum Assessments at Key Stage 1 in England, 2006 (Provisional)" *Statistical First Release* 30/2006 24 August

DfES (2006b) "National Curriculum Assessments at Key Stage 2 in England, 2006 (Provisional)" *Statistical First Release* 31/2006 24 August

DfES (2006c) "National Curriculum Assessments at Key Stage 3 in England, 2006 (Provisional)" *Statistical First Release* 34/2006 13 September

Green S, Johnson M, O'Donovan N and Sutton P (2003) "Changes in Key Stage Two Writing From 1995 to 2002" A paper presented at the United Kingdom Reading Association Conference at University of Cambridge, 11–13 July

Green S (2006) in Reed J and Tough S (eds) (2006) *Curriculum, Assessment and Pedagogy: Beyond the "standards agenda"* London: Institute for Public Policy Research

Hutchings M, Smart S, James K and Williams K (2006) *General Teaching Council for England Survey of Teachers 2006* London: Institute for Policy Studies in Education, London Metropolitan University. Available at: www.gtce.org.uk/research/tsurvey/tsurvey06/

Laming D (2004) *Human Judgement: The Eye of the Beholder* London: Thomson

Margo J and Dixon M with N Pearce and H Reed (2006) *Freedom's Orphan's: Raising youth in a changing world* London: Institute for Public Policy Research

Massey A, Green S, Dexter T and Hamnett L (2003) *Comparability of national tests over time: Key stage test standards between 1996 and 2001: Final report to the QCA of the Comparability Over Time Project* Cambridge: University of Cambridge Local Examinations Syndicate. Available at: www.qca.org.uk/6301.html

Micklewright J and Schnepf S V (2006) "Response Bias in England In PISA 2000 and 2003" *DfES Research Report* RR771. Available at: www.dfes.gov.uk/research/

Miliband D (2003) "School Improvement And Performance Management" A speech by David Miliband to Perfomance Management Conference, Bournemouth, 27 January. Available at: www.dfes.gov.uk/speeches/search_detail.cfm?ID = 59

Ofsted (2005a) *The Secondary National Strategy: An evaluation of the fifth year* Reference No. HMI 2612. Available at: www.ofsted.gov.uk/assets/4118.pdf

Ofsted (2005b) *Primary National Strategy: An evaluation of its impact on primary schools 2004/2005* Reference No. HMI 2396. Available at: www.ofsted.gov.uk/assets/4117.pdf

Ofsted (2006) *Evaluating mathematics provision for 14–19 year olds* Reference No. HMI 2611. Available at: www.ofsted.gov.uk/assets/4207.pdf

Pollard A and Triggs P with Broadfoot P, McNess E and Osborne M (2000) *What pupils say: changing policy and practice in primary education—findings from the PACE project* London and New York: Continuum

QCA/PwC (2004) *Financial Modelling of the English Exams System 2003–04* London: QCA. Available at: www.qca.org.uk/12130.html

Reed M and Lewis K (2005) *Key Stage 1 Evaluation of New Assessment Arrangements* London: NAA. Available at: www.qca.org.uk/downloads/pdf_05_18931.pdf

Richards C (2005) *Standards in English Primary Schools: are they rising?: A contribution to the debate from the Association of Teachers and Lecturers* London: Association of Teachers and Lecturers. Available at: www.askatl.org.uk/atl_en/resources/publications/research/Primarystandards.asp

Ruddock G, Sturman L, Schagen I, Styles B, Gnaldi M and Vappula H (2004) *Where England stands in the trends in international mathematics and science study (TIMSS) 2003: Summary of national report for England* Slough: NFER. Available at: www.nfer.ac.uk/publications/other-publications/downloadablereports/where-england-stands-in-the-trends-in-international-mathematics-and-science-study-timss-2003-national-report-for-england.cfm

Ruddock G, Clausen-May T, Purple C and Ager R (2006) "Validation Study of the PISA 2002, PISA 2003 and TIMSS-2003 International Studies of Pupil Attainment" *DfES Research Report* RR772. Available at: www.dfes.gov.uk/research/

Shorrocks-Taylor D, Swinnerton B, Ensaff H, Hargreaves M, Homer M, Pell G, Pool P and Threlfall J (2004) *Evaluation of the trial assessment arrangements for key stage 1: Report to QCA* London: QCA. Available at: www.qca.org.uk/8994.html

Statistics Commission (2005) "Measuring Standards in English Primary Schools: Report by the Statistics Commission on an article by Peter Tymms" *Research Report* 23 London: Statistics Commission. Available at: www.statscom.org.uk/C_402.aspx

Training and Development Agency (TDA) (2006) *Draft revised standards for teachers*. Available at: http://www.tda.gov.uk/teachers/professionalstandards.aspx

TeacherNet (2006) National curriculum assessment arrangements. Available at: www.teachernet.gov.uk/management/atoz/n/nationalcurriculumassessmentarrangements/

Tymms P (2004) "Are standards rising in English primary schools?" *British Educational Research Journal* 30:4 477–494

Wiliam D (2000a) "The meanings and consequences of educational assessments" *Critical Quarterly* 42(1)

Wiliam D (2000b) "Reliability, validity and all that jazz" *Education* 29 (3)

About ippr

The Institute for Public Policy Research (ippr) is the UK's leading progressive think tank. Its role is to bridge the political divide between the social democratic and liberal traditions, the intellectual divide between academia and the policymaking establishment and the cultural divide between government and civil society. It is first and foremost a research institute, aiming to provide innovative and credible policy solutions. Its work, the questions its research poses, and the methods it uses are driven by the belief that the journey to a good society is one that places social justice, democratic participation, economic and environmental sustainability at its core.

About the Authors

Richard Brooks is an Associate Director at ippr and leads the institute's public services work. He is a frequent contributor in both the broadcast and print media, and has published widely on public services and welfare policy issues. Prior to his current role he was Research Director at the Fabian Society, and he has also worked in the Prime Minister's Strategy Unit and the Labour Party Policy Unit. Between 2002 and 2006 he was a Tower Hamlets Councillor, and was Cabinet Member for Resources for three of these years. He is a governor of Bethnal Green Technology College.

Sarah Tough is a Research Assistant in the Public Services Team at ippr. Before joining ippr Sarah worked as a mathematics teacher in a London secondary school as part of the Teach First programme. Sarah has a first class honours degree in Economics from the University of Bristol. She is a governor of a Hackney primary school.

June 2007

Memorandum submitted by the National Association of Schoolmasters Union of Women Teachers (NASUWT)

Executive Summary

The Importance of a National Framework for Testing and Assessment

The need to sustain and develop a shared set of understandings about learning necessitates the maintenance and development of a nationally managed and regulated system of assessment. Such a system secures the common entitlement of all learners to effective and meaningful assessment and accreditation of their learning and ensures that Government policy on testing and assessment complements other elements of national education strategy.

Workforce Remodelling and Action to Tackle Teacher Workload

The remodelling of the school workforce and action to tackle teacher and headteacher workload are at the heart of the Government's strategy to raise standards and to ensure that all learners can gain access to the high-quality learning experiences to which they are entitled. Approaches to assessment that add to the workload and bureaucratic burdens of teachers and headteachers and ascribe responsibilities to teachers that do not make the best possible use of their professional skills, talents and expertise will work to reduce the quality of pupils' learning experiences and will undermine efforts to raise standards of educational achievement.

The Impact on Teaching and Learning of the Current High-stakes Testing and Assessment Regime

In relation to the current arrangements for end of Key Stage statutory assessment and testing, the Committee will be aware from its previous inquiries in this area of the ongoing debate about the implications of this system for teaching and learning at Key Stages 1, 2 and 3. The NASUWT acknowledges that much attention is focused on the apparent difficulties associated with the continued use of centrally set and externally marked tests at the end of Key Stages 2 and 3 and notes that the DfES faces continued pressure from a significant number of stakeholders and other agencies and organisations for the discontinuation of tests and their replacement by a system of moderated teacher assessment.

The NASUWT reasserts to the Committee its longstanding view that such views are based on an incomplete appreciation of the broader context within which National Curriculum tests are undertaken and the implications of this on the current system for statutory end of key stage assessment. The NASUWT believes that National Curriculum tests are, of themselves, generally unproblematic in terms of their impact.

The Use of Performance Data to Support Pupil Progress

It is essential that the use of performance data to inform policy and practice is moderated by an understanding of the context within which such data is generated and the extent to which test data is able to capture the full range of educational progress and development achieved by pupils.

Appropriate Use of Teacher Assessment, Assessment for Learning and Coursework

School-level arrangements for formative assessment, assessment for learning and coursework can become problematic when schools feel under pressure to implement approaches that are unnecessarily bureaucratic and time-consuming as a result of a perceived need to demonstrate to Ofsted inspection teams or other external auditors of school performance that practice in relation to assessment is effective.

Effective Public Administration and Management of the Assessment and Testing System

The Government, in collaboration with the NASUWT and other social partners, should undertake an assessment of the extent to which current national-level organisational arrangements for the management and administration of the testing and assessment system, including the provision of general qualifications, secures overall policy coherence and delivers effective use of finite public resources.

Introduction

1. The NASUWT welcomes the opportunity to submit evidence to the House of Commons Education and Skills Select Committee Inquiry into assessment and testing.

2. The NASUWT is the largest union representing teachers and headteachers throughout the UK.

3. The NASUWT's response sets out the Union's perspective on the key issues identified by the Committee in the terms of reference for the inquiry and highlights aspects of the current assessment and testing regime that the NASUWT believes needs to be reviewed in order to ensure that future policy in this area builds upon existing good practice and addresses less effective features of the present system.

The Importance of a National Framework for Testing and Assessment

4. Any effective national system of testing and assessment must secure three important objectives:

- Reliable—assessment information and data must give an accurate indication of the level of skills, knowledge and understanding attained by learners during a course of study;
- Valid—an assessment system must ensure that the nature of the learning subject to assessment is readily identifiable from associated information and data; and
- Comparable—assessment information and data should allow comparisons to be made between different levels and forms of learning.

5. The absence of any of these essential elements of assessment would prejudice the effective and coherent accreditation and recognition of learning on a system-wide basis. The ability of learners, parents, employers, teachers and other relevant stakeholders to interpret assessed learning using a shared set of terms and descriptions of the type of learning undertaken and the standards learners have attained, on both an individual learner and aggregated basis, is essential if meaningful dialogue and exchange about learning is to be made possible.

6. The need to sustain and develop a shared set of understandings about learning necessitates the maintenance and development of a nationally managed and regulated system of assessment. Analysis of comparable education systems, both within the United Kingdom and elsewhere, demonstrates that while features of the assessment system may take different forms, the existence of a system-wide approach to the formal recognition and accreditation of learning is a common feature.

7. However, it is important to acknowledge that a national framework for the assessment of learning also secures other key policy objectives. Of particular importance is the right of learners across the education system to have access to a minimum set of entitlements in respect of the way in which their learning will be assessed. In this respect, the ability of the Government to set out the basis upon which these entitlements will be delivered in practice is an important means by which equity and consistency within the education system can continue to be secured.

8. The legitimate rights and expectations of learners in relation to their employment and lifelong learning opportunities also underscore the importance of national frameworks for the accreditation and recognition of learning. In an economic and social context where the migration of labour between nation states is of increasing significance and scale, it is essential that effective means are in place to ensure the international transportability of qualifications. The work being undertaken by the European Commission on the European Qualifications Framework depends critically on the existence of consistent national examination and assessment systems against which qualifications originating in other countries can be compared. The maintenance of an effective national qualifications system therefore enables learners to access their labour mobility rights as EU citizens and supports the economic and social life of the UK by facilitating the inward migration of qualified workers.

Workforce Remodelling and Action to Tackle Teacher Workload

9. The ability of the Government to establish and develop a national framework to recognise and accredit learning has important implications for the effective implementation of policies affecting the employment and professional status of teachers, headteachers and other members of the wider school workforce.

10. The implications of the contractual changes introduced by the National Agreement *"Raising Standards and Tackling Workload"*, and the wider workforce remodelling agenda are particularly important in this respect. The aim of the Department for Education and Skills (DfES), the NASUWT and other social partners in taking forward these important reforms is not only to ensure that effective action can be taken to drive down teacher workload but also to ensure that work to raise standards can be sustained through the creation of an environment that actively supports the ability of teachers and headteachers to concentrate on their core responsibilities for teaching and learning and leading and managing teaching and learning.

11. The remodelling of the school workforce is at the heart of the Government's strategy to raise standards and to ensure that all learners can gain access to the high-quality learning experiences to which they are entitled. If the assessment system is to play an effective part in the achievement of the aims of this strategy, it must operate in a way that is consistent with the changes being introduced by the workforce

remodelling agenda. Approaches to assessment that add to the workload and bureaucratic burdens of teachers and headteachers and ascribe responsibilities to teachers that do not make the best possible use of their professional skills, talents and expertise will work to reduce the quality of the learning experience of pupils and will undermine efforts to raise standards of educational achievement.

12. In this context, the existence of a national framework for assessment and testing is essential as it provides the means by which a necessarily diverse and complex national education policy agenda can be taken forward coherently. While the need for an appropriate level of local and school-level flexibility in relation to the implementation of aspects of the assessment system remains an important consideration, particularly in respect of efforts to embed effective approaches to the personalisation of learning, an overemphasis on local autonomy and control over assessment at the expense of the ability of central Government to manage and administer key elements of the national education system, including assessment and testing priorities, risks undermining the effective implementation of policies legitimately determined at national level.

The Impact on Teaching and Learning of the Current High-stakes Testing and Assessment Regime

13. In relation to the current arrangements for end of key stage statutory assessment and testing, the Committee will be aware from its previous inquiries in this area of the ongoing debate about the implications of this system for teaching and learning at Key Stages 1, 2 and 3. The NASUWT acknowledges that much attention is focused on the apparent difficulties associated with the continued use of centrally set and externally marked tests at the end of key stages 2 and 3 and notes that the DfES faces continued pressure from a significant number of stakeholders and other agencies and organisations for the discontinuation of tests and their replacement by a system of moderated teacher assessment.

14. The NASUWT reasserts to the Committee its longstanding view that such views are based on an incomplete appreciation of the broader context within which National Curriculum tests are undertaken and the implications of this on the current system for statutory end of key stage assessment.

15. The NASUWT believes that National Curriculum tests are, of themselves, generally unproblematic in terms of their impact. National level testing and sampling is a well-established feature of approaches to the assessment of the progress and attainment of individual pupils and cohorts and is the basis upon which studies such as those currently undertaken by the Organisation for Economic Co-operation and Development's (OECD) PISA programme are based and were also used by Her Majesty's Chief Inspector of Schools to inform the work of the former Assessment and Performance Unit in judging the overall effectiveness of the education system. It should also be noted that the use of tests remains a well-established and legitimate element of the assessment repertoire of teachers in determining effective approaches to the assessment of pupils' learning.

16. However, the purposes to which the performance data generated by National Curriculum testing are put remain the principal cause of the problems for teachers, headteachers and learners created by the current statutory arrangements for assessing pupil progress at the end of Key Stages 2 and 3. In particular, the publication of tables of pupil performance, with their attendant significant consequences for perceived failure in respect of pupil outcomes, has led to the creation of a destructive and high-stakes environment within which National Curriculum tests remain located. While the publication of tables of school-level performance data is intended to hold schools and teachers publicly accountable for the attainment of pupils in the core subjects of the National Curriculum, their practical effect is to contribute to a skewing of the curriculum, generate unacceptable levels of pressure and workload at school level and entrench a competitive rather than collaborative culture between schools. They are also responsible for many of the pressures that inhibit the ability of teachers to exercise an appropriate level of professional discretion and autonomy. Although this was recognised to an extent in *Excellence and Enjoyment*, the Government's Green Paper on teaching and learning in the primary sector, the reported experience of NASUWT members working with pupils at Key Stage 3 suggests that performance tables have a comparable impact in the secondary sector.

17. The use of tests to generate performance tables has seriously affected the potential value of tests as an educational tool and has given them a prominence that distorts the real contribution they can make to pupils' learning. The Committee is right to question to extent to which the apparent decline in rates of pupils' progress between Years 6 and 7 is explained by the high-stakes context within which pupil data for pupils at the end of Key Stage 2 is generated. It is for this reason that the NASUWT continues to have serious reservations about calls to replace end of key stage testing with a system of externally moderated teacher assessment. Proponents of this view often fail to appreciate that many of the problems associated with testing relate to the high-stakes environment within which end of key stage assessment takes place. The NASUWT maintains that there is a significant danger that such an approach would result only in the replacement of one high-stakes assessment system with another and generate significant increases in workload for teachers and headteachers with no commensurate educational benefit. It is for this reason the NASUWT undertook national action in 1994 at the launch of the tests at Key Stage 2 to secure external marking which the Dearing Review of the National Curriculum confirmed should be the approach adopted to statutory end of Key Stage assessment beyond Key Stage 1.

18. Experience of similar changes in Wales is instructive in this respect. Following changes to assessment arrangements in Wales, a survey of members of the NASUWT undertaken in 2006 indicted that 72% of respondents in the primary sector felt that these changes had lead to increases in teachers' working hours and workload overall.

19. In this respect, the consequences of the changes to the arrangements for statutory assessment at Key Stage 1, introduced in 2004–05, are particularly instructive. While the previous system of dual reporting of the results of centrally set tasks and teacher assessment were associated with higher levels of teacher workload than at Key Stages 2 and 3 as a result of the lack of external marking arrangements, the move to teacher assessment-only arrangements has placed significant additional demands on teachers as a result of the increased emphasis on the results of teacher assessment of pupils' work undertaken during the course of the year. The NASUWT's survey of schools involved in statutory assessment at Key Stage 1 found that over 60% of teachers had experienced an increase in workload as a result of requirement to engage in more intensive moderation of pupils' work as part of the revised arrangements. It is likely that workload increases would be even more pronounced by the introduction of comparable arrangements at Key Stages 2 and 3 given the relatively higher stakes generated by the publication of performance tables for pupils at ages 11 and 14.

20. It should be recognised that similar difficulties are generated by the publication of tables of pupils' performance in respect of general qualifications. While it is inevitable that the pressures faced by pupils and teachers in this respect will be influenced by the implications for the future education and employment choices of candidates of the results of general qualifications, the incorporation of GCSE, A-level and equivalent qualifications into performance tables exacerbates needlessly the stress and anxiety faced by pupils, teachers and headteachers.

21. The NASUWT has no objection to an effective and development-focused approach to school accountability. However, the continued use of performance tables in England for this purpose works to undermine rather than support the efforts of teachers, headteachers and members of the wider school workforce in sustaining rates of pupil progress. There is little evidence that performance tables have contributed to raising standards of attainment. A growing number of international studies show that other comparable education systems, including those in Wales, Scotland and Northern Ireland, have reached and maintained high educational standards without use of the performance tables. The NASUWT believes it essential that the DfES, in collaboration with the NASUWT and other social partners, commence work to examine effective and viable alternative approaches to school accountability and that an investigation of the operation of such systems in other education systems would be an effective basis upon which the development of a more progressive and supportive system of accountability and school improvement could be based.

22. However, the NASUWT recognises that positive reforms to the current testing and assessment regime are being examined. Most significantly, in response to the themes set out in the Report of the Teaching and Learning in 2020 Review Group, which sought to identify effective approaches to the personalisation of learning, the Government has set out its intention as part of its *Making Good Progress* proposals to trial approaches to National Curriculum testing where pupils are entered for assessment on the basis of teachers' professional judgements about their rate of progress rather than their chronological age.

23. The commitment of the Government to retain external marking as a central element of reformed testing arrangements and to give greater emphasis to the skills and expertise of teachers in national testing arrangements is welcome and highlights other opportunities for reform. In particular, the NASUWT contends that the division of the National Curriculum into fixed key stages to mark points at the end of Years 6 and 9 at which pupils' test results are reported in performance tables contributes significantly to inconsistency across the National Curriculum as a whole. In light of the proposals to pilot testing on the basis of pupils' stage of development rather than chronological age, there is a case for examining the potential advantages of removing key stages from the National Curriculum and investigating the extent to which this would allow for the development of more consistent and coherent approaches to the assessment of pupil progress.

The Use of Performance Data to Support Pupil Progress

24. Notwithstanding the NASUWT's concerns about the context within which National Curriculum assessment is undertaken, the Union is not opposed to the use of national tests as a means of providing data in respect of pupil performance. The data provided by such tests can yield valuable evidence to inform the development of national education policy and can act as a powerful tool to support the work of teachers and headteachers in meeting pupils' learning needs.

25. Nevertheless, it is essential that the use of performance data to inform policy and practice is moderated by an understanding of the context within which such data is generated and the extent to which test data is able to capture the full range of educational progress and development achieved by pupils.

26. For this reason, the NASUWT continues to have concerns about potential overemphasis on performance data in analyses of pupil progress where insufficient account is taken of the impact of the school accountability regime on the context within which such data is generated. This feature of end of statutory

National Curriculum assessment highlights the need for the DfES, the NASUWT and other social partners to develop a shared understanding of the potential impact of the school accountability system on Key Stage 2 and 3 performance data. The NASUWT takes the view that it would be timely to use the opportunity created by the *Making Good Progress* pilots to select a relatively small number of schools where National Curriculum testing would be carried out in line with arrangements in other pilot schools but where the results of the tests would not be reported in DfES performance tables. This would provide an opportunity to examine the impact of aspects of the school accountability regime on the approaches adopted by schools to National Curriculum assessment and the effects of the accountability regime on support for pupil progress.

27. The NASUWT recognises the concerns expressed by the Government and other stakeholders about the rates of progress of specific groups of pupils, particularly in respect of pupils with special educational needs, in receipt of free school meals or with patterns of poor attendance. However, the Union has reservations about the Government's proposed use of existing performance data to establish expectations in respect of acceptable rates of pupil progress as part of its *Making Good Progress* proposals. In particular, the expectation that pupils should be able to demonstrate progress of least two National Curriculum levels within each key stage is based on an assumption that the rate of progress required to move between levels is consistent across the National Curriculum as a whole and that level descriptors in the National Curriculum are consistent between different key stages.

28. There are important and longstanding concerns about the credibility of this assumption in the light of teachers' experience of assessing pupils in Key Stages 2 and 3. While these concerns do not undermine the importance of the emphasis in *Making Good Progress* that work must continue to ensure that all pupils can reach the standards of which they are capable, they serve to highlight the importance of exploring approaches to the development of effective progression measures that avoid the potential difficulties associated with the use of National Curriculum level descriptors for this purpose.

29. In light of the increased levels of interest in the use of performance data, generated to a large extent by the proposals set out in *Making Good Progress*, to support the work of schools and local authorities in continuing to work to raise standards, it would be timely for the Government, in collaboration with the NASUWT and other social partners, to examine in more detail the ways in which data can be used effectively and in an appropriate context.

Effective Use of Teacher Assessment, Assessment for Learning and Coursework

30. The NASUWT shares the view of the Government, set out most recently in its *Making Good Progress* proposals, that formative assessment is an integral feature of effective teaching practice. Formative assessment and assessment for learning exemplify the benefits for children of teachers being given the scope to make appropriate use of their professional judgement to shape approaches to curriculum content and assessment to meet learning needs.

31. However, school-level arrangements for formative assessment can become problematic when schools feel under pressure to implement assessment for learning in ways that are unnecessarily bureaucratic and time consuming as a result of a perceived need to demonstrate to Ofsted inspection teams or other external auditors of school performance that effective approaches to formative assessment have been put in place. Assessment for learning conducted in this way is of limited use to parents or pupils in identifying and working towards future learning goals and distracts teachers from activities directly related to supporting pupils to make progress.

32. The NASUWT believes that the pilots of the *Making Good Progress* proposals provide an opportunity for the DfES to work with the NASUWT and other social partners to develop approaches to assessment for learning that do not lead to inappropriate burdens being placed on teachers but which serve professional assessment needs and support the work of teachers in engaging pupils more effectively in evaluating their own progress and areas for future development.

33. Teacher assessment also plays a significant role in the general qualifications system in respect of coursework requirements in subject specifications. The NASUWT believes that assessment procedures for all general qualifications should be designed so that candidates' learning can be assessed in a valid, reliable and manageable way. An essential feature of effective assessment systems in this respect is the extent to which they minimise bureaucracy and workload for teachers, students and other members of the school and college workforce, and candidates.

34. The NASUWT remains concerned that the way in which coursework is used to assess learning in many GCSE and AS/A-level specifications fails to meet these key criteria. This view is also reflected in the findings of the Qualification and Curriculum Authority's (QCA) review of GCE and GCSE coursework published in November 2005.

35. For GCSE examinations in particular, the weighting of coursework, which is not lower than 20% in any of the specifications of the 10 most popular subjects, has placed an emphasis on coursework that is problematic in many key respects. While the general qualifications system must assess learning in a way that is valid, reliable and comparable, the NASUWT agrees with the QCA review's finding that the pressures on teachers to ensure that these objectives are met in relation to coursework are significant.

36. The difficulties associated with plagiarism highlighted by the QCA review are a particular cause for concern in this respect. The confidence of teachers, candidates, employers and higher and further education institutions in coursework as an effective and reliable method of assessment is undermined by the considerable problems that plagiarism creates. The NASUWT is concerned that attempts to tackle plagiarism through the adoption of increasingly elaborate and time-consuming anti-plagiarism systems will not only add to the bureaucratic burdens on school and college staff but will also fail to resolve concerns about the fitness for purpose of coursework as an assessment technique. Regardless of the sophistication of the systems employed to counter plagiarism, the fact that coursework is open to significant abuse by learners and their parents suggests that, where possible, the use of coursework should be limited to learning that cannot be assessed effectively in any other way.

37. The analysis by the QCA review of the areas of learning assessed by coursework in the most popular subjects at GCSE demonstrates that coursework is being used inappropriately as a method of assessment in an excessive number of specifications. It is imperative that the QCA accepts the recommendation in the review that the syllabus and specifications of all GCSE and GCE examinations should be re-assessed to ensure that coursework is only used where it can be shown to be the only effective method of assessing learning and the NASUWT welcomes work that is ongoing in this respect. However, the Union has increasing concerns that where it is felt that coursework is no longer appropriate, it is being replaced by "controlled assessments" that, while resembling examinations in many respects, are not described as such by awarding bodies. Not only do such assessments fail to tackle workload burdens effectively, given the continuing requirement on teachers to mark candidates' work, they also invite pressure to be placed on teachers to be present during such assessments where this would contradict their contractual right not to invigilate examinations. It is essential that effective action is taken to ensure that controlled assessments are only used where they are necessary to ensure effective assessment and that it is made clear that teachers are not required to be present during such assessments for the purposes of invigilation.

38. Limiting the use of coursework to learning that cannot be assessed by alternative means would not only serve to support public confidence in the examinations system but would also reduce the burdens that teachers and centres face as a result of requirements associated with the assessment of coursework.

39. However, given the fact that an objective review of the effectiveness of coursework would be almost certain to recommend a reduction in its use, it is clear that in some areas of learning in some subjects an element of coursework will need to be retained where the ability of students to generate pieces of work over an extended period needs to be assessed. It is also clear that coursework can be of particular value for assessing and accrediting the achievements of a significant number of pupils with special educational needs. In reviewing the use of coursework in general qualifications, it is important that the important steps being taken to tackle workload and to remodel the work of teachers so that they are more able to concentrate on their core responsibilities for teaching and learning are recognised appropriately.

40. A clear example of concerns raised on a consistent basis by members of the NASUWT in this respect, including those involved in teaching GCSE mathematics, involves the requirement in some awarding body specifications for teachers to annotate candidates' work in unnecessary detail. The NASUWT is concerned not only by the variation in practice between awarding bodies in setting out requirements for annotation but also by the significant impact that this has on teacher workload. It is difficult to sustain the position that excessive annotation is necessary to ensure the validity and reliability of coursework in circumstances where some awarding bodies are able to accept a significantly lower level of annotation than others.

41. The degree of variation in practice between awarding bodies, exemplified by different requirements in relation to the annotation of coursework, is unacceptable and unnecessary. NASUWT believes that, as part of an overarching review of coursework, it is necessary for more robust minimum standards of practice to be established that ensure that all awarding bodies implement specifications that contribute positively and consistently to the embedding of remodelled approaches to professional practice within schools and colleges.

42. Similar concerns about the use of teacher assessment in relation to the extended project element of the 14–19 specialised Diplomas can also be identified. The proposed degree of learner collaboration and topic scope inherent in the extended project could present teachers with significant difficulties in ensuring that individual candidiates' learning is assessed in a way that is sufficiently valid, reliable and comparable. It is likely that the moderation demands on teachers in particular will be considerable and that, as a result, the benefits for learners and teachers of the reforms to coursework requirements as part of existing general qualifications could be undermined by bureaucratic and workload-intensive arrangements that will be required to make the extended project as proposed currently a viable and credible element of the diplomas. As part of ongoing going work to review the implementation of the reforms set out in the 14–19 White Paper, the Government should work with the NASUWT and other social partners to ensure that the extended project is embedded within the specialised Diplomas in a way that will not undermine work to drive down teacher and headteacher workload and school-level bureaucracy.

EFFECTIVE PUBLIC ADMINISTRATION AND MANAGEMENT OF THE ASSESSMENT AND TESTING SYSTEM

43. At present, executive management and administration of the curriculum and assessment system in England is delegated by the Government to the QCA and the National Assessment Agency (NAA) while general qualifications are provided currently by three competing awarding bodies.

44. These arrangements contrast to a significant extent with practice elsewhere in the United Kingdom. In Wales, the Welsh Assembly Government (WAG) has taken back into central control the curriculum and assessment responsibilities previously devolved to the Qualifications, Curriculum and Assessment Authority for Wales (ACCAC) in order to ensure greater coherence between the development and execution of policy in these areas. In respect of the role of awarding bodies, the system in Scotland operates on the basis of the Scottish Qualifications Authority (SQA) acting as the sole awarding body for Standard Grade and Higher Grade qualifications.

45. These distinctions between policy and practice in England and elsewhere in the United Kingdom invite reflection on the extent to which current arrangements would benefit from reform. The NASUWT believes that ensuring coherence in the development and implementation of education policy and making the most effective use possible of necessarily finite public resources must be core elements of an effective national strategy for education. In relation to the operation of the testing and assessment system, it would be timely for the Government, in collaboration with the NASUWT and other social partners, to consider the extent to which the experience and expertise currently located within the QCA and the awarding bodies might be best deployed within an organisational context that brings together the development and execution of policy under the direct authority of the DfES and examines possible alternatives to the provision of general qualifications through several competing awarding bodies in receipt of significant indirect state subsidies.

46. The need to ensure that effective arrangements are in place to secure the coherent and accountable management and administration of testing and assessment policy is highlighted by proposals set out by the Leitch Review of Skills in its report, *Prosperity for all in the Global Economy—World Class Skills*, to devolve significant decision making authority about the range and nature of qualifications eligible for public funding to the Sector Skills Councils (SSCs). It is entirely inconsistent with the establishment and maintenance of a co-ordinated and effective qualifications system for this essential function to be undertaken by an unelected set of employer-dominated quangos with no meaningful accountability to elected ministers, Parliament or other legitimate stakeholders. It will be essential for the Government in its response to the Leitch Report, due in the Summer of 2007, to set out clearly that such an approach to qualifications policy has no place in the development of coherent qualifications policy and to reject this proposal.

June 2007

Memorandum submitted by National Foundation for Educational Research (NFER)

EXECUTIVE SUMMARY

Purposes of Assessment

In essence, educational assessment serves two major purposes. The first of these is to provide immediate feedback to teachers and students, in order to facilitate the learning process. This is often termed formative assessment. Recently it is also frequently referred to as *Assessment for Learning*. The second major purpose is to provide information which summarises a particular phase of learning or education. This information may be for a variety of institutional uses such as monitoring progress, certification or selection, and it may be for a variety of users, such as parents, teachers, governmental bodies and the learners themselves. This type of purpose is termed summative assessment. Both these purposes are important in the educational system as a whole, but they have different requirements and different characteristics.

Formative assessment is vital to teaching and learning. It is embedded in effective classroom practice and it should be tailored to the individual and their own stage of learning. Such processes are essential for progress, providing a motivational effect for students as well as information on what has been recently achieved and the next steps in order to make progress. We are very supportive of the principles of *Assessment for Learning* and believe these must be further promoted. There is also a need for more, and more rigorous, research which explores the successful elements of *Assessment for Learning*.

However, we do not believe *Assessment for Learning* can or should provide summative information. Instead, we believe that this function should be a largely separate system with its own priorities, features and requirements.

Formal summative assessment can serve many purposes. This paper concentrates on one of these: summative assessment within schooling, principally through National Curriculum Assessment. This has had a consistent structure for about a decade, but there is currently renewed discussion on its purposes and methods. This has culminated in the department for Education and Skills' Consultation document *Making Good Progress* which proposes shorter, better focused "when ready" tests.

National Curriculum Assessment

In commenting on testing in the National Curriculum, it is apparent that it is now a complex system, which has developed many different purposes over the years and now meets each to a greater or lesser extent. It is a tenet of current government policy that accountability is a necessary part of publicly provided systems. We accept that accountability must be available within the education system and that the assessment system should provide it. However, the levels of accountability and the information to be provided are open to considerable variation of opinion. It is often the view taken of these issues which determines the nature of the assessment system advocated, rather than the technical quality of the assessments themselves.

Our review of National Curriculum Assessment shows that there are many purposes, and any calls for change needs to consider which are the most important and which can be downgraded. In the existing system, the current National Curriculum tests are a compromise which attempts to meet all these functions. The accountability role means that they must achieve high levels of reliability and validity. As one of the developers of National Curriculum tests, we are aware of the thorough development process they undergo and the underlying statistical data on their performance. In our view, the current tests achieve the necessary technical and psychometric requirements to a reasonable extent. Any development of the existing system and its tests for which the accountability purposes remain, would properly need to demonstrate that it has equivalent or higher reliability. We do not believe it would be defensible to introduce a system in which levels of reliability are not known or cannot be demonstrated.

We therefore believe that there should not be changes to the existing system without careful consideration of what the purposes of the system are and a statement of this. Any proposals for change should set out carefully which of the purposes they are attempting to meet and which they are not. The level of requirements for validity and reliability could then be elucidated and the balance with manageability and the resources required determined. If accountability is no longer to be required then a different assessment regime could be implemented. However, this should not be done without evidence that any replacement would meet its own purposes validly.

Our stance in relation to assessment is that there must be a clear statement of the intended purposes of the assessment system and that its processes and instruments should have an appropriate level of validity and reliability to provide sound evidence for those purposes. This implies that there should be a sound development process for instruments, and evaluative research to demonstrate that the judgements being reached on the basis of the system are soundly based.

National Monitoring

One of the current purposes of National Curriculum Assessment is the provision of central information on the education system as a whole, for monitoring standards over time and reporting on the curriculum in detail. There are difficulties in maintaining a constant standard for the award of a level in a high stakes system where tests or questions cannot be repeated. We do though believe that the methods used for this currently which include year on year equating and the use of a constant reference point through an unchanging "anchor test" are the best available. A second consideration is that the curriculum coverage each year is limited to the content of that year's tests. In response to these (and also other issues), there has been considerable advocacy of a light sampling model for monitoring the curriculum and changes in performance.

However, there are some problems with this approach, which should be recognised. The lightly-sampled low stakes assessment would provide one view of standards, but because it is low stakes it may well underestimate what students are really capable of when they are more motivated. The research literature shows that there is a large difference in scores on the same test in high and low stakes situations. This is a validity issue related to the view taken of standards. If we are interested in monitoring what pupils can achieve when not motivated to achieve, low stakes surveys would be acceptable. We also believe that the practical difficulties of conducting such surveys voluntarily have been underestimated.

Nevertheless, we would support the introduction of a properly planned regular national monitoring exercise, to examine changes in performance at regular intervals, on a sample basis, and to monitor the curriculum widely. To assess the full curriculum in a valid manner may well require assessment methods other than written tests (eg for speaking and listening, or science experimentation). Such surveys would need to be regarded as a proper research exercises with the collection of background data on pupils and schools, in order to examine educational and social questions. They should also ensure a wide agreement on the appropriateness of its methodology and analysis techniques, reducing the possibility of attacks on the results.

Making Good Progress Proposals

The *Making Good Progress* proposals range widely over assessment, personalised learning and target setting. These should properly be regarded as a whole. However, since this paper deals with assessment issues, we will concentrate on that part of the proposals.

In general, we are supportive of the notions of testing when ready and the close tie to teaching and learning. The concept of testing when ready can be a useful one, particularly if it is used formatively and incorporated into the teaching-learning process as in *Assessment for Learning*. As such, "progress tests" could provide a useful stimulus to teaching and learning. However, as described in *Making Good Progress*, we would doubt that they can fulfil that function. As a single level test, awarding a level, the test would generally show what a student could do but it would not be able at the same time to provide diagnostic information about the next steps since these would not be included in the test. Similarly, because it would have to cover the curriculum broadly at that level, and levels represent two years of teaching (on average), it could not identify the small next steps needed for personalised learning.

The argument advanced in *Making Good Progress* is that success at one level will stimulate progress toward the next level, acting motivationally. This will need to be evaluated in practice. It may be that the levels are so far apart (they are intended to cover two years of development) that achieving one may actually slow progress, since the next target may be too distant. This is particularly a concern because of the "one way ratchet" proposal. The achievement of a level and the knowledge that it cannot be removed may act to demotivate rather than motivate. We would advise that the "one way ratchet" is abandoned and that the system allows for re-testing of doubtful cases so that high levels of certainty are achieved and so that misclassification is minimised.

For these reasons, we do not believe that the tests as described could support teaching in any direct way. If this is the desired intention, a different model with a suite of short tests, relating to specific elements of the curriculum, and providing information both on what has been achieved and the next steps would be more appropriate. There would need to be a large bank of such tests available for testing when ready on an individual basis. To be most useful they would be marked by the teacher, immediately, rather than through an external system. Such tests would be low stakes and have little accountability function.

To summarise, we do believe that some version of Progress Tests may be a useful addition to the system, but believe that their purpose must be carefully defined. That purpose should then lead to a specification and a development process that produces tests which are fit for use in terms of their reliability and validity.

National Foundation for Educational Research (NFER)

The National Foundation for Educational Research (NFER) was founded in 1946, and is Britain's leading independent educational research institution. It is a charitable body undertaking research and development projects on issues of current interest in all sectors of education and training. The Foundation's mission is to gather, analyse and disseminate research based information with a view to improving education and training. Its membership includes all the local authorities in England and Wales, the main teachers' associations and a large number of other major organisations with educational interests, including examining bodies. It is overseen by a Board of Trustees and Council.

The NFER's Department for Research in Assessment and Measurement is one of two research departments of the Foundation. It specialises in test development and research into assessment-related questions. The work of the Department involves projects of importance to national educational policy and its implementation through research, the development of assessment instruments and the evaluation of assessment initiatives. It has a consistent track record of developing high quality assessment materials to meet the needs of a variety of sponsors. The Department's experience covers the whole range of tests and other assessments. NFER's work in assessment and surveys stretches back over its entire history, such that the Foundation has a unique experience of test development and the use of tests. In addition to developing assessments, we also carry out major evaluation studies, large scale surveys and international surveys for a number of sponsors including DfES, QCA, SEED and DELLS.

Experience in Assessment

The following list of projects illustrates the variety of experience in assessment matters:

National Assessment by Sampling the Cohort

NFER was responsible for the greater part of the work of the Assessment of Performance Unit (APU) in the UK. National monitoring of performance in mathematics, English and foreign language, in England, Wales and Northern Ireland, was undertaken by the Foundation from the late 1970s to the late 1980s, when testing all of the cohort replaced a sampling approach. as part of this work a range of assessments going beyond pencil and paper tests was involved.

National Assessment by Testing the Whole Cohort

Since 1989, the Foundation has undertaken much work in producing National Curriculum tests to be used by the whole cohort in England and / or Wales. Such work has encompassed English, mathematics and science for various ages: 7, 11 and 14 and has been undertaken under contract to QCA or its predecessors. This work has provided many insights into the relationships between reliability, validity and manageability. Each of these tests is taken by 600,000 students, and the results have high stakes for schools since they are published as part of the accountability of the education system.

UK Assessment in the International Context

The Foundation has had a long involvement with international assessment, and was a founder member of the International Association for the Evaluation of Educational Achievement (IEA), which was set up in the 1960s and organises international comparative studies of educational achievement. NFER has been responsible for managing the testing for all of the IEA surveys in which England has participated, including both TIMSS (Trends in International Mathematics and Science Surveys) and PIRLS (Progress in International Reading Literacy Survey).

In 2005, NFER became responsible for the OECD PISA (Programme For International Student Assessment) surveys in England, Wales and Northern Ireland for 2006, which will report this year and will also be undertaking the 2009 surveys in all four UK countries.

NFER Position Paper on Assessment (2007)

Introduction

This paper has been produced in order to inform some of the current debates on National Curriculum Assessment in England. The Education and Skills Committee of the House of Commons has announced an inquiry into Testing and Assessment. In part this will examine testing and assessment in primary and secondary education as a key issue. Currently, pupils in England take Key Stage tests at seven years-old, 11 years-old and 14 years-old in English, mathematics and science. This system has developed and evolved since its introduction in 1991. In January 2007 the Government announced that it would pilot several measures at Key Stages 2 and 3, including allowing pupils to sit national Curriculum assessments as soon as they are ready instead of waiting until the end of the key stage.

Our paper sets out the background to these debates, concentrating on the purposes of assessment, and the desirable characteristics which flow from these purposes. This leads to a statement of NFER's stance in relation to assessment. Finally, a commentary is given on two specific proposals for change currently under discussion: a national monitoring system based on a sampling approach; and the "Progress Tests" proposed in the DfES discussion paper *Making Good Progress*.

Dimensions of Assessment

In essence, educational assessment serves two major purposes. The first of these is to provide immediate feedback to teachers and students, in order to facilitate the learning process. This is often termed formative assessment, but may also be referred to as diagnostic assessment. Recently it is also frequently referred to as Assessment for Learning. The second major purpose is to provide information which summarises a particular phase of learning or education. This information may be for a variety of institutional uses such as monitoring progress, certification or selection, and it may be for a variety of users, such as parents, teachers, governmental bodies and the learners themselves. This type of purpose is termed summative assessment. Both these purposes are important in the educational system as a whole, but they have different requirements and different characteristics.

This dimension of purposes is only one categorisation. A different categorisation, which cuts across the assessment purpose, is between formal and informal processes of assessment. The distinction here is between, on the one hand formal processes such as exams, tests and other assessments in which students encounter the same tasks in controlled and regulated conditions and, on the other hand, those less formal activities that form part of on-going teaching and learning. This second group would encompass question and answer, teacher observations, group discussion and practical activities together with classroom and homework writing and assignments.

Using this two-fold classification (as shown in the table below), it can be seen that formal and informal assessments can each be used for both formative and summative purposes. The formal processes are often managed externally to the school, though they need not be, and the informal processes are often internal to the school though they may provide information which is reported externally. The four cells of the table can be used to discuss the role and requirements of assessment systems and instruments.

	Purposes	
Processes	Formative	Summative
Informal	*Questioning*	*Essays in uncontrolled conditions*
	Feedback	*Portfolios*
	Peer assessment	*Coursework*
	Self assessment	*NC teacher assessment*
Formal	*Analysis of tests, exams, essays*	*Tests*
	Target setting	*Exams*
		Essays in controlled conditions

Formative Assessment

Informal Formative Assessment

Formative assessment is vital to teaching and learning. It is embedded in effective classroom practice and it should be tailored to the individual and their own stage of learning. Such processes have long been regarded as essential for progress, providing a motivational effect for students as well as information on what has been recently achieved and the next steps in order to make progress.

Although such practices have always been intrinsic to successful teaching, recent research and policy has characterised them as a particular approach to assessment, leading to the principles that have been set out under the heading of "Assessment for Learning" and its characteristics are as follows:

- — sharing learning goals with the pupils;
- — helping pupils know and recognise the standards they must aim for;
- — providing feedback that helps pupils to know how to improve;
- — both teachers and pupils reviewing pupils' performance and progress;
- — pupils learning self-assessment techniques to discover areas they need to improve;
- — pupils helping each other to learn; and
- — including both motivation and self-esteem within effective assessment techniques.

Assessment for Learning is a simple idea which is far-reaching in its implications and quite difficult to put into practice. If teachers obtain information from assessment and use it to identify the next steps in learning, their teaching will be much more effective. Better still, if pupils are "let in on the secret", so that they, too, understand what the next steps are, they will be better motivated and more successful learners. However, putting this into practice well can make formidable demands on teachers in terms of their professional knowledge and skills.

We are very supportive of the principles of *Assessment for Learning* and believe these must be further promoted and new and better supportive materials must be produced and supplied to teachers. There is also a need for more, and more rigorous, research which explores the successful elements of *Assessment for Learning*. Our own work in this area has been concerned with providing helpful materials for teachers and in researching the possibilities of e-assessment in supporting *Assessment for Learning*.

There is a scarcity of good quality formative assessment materials to be used by teachers and students in classrooms. It seems to have been assumed that the very openness of formative assessment, and its devolving of responsibility to the student, renders such materials undesirable. In contrast, we believe that well-designed support materials can encourage the spread of formative assessment, and have undertaken projects to develop such materials, with the specific intention of fostering peer assessment in literacy for pupils. (See Twist, L. & Sainsbury, M. (2006) *Assessment for Learning: Literacy 10–11*. (London: nferNelson)).

It is often asserted that *Assessment for Learning* leads to greater gains in pupil's knowledge and understanding and these claims are impressive. We do though believe that there remains a need for more research evidence demonstrating what leads to such gains. There are limitations to *Assessment for Learning*, which arise from its classroom role. Because of its immediacy and the focus on what has just been learned and what is about to be learned, it is difficult to give information on the overall level of attainment or on the curriculum as a whole. The involvement of the teacher (and also the student as a self-assessor and other students as peer-assessors) introduces problems of reliability (and also bias) so that *Assessment for Learning* data is not necessarily good for comparing pupils.

A further difficulty is its detail. If it is to be used for summative purposes, then the essentially atomised data needed for *Assessment for Learning* alone needs collating in a systematic manner to allow an overall judgement which is reliable and comparable. This can be a time consuming task.

For all these reasons, we do not believe *Assessment for Learning* can or should provide summative information. Instead, we believe that this function should be a largely separate system with its own priorities, features and requirements.

Formal Formative Assessment

Formal assessments can also be used for formative purposes, whenever an analysis is made of performance in a formal test and the results used to plan teaching for classes, groups or individuals.

The national Key Stage tests over the last few years have been systematically analysed in order to draw out implications for teaching and learning, which have then been published on the QCA website; a large part of this work has been carried out by NFER teams. An investigation of patterns of performance over a large sample of pupils can provide indications for teachers of typical patterns of errors. This can aid overall curriculum planning, but does not, in itself, give formative information for particular individuals or groups.

An additional problem with this approach is its timing. Formative information of this kind is of most use when the teacher is at the beginning of a programme of study, whereas the national tests are taken at the end of the key stage. A change in the timing of the national tests would in itself introduce greater potential for formative value.

A major focus of NFER's current work is the formative use of assessment information gained by more formal means. We are researching the potential of e-assessment in low-stakes contexts and to support assessment for learning. It is clear that teachers are required to focus on the understanding and attainment of individual pupils in order to develop effective plans for personalised learning. This will involve the management of a great deal of assessment evidence for planning teaching, in the form of test data and information on progress through the ongoing curriculum. E-assessment can occupy a central role, first in gathering detailed information about the nature of individual pupils' understanding and attainment, and then in collating and analysing this data. Rather than supplanting the teacher's role in relation to the child, it could supplement it, reducing the marking and recording workload while increasing and easing the flow of genuinely useful information.

In order to explore this opportunity an NFER research project is currently testing some of these principles. Experimental prototype questions are being trialled with samples of pupils and a variety of exploratory statistical analyses are being undertaken. This work may give rise to a clearer understanding of how e-assessment can provide a sensitive and unobtrusive evidence base for classroom activities and informative progress records.

Summative Assessment

Informal Summative Assessment

Since teachers have, by the very process of teaching, a wealth of informal assessment information on each pupil, there is a strong incentive to find ways of summarising that information so that it serves a summative purpose. Ongoing informal assessment information covers pupils' performance across a range of contexts, and is thus potentially both more valid and more reliable than a single test result.

The National Curriculum assessment system recognises this by requiring teacher assessment judgements alongside test results. Although this is, and has always been, an intrinsic element of the system, it has tended to have been given less prominence than the test results. In the early 1990s, there were indications that the structured attainment targets and level descriptions were introducing a useful element of standardisation and a common language to teachers' informal assessments. This "community of practice" tended to decline around the beginning of the new century, however, because of the introduction of the national strategies, which had a strong focus on pedagogy but very little on informal assessment. However, over the last few years the balance has begun to change. The ideas of *Assessment for Learning* have been integrated into policy and with this has come a renewed interest in making use of informal assessment in more systematic and summative ways. Currently, the QCA initiatives on Assessing Pupils' Progress in secondary schools and Monitoring Children's Progress in the primary sector have reintroduced some of the original ideas and methods of the early National Curriculum, restructured in accordance with later thinking, technology and strategies. In Wales, the Key Stage tests have been replaced by a system of teacher assessment only, supported by publications in which standards are exemplified. NFER staff members have worked with DELLS[10] to develop optional assessment materials and exemplification to support summative teacher judgement.

In order to be used summatively, teachers' assessment information needs to be related to the standards which are provided by the National Curriculum level descriptions. However, the descriptions are broad and general, including many imprecise judgemental terms, so there is work to be done in reaching a shared interpretation of their meaning and application. This would involve a process of moderation between teachers, both within a school and between schools, which would require local leadership, possibly by a local authority adviser. Typically, the moderation process would involve discussion of specific pieces of pupils' work, chosen to represent the characteristics of a level, leading to agreement on the criteria to be applied. It would aim to result in an agreed, shared portfolio of exemplars. This process is professionally valuable but costly and extremely time-consuming.

[10] The Department for Education, Lifelong Learning, and Skills (DELLS) within the Welsh Assembly Government.

A further time-consuming and potentially unmanageable aspect of informal summative assessment is the collection of evidence to support a judgement for each pupil. The system can collapse under an avalanche of paperwork if this is not managed carefully. The provision of an e-portfolio for each pupil could help greatly in managing the storage of examples of work and access to these, but will do nothing to reduce the time necessary to select, store, label and annotate the examples.

There is currently a debate about how far this kind of summative information can be used instead of test results, as in Wales and like coursework in public examinations. On the one hand, it has strong advantages in terms of scope and teacher involvement. On the other, its manageability is in question and its reliability has not been demonstrated.

Our view is that such a system in England is conceivable, but distant. There are three conditions that must be fulfilled before it could be introduced successfully. Firstly, a major investment—comparable to the introduction of the national strategies—has to be made in professional development in order to bring about a shared understanding of criteria. This would be supported by published exemplification materials and could include the use of some formal tests (as is currently the case at Key Stage 1). Secondly, a part of this professional development would need to address teachers' and advisers' understanding of the nature and purposes of the four quadrants of assessment, as described in this paper. It is necessary to reach a point where teachers perceive high-stakes summative assessment as professionally useful and complementary to formative approaches before a system of sufficient robustness could be introduced. Rigorous piloting and evaluation would be necessary in order to demonstrate appropriate levels of reliability. Finally, the system would need an element of external monitoring and accountability that commanded public and professional confidence.

Formal Summative Assessment

Formal summative assessment can serve many purposes. Among these are certification of schooling (as with GCSE) and selection (as with A-levels for university entrance). We will not consider these purposes here but concentrate on summative assessment within schooling, principally through National Curriculum Assessment. This has had a consistent structure for about a decade, but there is currently renewed discussion on its purposes and methods. This has culminated in the department for Education and Skills' Consultation document *Making Good Progress* which proposes shorter, better focused "when ready" tests. This paper will give general observations on National Curriculum testing (for summative purposes) and specific comments on *Making Good Progress*.

In commenting on testing in the National Curriculum the purposes of summative information need to be set out. Here, we are taking them to be, as follows:

A. The provision of comparable reliable information for children and their parents on their current levels of attainment.

B. The provision of comparable reliable information for children and their parents on the progress being made.

C. The provision of individual and grouped information for teachers to inform them of national standards and expectations in their subjects and to assist them generally with teaching pupils in the future.

D. The provision of grouped information for school managers and governors to inform them of the quality of learning of their students (and by inference the quality of teaching with the school) through the study of progress of their classes.

E. The provision of grouped school information for the public, providing an accountability function and contributing to choice for parents.

F. The provision of grouped school information to accountability agencies, such as LAs and Ofsted, to contribute to their judgements and measure improvement and decline.

G. The provision of central information to government and others on the education system as a whole, for monitoring standards over time and reporting on the curriculum in detail.

These seven purposes move from individual information to grouped information. They also move from levels of personal accountability to system accountability. It is a tenet of current government policy that accountability is a necessary part of publicly provided systems. There is a broad consensus on this and we accept that accountability must be available within the education system and that the assessment system should provide it. However, the levels of accountability and the information to be provided are open to considerable variation of view. It is often the view taken of these issues which determines the nature of the assessment system advocated, rather than the technical quality of the assessments themselves.

It is worth remarking that in addition to the purposes set out above, National Curriculum tests have served other indirect but nevertheless important functions within the system.

H. For professional development of teachers, informing them of the nature of the National Curriculum and its interpretation. (This was particularly true of the early years of implementation, but continues to have a role. In some subjects, notably English, this has brought about a

community of practice among teachers such that their judgements are much more aligned and standardised than they were before at Key Stages 1, 2 and 3. It is not necessarily also the case in mathematics, where many teachers continue to prefer test outcomes.)

I. To introduce positive change into the emphasis of the curriculum as taught (the delivered curriculum)—sometimes called a "backwash" effect.[11] Examples of this have included mental mathematics, spelling at Key Stages 1 and 2, and science processes at Key Stage 2 and 3.

J. The accountability functions themselves contribute to a further indirect purpose for the assessment system, which has a political motivation: that of putting pressure on schools and teachers to maximise the attainment of pupils and students. The testing regime is intended to motivate students to perform to high standards, teachers to teach better and parents and school governors to raise the quality of schools. The underlying reason behind this is what is perceived as a stagnation in standards from the 1950s to 1980s at a time when educationalists alone were responsible for the curriculum and schooling. The rise of economic globalisation and the widespread belief that raising educational standards was vital to future economic survival, led to the accountability and pressure models of the current system. (Education of course, is not alone among public services in being subject to this sort of pressure.)

To these can be added some additional purposes which have arisen almost accidentally, but now have a useful function.

K. In recent years there has been an acknowledgement of the importance of using national test data for school self-evaluation and improvement, often in partnership with other agencies such as the School Improvement Partner. The provision of sophisticated indicators based on national testing data, such as DfES/Ofsted's Contextualised Value Added (CVA) measures or those provided by the Fischer Family Trust (FFT) has led to a significant improvement to schools' ability to evaluate their own performance. These indicators rely crucially on the current national testing system, and any replacement system proposed would need to offer equivalent or better measures if there is any desire not to lose the progress which has been made in this area.

L. The availability of comprehensive national data with attached to it detailed pupil information has provided a powerful tool for the evaluation of the impact of educational initiatives on attainment and performance. Examples include NFER's work on evaluating Excellence in Cities, the National Healthy School Standard, Playing for Success, and the Young Apprentices Programme. Such data provides an important instrument for informing educational policy.

This account of the purposes of National Curriculum Assessment shows that there are many of these, and any calls for change needs to consider which are the most important and which can be downgraded. In the existing system, the current National Curriculum tests are a compromise which attempts to meet all these purposes. The accountability functions mean that they must achieve high levels of reliability. This means that the results must be reliable and subject to a limited amount of error and misclassification. (It is important to recognise that all tests, indeed all judgement processes have some component of error—this includes examinations, interviews, teacher judgement, and legal processes.) Any development of the existing system and its tests for which the accountability purposes remain, would properly need to demonstrate that it has equivalent or higher reliability. We do not believe it would be defensible to have a system in which levels of reliability are not known or cannot be demonstrated.

As one of the developers of National Curriculum tests, we are aware of the thorough development process they undergo and the underlying statistical data on their performance. In our view, the current tests achieve the necessary technical and psychometric requirements to a reasonable extent. They have good to high levels of internal consistency (a measure of reliability) and parallel form reliability (the correlation between two tests). Some aspects are less reliable, such as the marking of writing, where there are many appeals / reviews. However, even here the levels of marker reliability are as high as those achieved in any other written tests where extended writing is judged by human (or computer) grades. The reliability of the writing tests could be increased but only by reducing their validity. This type of trade off is common in assessment systems with validity, reliability and manageability all in tension.

The present tests do provide as reliable a measurement of individuals as is possible in a limited amount of testing time. When results are aggregated over larger groups such as (reasonably large) classes or schools, the level of reliability is extremely high.

A second requirement of the National Curriculum tests (and all assessments) is that they should be valid for their purpose. According to current thinking[12], the validation of a test consists of a systematic investigation of the claims that are being made for it. In the case of National Curriculum tests, the claims are that the tests give an accurate and useful indication of students' English, science or mathematical attainment in terms of National Curriculum levels. The tests do have limited coverage of the total curriculum: the English tests omit Speaking and Listening, the science tests formally omit the attainment target dealing with scientific enquiry (though questions utilising aspects of this are included) and

[11] The term "backwash" is often used of the negative consequences of testing on the curriculum. The effects can though be either positive or negative.

[12] See, for example: American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.

mathematics formally omits using and applying mathematics. Outside of these the coverage of content is good. The fact that the tests change each year means that the content is varied and differing aspects occur each year. In general, the content validity of the tests can be regarded as reasonably good in relation to this coverage of the National Curriculum. However, a full validation has other aspects and these are seldom considered in relation to the National Curriculum tests, principally because of their numerous purposes. In general, the current tests adequately serve the accountability requirements, listed above as A to F. They may not meet the monitoring requirement (Purpose G) so well and we address that below.

We therefore believe that there should not be changes to the existing system without careful consideration of what the purposes of the system are and a statement of this. Any proposals for change should set out carefully which of the above purposes they are attempting to meet and which they are not. The level of requirements for validity and reliability could then be elucidated and the balance with manageability and the resources required determined. If accountability is no longer to be required then a different assessment regime could be implemented. However, this should not be done without evidence that any replacement would meet its own purposes validly.

NFER Assessment Philosophy

The NFER view of assessment is to acknowledge and embrace the variety of assessment purposes and processes that the discussion above has set out. Both broad purposes and both types of process have their place in the overall assessment enterprise. It is meaningless and unhelpful to dismiss summative assessment because it is not formative, or to dismiss informal assessment because it is not formal. Our work encompasses all four quadrants and it is important to recognise the distinctive features and requirements of each.

Correspondingly, the need is for education professionals and policymakers to develop the same kind of understanding. The classroom teacher, like the assessment researcher, is required to deal with all four quadrants. The best approach to this is to understand and accept the distinctions and relationships between them, and to give appropriate attention to each one. Similarly, policymakers, officials and teacher educators must recognise that teachers have this variety of assessment responsibilities and opportunities and give attention and respect to all of them.

Our stance in relation to assessment is that there must be a clear statement of the intended purposes of the assessment system and that its processes and instruments should have an appropriate level of validity and reliability to provide sound evidence for those purposes. This implies that there should be a sound development process for instruments, and evaluative research to demonstrate that the judgements being reached on the basis of the system are soundly based.

Specific Proposals for Change

National Monitoring

One of the current purposes of National Curriculum Assessment is the provision of central information on the education system as a whole, for monitoring standards over time and reporting on the curriculum in detail (purpose G). It is here that the present system may be less valid. First, there are difficulties in maintaining a constant standard for the award of a level in a high stakes system where tests or questions cannot be repeated. We do though believe that the methods used for this currently which include year on year equating and the use of a constant reference point through an unchanging "anchor test" are the best available and lead to the application of a consistent standard. A second consideration is that the curriculum coverage each year is limited to the content of that year's tests. In response to these (and also other issues), there has been considerable advocacy of a light sampling model for monitoring the curriculum and changes in performance.

National Curriculum Assessment currently has monitoring national performance as only one of its many purposes, and is probably not optimal for this, as is the case for most assessment systems which attempt to meet many purposes. NFER conducted a review of educational statistics across the UK for the Statistics Commission, which was included in their report on the subject.[13] They concluded that the current national monitoring system in England was sufficiently fit for purpose that an additional survey would not be cost-effective.

We believe that, in principle, if the sole goal of an assessment system is to derive comparable measures of national attainment at different time points, then a low-stakes, lightly-sampled survey is probably the best way of meeting this one aim. Low-stakes testing has the advantage that there is no incentive to "teach to the test", reducing the effects in schools. (Though, as we have seen a positive backwash effect is one of the current uses of National Curriculum tests.) Because of reduced or negligible security issues it is possible to repeat substantial numbers of items from survey to survey, thus enabling relatively reliable measures of change over time to be adduced. It may not be necessary to monitor national performance on a yearly basis, and in this case less frequent surveys would be possible. A well-stratified national sample should enable good

[13] See: http://www.nfer.ac.uk/publications/other-publications/downloadable-reports/pdf_docs/serfinal.PDF

estimates of the uncertainty in the national performance measures to be made. A matrix sampling design, in which different pupils take different combinations of test items, would enable a wide coverage of curriculum areas to be maintained while minimising the burden on individual pupils.

However, there are some problems with this approach, which should be recognised. The lightly-sampled low stakes assessment would provide one view of standards, but because it is low stakes it may well underestimate what students are really capable of when they are more motivated. Our experience and the research literature shows that there is a large difference in scores on the same test in high and low stakes situations. This is a validity issue related to the difference between performance in motivated and unmotivated conditions. If we are interested in monitoring what pupils can achieve when not under motivated to achieve, low stakes surveys are well and good. If we are interested in performance when the results matter, this approach would not give it. It would also mean that such survey results would not align with any high stakes measures that continue eg GCSE.

There is considerable opposition in schools to taking part in optional assessment exercises, particularly secondary schools. However, anything other than a very high school response rate would cast serious doubts on the results, due to non-response bias, but it would be hard to find suitable incentives for schools to take part. Problems with response rates in international studies such as PISA, TIMSS etc. illustrate this—considerable efforts have been put into the attempt to persuade enough schools to take part to achieve the sample response rate constraints. It would probably be necessary in the modern climate to make participation in the survey compulsory for the selected schools in order to assure proper representative national samples.

Nevertheless, we would support the introduction of a properly planned regular national monitoring exercise, to examine changes in performance at regular intervals, on a sample basis, and to monitor the curriculum widely. To assess the full curriculum in a valid manner may well require assessment methods other than written tests (eg for speaking and listening, science experimentation). Such methods were attempted in the Assessment of Performance Unit (APU) surveys in the 1980s, conducted by the NFER and others, but proved difficult and expensive to implement. The lessons of the experience of that monitoring exercise also need to be learned. It would need to be regarded as a proper research exercise with the collection of background data on pupils and schools, in order to examine educational and social questions. It should also ensure a wide agreement on the appropriateness of its methodology and analysis techniques, reducing the possibility of attacks on its results.

Making Good Progress Proposals

The *Making Good Progress* proposals range widely over assessment, personalised learning and target setting. These should properly be regarded as a whole. However, since this paper deals with assessment issues, we will concentrate on that part of the proposals. Within the *Making Good Progress* document, it is proposed that there should be a new type of tests. The features of these are described briefly, and appear to be as follows:

- Single level tests.
- Testing when ready—shorter more focused and more appropriate tests.
- Externally set and marked.
- "One way Ratchet"—never going back, only forward.

In general, we are supportive of the notions of testing when ready and the close tie to teaching and learning. This fits within the context of Personalised Learning/ Assessment for Learning. As such, "progress tests" could provide a useful stimulus to teaching and learning. However, as described in *Making Good Progress*, we would doubt that they can fulfil that function. As a single level test, awarding a level, the test would generally show what a student could do but it would not be able at the same time to provide diagnostic information about the next steps since these would not be included in the test. Similarly, because it would have to cover the curriculum broadly at that level, and levels represent two years of teaching (on average), it could not identify the small next steps needed for personalised learning.

For these reasons, we do not believe that the tests as described could support teaching in any direct way. If this is the desired intention, a different model with a suite of short tests, relating to specific elements of the curriculum, and providing information both on what has been achieved and the next steps would be more appropriate. There would need to be a large bank of such tests available for testing when ready on an individual basis. To be most useful they would be marked by the teacher, immediately, rather than through an external system. Such tests would be low stakes and have little accountability function.

In fact, *Making Good Progress* makes clear that the proposed progress tests would be used for accountability purposes, with the levels awarded being retained and reported. This means that the tests will need to have the characteristics of tests for accountability: high levels of reliability and validity. The following sections examine the proposals from this viewpoint.

The meaning of the phrase "single level tests" will need some exploration. In a sense, the existing tests (or tests in the same style) could be utilized as tests which simply give a pass/fail at a single level. Given their length and their coverage of the curriculum this would lead to results with a reasonable (and measurable) level of reliability. However, *Making Good Progress* states that the tests will be "shorter and more focused." There is a strong relationship between reliability and test length, so an unfortunate implication of this is that the tests will have lower levels of reliability and reduced curriculum coverage.

In this context, the important aspect of reliability is the consistency of the decisions made. It there were two progress tests at the same level, what would be the percentage of students classified the same way on both occasions? For the tests to be shown to be useful, this needs to be considerably above chance levels. In the current reading and writing tests at Key Stage 2, the degree of decision consistency for each level is at least 80% and for some levels is as high as 98%. The progress tests would have to match these levels of consistency. This would need careful examination during development, as reducing the length of test inevitably leads to lower levels of reliability.

A second aspect of the "shorter more focused" approach is curriculum coverage. In the current National Curriculum tests considerable efforts are made to include as wide a representation of the curriculum as is practically possible in a written test. This is essential for demonstrations of validity. Moreover, the annually changing tests mean that, over time the tests have even wider coverage. In writing for example, different text types/genres are sought from children each year and, within the test each year, two different tasks are required. Hence, reducing the length of the tests could also reduce the validity of the test.

The concept of testing when ready can be a useful one, particularly if it is used formatively and incorporated into the teaching-learning process as in *Assessment for Learning*. However, its utility within a summative system may not be as apparent. The provision of information from the "progress tests" (which are aimed at making judgements about a single level) is unlikely to have the diagnostic element useful for Assessment for Learning. The argument advanced in *Making Good Progress* is that success at one level will stimulate progress toward the next level, acting motivationally. This will need to be evaluated in practice. It may be that the levels are so far apart (they are intended to cover two years of development) that achieving one may actually slow progress, since the next target may be too distant. This is particularly a concern because of the "one way ratchet" proposal. The achievement of a level and the knowledge that it cannot be removed may act to demotivate rather than motivate.

We have further concerns about the "one way ratchet". Its underlying assumption seems to be that children's learning is an ordered progression and that movement is always forward. This is not in fact the case, and children can decline in terms of skills or knowledge. It is therefore useful to have later checks that a level previously achieved has been maintained. If this is not the case, we do not believe the "one way ratchet" should be implemented.

This issue may interact with that of the reliability of the test. If the decision consistency of the tests at a given level is low, then a large proportion of candidates could be misclassified as achieving the level when they should not. If this is coupled with the "one way ratchet", the misclassification would become enshrined, possibly being harmful to such children's progress as they would be being treated (and taught) as if they were at a higher level than was actually the case.

It is not the case that the levels of the National Curriculum are, in practice, as even and well ordered as the underlying model would suggest. In a given strand of a subject, the difficulty of the content may not increase in regular steps. Similarly, in different strands, the difficulty of the processes or skills at a given level may not be the same. It was this type of difficulty that led to the abandonment of the strong criterion referencing model of the early national curriculum assessment in the 1990s. This was replaced by a weak criterion referencing model in which content from various levels and across a broad range has been included in the National Curriculum tests, leading to the setting of an overall subject level (or within English, reading and writing levels). It also marked a return to a traditional psychometric principles and a mark-based scoring system.

There is a naïve view that questions can be written at a single level, derived from the level descriptors and these will have comparable difficulty. Taken to its extreme, it is sometimes thought that a single level test could be constructed by having material drawn from the level descriptor at that level. Candidates would then be expected to answer a set proportion of this correctly. This might be 50%, or more usually 80%, or sometimes all. There have been examples of such systems which have been constructed with these principles and in which the consequence has been very low pass rates. We therefore would advise that although the Progress Test may award a single level, they should have the following characteristics:

— Sufficiently long (in terms of numbers of questions and marks awarded) to have a good curriculum coverage, leading to good evidence of validity.

— Sufficiently long (in terms of numbers of questions and marks awarded) to have high levels of reliability, so that decision consistency is good and the number of misclassifications (particularly false positives) is small.

— Include content from the level below as well as the target level in order to elicit a range of outcomes, and also to allow some simple questions to give pupils confidence and to motivate them.

— Include content from the level above as well as the target level in order to elicit a range of outcomes, and to allow some formative information to be provided for next steps.

— For writing, continue to allow a range of levels to be demonstrated, through differentiation by outcome.

— Set the criterion for achieving the level through soundly based equating or judgemental processes, not through the application of strict algorithms which assume equal difficulty in questions and in tests.

In addition, we would advise that the "one way ratchet" is abandoned and that the system allows for re-testing of doubtful cases so that high levels of certainty are achieved and so that misclassification is minimised. A useful refinement would be to have a system in which there are three levels of outcome: level X awarded; level X not awarded; and a band of uncertainty in which a retest is advised in the following test round. Hence teachers could report only success which is assured to a high probability, requiring pupils with scores in a defined range of uncertainty to be retested.

To summarise, we do believe that some version of Progress Tests may be a useful addition to the system, but believe that their purpose must be carefully defined. That purpose should then lead to a specification and a development process that produces tests which are fit for use in terms of their reliability and validity.

June 2007

Memorandum submitted by the National Union of Teachers (NUT)

Introduction

1. The National Union of Teachers welcomes the Education and Skills Committee's decision to undertake an Inquiry into testing and assessment. The nature and purpose of assessment has been the subject of intense debate for nearly 20 years; a fact acknowledged by the DfES in its recent consultation document, *Making Good Progress*. Yet, despite the DfES's acknowledgement that, "over the nearly two decades of the National Curriculum and its assessment regime, the end-of-key tests have often stimulated controversy", the Government remains in denial about the continuing impact of high stakes National Curriculum tests.

2. Research evidence continues to conclude overwhelmingly that the current high stakes system of testing and assessment undermines children's learning. Other countries in the United Kingdom have acted on this evidence and only England remains with an end-of-key assessment system which is fundamentally flawed.

3. It is clear also, from the consultation conducted by the NUT on *Making Good Progress*, that teachers remain firmly opposed to the current National Curriculum testing arrangements and their use as accountability measures. The fact that successive governments have chosen to ignore the teaching profession's concerns about the impact on teaching and learning of the current National Curriculum testing arrangements is an indictment of Government attitudes to teachers' professional judgement.

4. This fissure between the teaching profession and the Government led, in 1993, to what the weekly magazine, *Education*, called, at that time, "the greatest act of civil disobedience in the history of education". The boycott of the arrangements ended but the massive gap between teacher attitudes to the National Curriculum assessment arrangements and Government policy has continued to this day, triggering resentment amongst the profession about the refusal of successive governments to recognise the continuing damage of high stakes tests to the curriculum and children's learning.

5. Indeed, if the gap between teacher professional opinion and Government policies were not stark enough, the fact that successive governments have chosen to ignore overwhelming research evidence is a fundamental failure of evidence informed policy-making.

6. The National Union of Teachers has been at the centre of seeking a coherent, reliable and valid alternative to the current arrangements. It has taken the position consistently that teachers must be at the centre of defining the nature and purpose of assessment.

7. The NUT has supported and contributed to the development of assessment for learning; including through its professional development programme. With other teacher organisations, it has sought an independent review of summative assessment. It has formulated rounded proposals which it believes can be adopted by Government, are supported by schools and which would secure the support of the wider public. These proposals are set out within this submission. For this reason, the National Union of Teachers requests the opportunity to give oral evidence to the Select Committee on this Inquiry.

Assessment and Evaluation

8. The Government in England has failed consistently to adopt a coherent approach to assessment. Current systems for evaluation, from individual pupils to the education service at a national level, are extraordinarily muddled. There is no clear rationale of why various systems of summative evaluation and accountability exist. Consequently, schools experience over-lapping forms of high stakes evaluation systems, including institutional profiles based on test results and Ofsted judgements, which are often in contradiction with each other. These over-lapping systems of accountability are made worse by Government national targets for test results and examination results and by the publication on an annual basis of school performance tables.

9. Recently, the Government asserted within its *Making Good Progress* consultation that the, "framework of tests, targets and performance tables have helped drive up standards in the past decade". There is no evidence that such a framework has achieved this objective. Indeed, the same document contains the DfES's view that, "The rate of progress . . . has slowed in the past few years". Indeed, the reality is that national targets based on test results have damaged the record of Government on education, giving the impression of failure, not success.

10. It is vital that the Government, under a new Prime Minister, initiates an independent review of its school accountability arrangements. Accountability for the effective functioning of the education service is a legitimate requirement of both local communities and Government. Parents have the right to expect fair and accurate systems of accountability. The accountability system in England is permeated, however, by a lack of trust. The Government's assertion, in its recent document, *Making Good Progress*, that, "most schools now regard an externally validated testing regime as an important accountability measure", is completely without basis in fact. Teacher initiative and creativity is undermined by uncertainties created by multiple and often conflicting lines of accountability.

11. Through an independent review, the Government must act to clarify the distinction between assessment for learning and evaluation for accountability purposes. Recently, the National Union of Teachers agreed with other teacher organisations a proposal to Government to conduct an independent review of summative assessment.

12. This proposal remains valid, but, even more importantly, the Government should clarify the nature of its legitimate need to evaluate how well the education service it funds is operating, separately from the use of assessment by teachers as part of their teaching.

Summative Assessment for the Purposes of School Accountability

13. The weight of research evidence against the use of summative assessment for the purposes of school accountability is overwhelming. Over the last decade, the NUT has conducted a range of studies focusing on the impact of high stakes testing for the purposes of accountability. The agreement that the NUT reached with the Conservative Government, in 1994, included a commitment to review the impact of National Curriculum testing. This was a recommendation which Ron Dearing developed, but work on this area was halted after the election of the Labour Government, in 1997. The Government chose not to develop the work carried out by the School Curriculum and Assessment Authority (SCAA) and subsequently the Qualifications and Curriculum Authority (QCA) on alternatives to the assessment and testing arrangements, although, rightly, the QCA has continued occasionally to offer a radical critique of the current assessment arrangements.

14. The Committee should reflect on the battery of uses for which test results are now used for the purposes of institutional accountability; uses which encourage crude and inaccurate summative judgements or create obsessional bureaucratic recording procedures. Some examples are set out below.

15. Ofsted now relies on contextual value-added test result data as the baseline measure for school evaluation despite the fact that the usefulness of CVA data is limited by each school's unique demography and level of pupil mobility. Level 5 plus test results are being used as the gatekeeper for entry to the Gifted and Talented scheme. Despite the National Assessment Agency's guidance, local authorities are placing pressure on early years settings to use National Curriculum numeric levels in preparation for National Curriculum test result benchmarks at Level 1. Local authorities are encouraging the use of commercial schemes for special educational needs assessment which sub-divide each one of the National Curriculum P levels at Level 1 into a further five sub-levels.

16. This focus on the use of test results for the purposes of high stakes accountability has damaging effects on young people and teachers alike. The NUT's study, *National Curriculum Tests* (2003), conducted by Dr S Neill, of the University of Warwick's Institute of Education, found that teachers felt strongly that testing narrowed the curriculum and distorted the education experience of children. They said that the excessive time, workload and stress for children was not justified by the accuracy of the test results on individuals. Teachers did not feel that the tests accurately affected school achievements.

17. The study found that experienced teachers and teachers of younger children stressed two negative factors; the effect of the tests in distorting the curriculum and educational experience available to children, and the tests' effects on their workload, which were largely due to their own efforts in supporting children through the tests.

18. While the NUT's study found that high stakes end-of-key testing continued to cause much greater concern in primary than in secondary schools and that the longstanding experience of testing and examinations at secondary level tended to lead to greater acceptance amongst teachers and parents, teachers of older children reported also that the testing period was more stressful for students than for children in primary schools.

19. The critical piece of evidence that the NUT believes that the Government has almost completely ignored is the research review conducted by the Government funded body, EPPI (Evidence for Policy and Practice Information) (2004), on the impact of repeated testing on pupils' motivation and learning. The review concluded that repeated testing and examination demotivated pupils and reduced their learning potential, as well as having a detrimental effect on educational outcomes. Other key findings included evidence which showed that teachers adapt their teaching style to train pupils to pass tests, even when pupils do not have an understanding of higher order thinking skills that tests are intended to measure and that National Curriculum tests lower the self-esteem of unconfident and low achieving pupils.

20. The Government appears to have ignored other key research. The Assessment Reform Group's (ARG) (2005) study of the role of teachers in assessment of learning came to key conclusions based on the exhaustive research reviews it has carried out in conjunction with EPPI. Crucially, the Assessment for Reform Group concluded that:

> *"It is likely that opportunities to use assessment to help learning and reduce the gap between higher and lower achieving children are being missed . . . Many schools give the impression of having implemented assessment for learning when, in reality, the changes in pedagogy that it requires has not taken place . . . This may happen, for example, when teachers feel constrained by external tests over which they have no control. As a result, they are unlikely to give pupils a greater role in directing their own learning, as is required in assessment for learning, in order to develop the capacity to continue learning throughout life. The nature of classroom assessment is (being) dictated by the tests".*

21. The NUT believes that this finding is crucial to the future of the development of personalised learning. Personalised learning is fundamentally dependent on teachers integrating assessment for learning in their teaching practice. If the evidence is that National Curriculum tests are a barrier to the development of personalised learning, then the Government has to choose between the two approaches; one cannot complement the other.

22. The Select Committee will be fully aware of the work conducted by Paul Black and Dylan Wiliam. It believes that *Inside the Black Box* (1998) and *Working Inside the Black Box* (2002) contain a fundamental critique of "an ill-judged confidence in the reliability of short, external tests, which are the dominant instrument of public policy" (1998).

23. There are many other studies too numerous to refer to. It is, however, highlighting three which the NUT believes the Select Committee must address.

- Research by the Institute of Public Policy Research (2001) found that pupils' mental health problems were directly linked to pressures connected with testing and recommended that the Government should take a less prescriptive approach if it was to halve the increase in mental health problems in schools.
- The Demos publication, *Beyond Measure* (2003), concluded that the assessment system measured recall of knowledge, rather than depth of understanding and tested only a narrow section of the curriculum and that it demotivated and lowered the self-esteem of learners.
- The studies by Cambridge University, on the lives of primary (2002) and secondary school teachers (2004), by John Macbeath and Maurice Galton, concluded that high stakes National Curriculum tests had almost wiped out the teaching of some Foundation subjects at Year 6. At secondary level, they found that the use of high stakes testing for the purposes of institutional evaluation, alongside high class sizes, inappropriate curriculum, pressure to meet targets and keeping up with initiatives, exacerbated unacceptable pupil behaviour.

24. Internationally, the Government in England is isolated in its approach to assessment. The Select Committee should draw on the evidence available through a study of 21 countries which have participated in the OECD's PISA programme (Programme for International Student Assessment) and the PIRLS (Progress in International Reading Literacy Study) study. Most developed countries within the OECD use forms of assessment for which are integrated within teaching and learning. Use of tests for evaluating the performance of teachers and institutions is rare but can have a powerful impact. The Select Committee would be well advised to look at the developments in the United States in relation to the use of standard assessment tests and compare those with the effectiveness of a high performing country education system, such as Finland, Sweden, Korea and Canada.

25. It is the steadfast refusal of the Government to engage with the evidence internationally about the impact of the use of summative test results for institutional evaluation which is so infuriating to the teaching profession. Despite the cautious and timid moves made by the Government on new types of testing within its *Making Good Progress* proposals, nothing rarely has changed in the psychological make-up of the Government on this issue.

26. Even more irritating for teachers is that, within the United Kingdom, developments have taken place to which the only Westminster Government response is that since both countries have responsibility for education, decisions on education remain with them. The Select Committee should certainly ask why the Government has shown no apparent interest in why the Welsh Assembly Government abolished testing for seven year olds, in 2002, and national tests for 11–14 year olds, in 2004, and why, in Scotland, teachers draw national assessment tasks from an electronic bank to support their judgements about pupils' attainment and why test scores are no longer collected by the Scottish Government.

27. In short, up until now, the Government has approached the whole issue of National Curriculum assessment with a curious mixture of blinkered stubbornness and timidity.

"Making Good Progress"—A Commentary on the Government's Proposals to Pilot New Forms of Assessment in the Primary and Secondary Phases

28. *Making Good Progress* contains potentially radical proposals for the future of assessment and personalised learning. Despite the DfES' assertion in *Making Good Progress*, that *"the issues . . . should be the subject of a larger and wider agenda which should involve debate across the school system"*, any potential for such a debate is diminished by its insistence on maintaining a high stakes approach to assessment and accountability. Indeed, any potential for positive reform in its statement that, *"ultimately, existing end-of-key arrangements could be replaced by tests for progress"*, is severely limited by a refusal to abolish school performance tables and imposed national targets.

29. The NUT believes that only if the proposals in *Making Good Progress* are amended on the following lines will they provide the basis for a valid and reliable pilot.

- The *Making Good Progress* pilot should be developed on the basis that it is a genuine experiment. Schools could either be granted permission to innovate through the Education and Inspections Act or be disapplied from the current assessment arrangements through existing legislation.
- Requiring schools to continue with end-of-key tests as well as expecting them to conduct progress tests, not only imposes an excessive burden on schools, it contaminates the ability to evaluate the relative merits of the progress tests in the pilot against the use of end-of-key stage tests in other schools. Teachers in the pilot will continue to have in their minds the fact that they have to conduct end-of-key stage tests as well. For national target and performance table purposes, teachers will remain constrained to focus on the borderline pupils in Year 6 and Year 9. The requirement in the pilot to conduct end-of-key stage tests as well should be dropped, therefore.
- Any expectation for schools in the pilot to conduct existing "optional" tests for Years 4, 5, 7 and 8 should also be dropped for the reasons given above.
- Decisions about when the progress tests should be conducted should rest entirely with teachers. The fact that *Making Good Progress* offers schools, "regular—perhaps twice yearly—opportunities" should not create any expectation that the tests should be conducted on this basis.
- Teachers should be involved in developing the progress tests. The Qualifications and Curriculum Authority should be asked to establish a forum for teachers involved in the pilot authorities to discuss the nature of progress tests.
- Schools in the pilot should be exempted from the national target setting and performance table regime.

— Independent research on the impact of the pilots on teaching and learning and teachers' workload should be commissioned by the DfES. The results of the pilot should provide the basis for a national debate initiated by the Government on the future of the assessment arrangements. The NUT agrees with the proposal within the 2020 Vision report that there should be, *"a shift of focus towards the progress of every child and away from thresholds and average attainment levels"*. It agrees also with the report's proposal for an investigation into the impact of national tests and examination preparation on the quality of learning.

— The Government asserts within *Making Good Progress* that national targets have driven up *"absolute levels of attainment"*. There is no evidence that national targets in themselves have achieved any such thing. Instead, the Government's national targets have led to an excessive skewing of resources towards children on the borderline of national target levels. This same skewing effect will take place if the proposed progression targets are introduced. Again, resources will be skewed towards borderline pupils. Pupils for whom achievement of an additional level outside the national target levels at Year 6 and Year 9 is possible with additional support may well suffer as a result.

— Adding progression targets to measures which have only just incorporated contextual value added targets will only increase the unstable mix of pressures on schools and increase the possibility of groups of pupils missing out on the support they need.

— The NUT welcomes the proposal within the pilot that pupils should receive funded, individual tuition and that qualified teachers should deliver it. If individual tuition is to be offered fairly to those pupils whom teachers judge will benefit, then it cannot be confined to pupils who may achieve two levels. The criterion which should apply in the pilot is that resourcing for tuition should go to the pupils who would benefit the most, particularly those from socially deprived families and/or who have English as an additional language. Judgements on which pupils they would be can only be made by schools themselves.

— Resources for individual tuition should be ringfenced. In addition, responsibility for providing individual tuition should not be simply added to the responsibilities of existing teachers. Additional qualified teachers should be employed within the pilot to deliver the additional personal tuitions.

— The NUT opposes the introduction of the progression premium for schools in the pilot. It is not clear what such a premium is for. There are obvious questions about its purpose. Is it a bribe for schools in the pilot to deliver success? Is the progression premium a reward? Would such a premium be available to all schools if the pilot was rolled out nationally?

30. In summary, the NUT has emphasised to the DfES that it should:

— remove the current optional and end-of-key stage test from the pilot; and

— drop the progression premium and re-allocate the funding set aside for the premium to increasing the number of days available for individual tuition.

The NUT's Strategic Proposals

31. Committee members will be aware of the NUT's Education Statement, *Bringing Down the Barriers*. Part 2 of *Bringing Down the Barriers* contained the NUT's strategic proposals on the National Curriculum and its assessments. Those proposals can be accessed in full by Committee members. For the purposes of this submission, the NUT believes it is worthwhile reminding the Committee of the direction the NUT believes the Government should take.

> *"Personalised learning has a long history based in part on child centred learning and the need to differentiate teaching according to need. Meeting the individual needs of each child and young person is an aspiration which all those involved in education can sign up to. The NUT believes that two conditions need to be established for personalised learning to succeed. A fundamental review of the National Curriculum and its assessment arrangements are essential to meeting the aspirations of personalised learning. Young people need to be able to experience, and teachers need to be able to provide, much more one to one teaching.*
>
> *The Government in England should recognise the major developments which have taken place in reforming assessment in Wales, Scotland and Northern Ireland. The review of assessment in Wales, conducted by Professor Richard Daugherty and his team, is a model from which the DfES should learn. An independent review of testing and assessment of children should be commissioned by the*

Government. Such a review should encourage and support assessment for learning and should examine the role of summative assessment. It should cover the current Foundation Stage profile and testing and assessment in the 5–14 age range.

There are no performance tables or national targets linked to test results in Scotland, Wales or Northern Ireland. School performance tables and national targets have the capacity to damage the record of Government on education as well as schools. The next Government should abolish both tables and targets."

"The data available from summative assessment and examination results should feed into school evaluation reports as they do in current inspection reports. To meet the country's need for a summative picture of the effectiveness of the education service, it should re-establish the Assessment of Performance Unit. The Unit would be able to summarise data and ask questions through studies based on sampling. Such a Unit would operate independently with an advisory board involving teacher and support staff unions, the TUC, the CBI, Government and relevant agencies. It would respond to requests for national evidence on standards within schools and colleges."

32. In *Which Way Forward?*, Professor Peter Mortimore's comparative review of the Government's White Paper, *Higher Standards: Better Schools for All*, the Education and Inspections Act, and the NUT's Education Statement, *Bringing Down the Barriers*, he commented on the NUT's proposals with respect to assessment and the APU.

"The proposal by the NUT for an independent review (of National Curriculum testing and assessment) seems imminently sensible. . . . it would be important for a review of assessment to be undertaken by a panel drawn from those involved with the system, advised by experts in the technical details. The panel might wish to study the arrangements adopted in Scotland and Wales as well as other international practice."

33. Peter Mortimore welcomed the NUT's proposal for, *"an APU-type body which would monitor national standards"*, which he believed would provide, *"a reliable national picture of standards of achievements"*.

34. In the context of the position outlined above, a position which has been subjected to reliable independent evaluation, the NUT believes that the Select Committee should adopt the following proposals.

— The NUT has taken the position consistently that teachers must be at the centre of defining the nature and purpose of assessment. Through its professional development programme, the NUT has supported and contributed to the development of assessment for learning. The Committee should consider proposing a national bank of teacher developed assessment tasks, which can be drawn down by teachers when they need to assess pupils' learning.

— There should be a major funding boost for professional development in assessment for learning. Teacher organisations could play a major part as providers of such professional development. Funding should be restored for inter-school moderation of assessment judgements. That funding could be transferred from the current major printing and distribution costs of National Curriculum end of key stage and optional tests.

— As part of an independent review of the National Curriculum assessment arrangements, the review of summative assessment should focus on how to achieve the most efficient and economic way of summarising and reporting pupil achievement within the context of the framework of the National Curriculum. One requirement of the review should be to focus on separating summative assessment from arrangements for institutional accountability.

— A review of National Curriculum assessment must be conducted by an independent group. Part of its remit should be to evaluate the arrangements in Wales and Scotland and explore developments in Northern Ireland. The *Making Good Progress* pilot would have value if it adopted the NUT's proposed modifications set out in this submission. The pilot could then provide substantive evidence to an independent review.

— The Government should re-establish the Assessment of Performance Unit so that a summative picture of trends in pupil achievement can be achieved nationally without subjecting schools to the vagaries of school performance tables. The Unit would sample pupil achievement from 0–19.

— In parallel with an independent review of National Curriculum assessment, the Government should review the measures in place it has for school accountability. Such a review would cover the current inspection arrangements, national targets and school performance tables. Its focus would be on achieving public accountability of schools whilst removing the warping and distorting effects of current high stakes accountability measures.

In Summary

35. The NUT submission to the Select Committee has focused necessarily on National Curriculum testing and assessments. The NUT has submitted evidence to the investigations conducted by the Select Committee on 14–19 education. The NUT backed fully the Tomlinson Report on 14–19 education. An opportunity is available to the Government to explore the development of the current diplomas and whether accompanying arrangements such as the changes to GCSE coursework are relevant and reasonable. That

opportunity presents itself as a result of previous Secretary of State, Ruth Kelly's commitment to review the progress of the post-Tomlinson proposals in 2008. The Select Committee should continue to press for such an approach despite the Secretary of State's view that the 2008 review will only focus on A levels.

36. There is one further proposal which the NUT believes the Select Committee should recommend to Government. It was contained in *Bringing Down the Barriers*.

> *"As a result of the 10-year lead time, no single Government can have responsibility for implementing the post-Tomlinson arrangements. It is essential, therefore, that 14–19 reforms should have continuity over time. The NUT would propose, therefore, the establishment of an implementation body which covers the full term of the post-Tomlinson arrangement. A broad range of representation from teacher organisations, the TUC, Learning and Skills Councils, universities and industry to the Qualifications and Curriculum Authority and Government would be included on its membership. Its job would be to provide a forum and sounding board for any potential problems arising from the practical implementation of change. Its existence would assist in establishing a social consensus for progressive change".*

37. This approach to 14–19 education is long overdue.

May 2007

Memorandum submitted by Professor Colin Richards

1. This memorandum argues that the issue of testing cannot be seen in isolation but needs to be considered as part of the re-formation of accountability in English education. It proposes a new style of accountability focussed at national, school and parental levels and involves reconsideration of the place and nature of national tests in the education of young people before the age of 16.

2. The views expressed are the result of experience of, and reflection on, national testing since its inception—in my roles as senior HMI, as Ofsted's specialist adviser for primary education and latterly as visiting professor at a number of universities in the United Kingdom and abroad.

3. The proposals for the future of accountability(including testing) assume that:

   (a) some form of national curriculum continues to exist;

   (b) some form of national testing is a political (and public) necessity;

   (c) some form of national inspection system is a political(and public) necessity;

   (d) the prime task of teachers is teaching their pupils;

   (e) autonomy must be balanced by accountability; and

   (f) in academic terms parents are particularly concerned with their children's achievement and progress in reading, mathematics and basic writing skills, though they also support their children's entitlement to a broad and balanced curriculum.

4. Accountability in pre-16 education needs to be rendered at national, school and parental levels.

At National Level

5. At national level the Government needs to keep standards under review and to devise a non-intrusive system for assessing pupils' performance in relation to those standards over time. It needs to be able to determine whether that performance is improving or deteriorating over time—preferably in relation to all the major components of any national curriculum, not just the core subjects. The current system of national assessment at ages seven and 11 does not provide a valid or reliable assessment of national performance over time. The Assessment of Performance Unit (operative during the 1980s) might provide a possible model on which to build.

6. The Government should set up an independent national body to review standards and to devise national tests which reflect those standards. Such a body would have to make publicly defensible decisions about which age groups would be involved, which components of the curriculum should be tested and which aspects of these components would be tested. The same set of tests would be administered year on year to a very small but representative sample of the school population. The tests would have to be administered confidentially to avoid pressures on schools for test preparation. Data at the national level would be published annually.

At School Level

7. The Government needs a system which assures that individual schools are providing a suitable quality of education and which triggers action should that quality not be evident. This requires some system of school inspection which assesses standards and quality and retains the confidence of parents and teachers. The current Ofsted inspection model does not provide this. However, parents have come to expect publicly available periodic assessments of the quality of and standards in, individual schools. It would be political folly to abandon the notion of regular inspection.

8. The current Ofsted model would be modified in a number of ways to make "inspection fit for purpose" Inspections would be lengthened (compared with the current "light-touch" model) but not to the same extent as the earlier Ofsted inspection models. This would probably involve lengthening the time between inspections from three to perhaps five years. Such enhanced inspections would focus on the classroom, not on the school's paperwork, and would report on (a) the performance of children in the work actually observed by inspectors; and (b) the quality of teaching and (as far as is possible) learning based on far more classroom observation than the current "light-touch" inspection model allows. Enhanced inspections should also report on the effectiveness of the school's procedures for self-evaluation and improvement. A summary of these judgements would be reported publicly to parents, along with a summary of the school's reactions to the inspection judgements. A very adverse report might trigger a full inspection or the bringing forward of the timing of the next enhanced inspection.

9. Governors and parents would have the right to request an inspection during the five year-year period between inspections and this request would be considered by either Ofsted or HMI (see 13)

10. Inspection teams would include the individual school's improvement partner (ie its S.I.P or its future equivalent) who would advise the inspection team, might (or might not ?) contribute to the team's judgements and would take responsibility with the head and governors of the school for any follow-up work consequent on the inspection.

11. The system of enhanced inspections would be administered by a reconstituted Ofsted whose inspectors would be drawn from the current cadre of additional inspectors (along with suitably trained headteachers on secondment) and whose management would be drawn from that same cadre. Whatever the failings of the current and post Ofsted models of inspection, the system introduced in 1992 has identified and developed the expertise of enough capable Ofsted inspectors to manage, "man" and regulate the proposed system of enhanced inspections.

12. HMI would revert to a role similar to that of pre-Ofsted days. They would be members of, and act as advisers to, a reconstituted Department of Life-Long Learning, would liaise with local authorities and would also carry out their own programme of survey inspections. In exceptional circumstances they might also inspect individual schools at the request of ministers. They might (or might not) consider inspection requests from parents (see 10 above).

At Parental Level

13. Parents need to be assured that the education system as a whole is performing well, that the schools to which they send their children are providing an education of appropriate quality, and that their children are making appropriate progress. The first two of these considerations would be met by the systems outlined above. It would be political folly not to provide parents with reliable information on how well their children are progressing in the so-called but "mis-named" basics.

14. To provide parents with information about individual progress teachers need to engage in ongoing assessment and to report its results.This would be provided in part by approaches to assessment for learning and in part by testing. It would be important that tests should serve, not dominate, good quality teaching.

15. There would be one or two kinds of testing. One would involve adopting the Scottish model of having a national data-bank of test items linked to progression particularly in English and mathematics and of teachers drawing, as appropriate, on this bank when seeking to determine or confirm their judgements of individuals' progress. These judgements would then be reported to parents on an individual basis. They would not be reported on a school by school basis (thereby helping to prevent "teaching to the test" or excessive pressure being placed on teachers for results) but they could be reported at an LA level (if thought desirable).

16. If the first type of testing is not considered sufficient, a second type would complement it—focusing on parents' main concerns: their child's performance in reading, mathematics and basic writing skills. National standardised tests would be devised to provide both summative and (if possible) diagnostic information which would be reported to parents on an individual basis, not on a school by school basis. Such national tests would be administered twice in a child's primary career—once on a one-to-one basis at the end of year 1 (followed where necessary by programmes of "reading recovery" and "number recovery") and once collectively at the end of year 5 (followed, where necessary, by more remedial or more challenging work to be provided within the same school in year 6). This slimmed-down programme of testing would replace the current end-of-key-stage and "optional" tests which too often dominate both the teaching and the curriculum, especially, but not only, in years 2 and 6.

CONCLUSION

17. Such a three-fold system would remove much (though not all) of the burden currently placed on schools by over-controlling regulatory measures—in particular national testing. It would provide government, schools and parents with appropriate information about progress and performance and provide an appropriate balance between professional autonomy and public accountability.

June 2007

Memorandum submitted by The Wellcome Trust

1. The Wellcome Trust is pleased to have the opportunity to respond to the Education and Skills Committee inquiry into Testing and Assessment.

2. The Wellcome Trust is the largest charity in the UK. It funds innovative biomedical research, in the UK and internationally, spending around £500 million each year to support the brightest scientists with the best ideas. The Wellcome Trust supports public debate about biomedical research and its impact on health and wellbeing. As part of this public engagement work, our education programme sets out to stimulate interest and excitement in science amongst young people, increase the quality of young people entering biomedical related careers and support scientific literacy more broadly.

3. Our response to this consultation will concentrate mainly on assessment in science education, which is of most relevance to the work of the Wellcome Trust. We argue that assessment systems in school science should ensure young people are tested on scientific understanding and skills rather than simply factual recall. In 2005, the Trust commissioned a study to explore primary teachers' views and experiences of science education across the UK. The resulting report, *Primary Horizons: Starting out in science*, highlighted issues relating both to formative assessment and statutory national tests. In terms of statutory testing, the study found that primary teachers felt that national tests had a negative effect on children's enjoyment of science, because of the increasing tendency to "teach to the test". An over-emphasis on curriculum content and pressure to prepare for national tests were felt to reduce opportunities for investigative work and lead to science frequently being taught as a collection of facts.

4. Young people often respond very positively to constructive feedback that tracks their progress and informs their development. Teachers in the *Primary Horizons* study agreed that both teachers and children find this sort of "formative assessment" ("assessment for learning") more enjoyable than "summative assessment", where the main purpose is to give a quantitative grading. Pressure to perform well in national tests in England and Northern Ireland appeared to have a considerable negative influence on the use of formative assessment in the later primary years.

5. The report recommended that further research should be carried out into the effects of national tests on young people's attitudes to science, and on opportunities for children to develop investigative, questioning and thinking skills. The Wellcome Trust is in the process of commissioning research into the effects of compulsory national testing in science on teachers and teaching at Key Stage 2. The study will include a comparison between teachers in England and Wales, where compulsory national testing has now been removed at Key Stage 2, providing the rare opportunity to compare two different assessment systems operating with the same curriculum. We have discussed the research with the Clerk of the Committee and hope to be able to contribute interim findings to the Committee's inquiry by the end of the year.

6. We would also like to highlight the role of continuing professional development (CPD) in supporting assessment. The *Primary Horizons* report found that those teachers who had undertaken CPD in science were more confident in nearly all aspects of science teaching, including the use of formative assessment. Teachers who had participated in science-specific CPD were significantly more likely to include investigations, mind mapping and individual target setting as part of their assessment in science. Provision is now in place to deliver high quality CPD for teachers of science through the national Science Learning Centres network; it will be important to realise the benefits of this network to support and enhance assessment skills.

June 2007

Supplementary memorandum submitted by the Wellcome Trust

1. The Wellcome Trust is the largest charity in the UK. It funds innovative biomedical research, in the UK and internationally, spending around £600 million each year to support the brightest scientists with the best ideas. The Wellcome Trust supports public debate about biomedical research and its impact on health and wellbeing.

2. The Wellcome Trust responded to the (then) Education and Skills Committee inquiry into Testing and Assessment in June 2007. In our submission we referred to research that the Trust was commissioning into the effects of compulsory national testing in science and said that we would be able to provide interim findings to the Committee.

3. The summary report from the research is now available and is attached as an Annex.[14] The research was conducted by the Institute of Education and considered the effects of compulsory national testing on the teaching of science, and teachers, at Year 6 (Y6) in England; and the impact of the abolition of statutory testing in science at Key Stage 2 (KS2) in Wales on Y6 science teaching and teachers.

4. The research was conducted through a telephone survey of 600 teachers, science coordinators and headteachers (300 from England, 300 from Wales) and sixteen focus groups (eight in England, eight in Wales) involving a total of 74 Year 6 teachers, science coordinators, headteachers and secondary science teachers. The summary report is therefore drawn from teachers' perceptions of the impact of compulsory testing.

5. The key points from the report are:

— teachers consider statutory testing to be leading to a narrowing of the science curriculum and limiting approaches to teaching;

— teachers in England find it difficult to maintain positive attitudes to science among pupils and suggested that abolishing testing would make science more enjoyable;

— teachers in England feel that abolishing statutory testing would enable them to respond more to individual pupils' needs and would allow pupils to develop greater independence in learning;

— teachers support retaining optional test materials to inform teacher assessment (as they have been in Wales); and

— summative teacher assessment is seen to provide a more accurate assessment of pupils' level of attainment in science than national test results and teachers are concerned about how data from school achievement tables may be used.

6. While it is still too early for the effects of the abolition of testing in Wales to be fully realised, Welsh teachers suggest that the emphasis in science teaching is beginning to shift away from the transmission of factual knowledge towards the development of pupils' skills to support their learning in science, and there has been an increased emphasis on small group work and practical activities.

7. The report makes three recommendations:

— there should be a review of classroom support for science in Y6 classes to enable practical whole class activities and to support summative teacher assessment of pupil attainment in England and Wales;

— consideration is needed of how to improve progression in learning during KS2 in England and Wales; and

— there should be an evaluation of how appropriate it is to use school achievement tables based on KS2 test results for purposes of accountability in England.

8. The Trust would urge the Committee to consider these recommendations in its review of testing and assessment in England.

March 2008

Memorandum submitted by Hampshire County Council

This response is provided on behalf of Hampshire Local Authority (LA). It does not claim to offer a comprehensive representation of the views of all involved in the processes of testing and assessment. However it draws extensively on the expressed opinions and experiences of many colleagues in the inspection and advisory service and of practitioners in the primary and secondary phases of education. It focuses mainly on issues at Key Stages 1, 2 and 3.

SUMMARY

— The assessment regime has become enormously burdensome for schools.

— National Curriculum (NC) tests have now expanded out of all proportion to their usefulness.

— NC testing underpins a system of school accountability, self- evaluation and regular inspection through the provision of NC data on pupil attainment and progress. However, the validity and reliability of some of the data continues to be questioned by practitioners and researchers.

— The tests skew teaching by focusing on core subjects and encouraging widespread teaching to the tests.

— The system has encouraged the growth of an expensive assessment industry.

[14] *Research into the Effects of Compulsory National Testing in Science on Teachers and Teaching at KS2*, Dr Sue Collins, Professor Michael Reiss, and Professor Gordon Stobart, Commissioned by the Wellcome Trust in association with the Association for Science Education. Not printed. Available on Committee website.

— There is a case for retaining testing at the end of each phase of statutory education (ie at ages 11 and 16) but any assessment at other times should be used formatively by teachers and not as an accountability tool.

— Important aspects of pupils' abilities and skills are often ignored because the emphasis on tests gives higher status to more easily measured aspects of the curriculum.

— The recent DfES proposal for more frequent tests for each NC level ('Making Good Progress' pilot) is counter to the extensive research on effective formative assessment.

— The development of a rigorous approach to teacher assessment offers greater opportunity for professional development and a fairer and more valid way of monitoring pupil progress.

National Key Stage Tests

1. *The uses and abuses of assessment data*

— Tests are acknowledged to be just 'snapshots' of pupil attainment but the numerical data from them is increasingly treated as definitive in relation to individuals' progress and in judging a school's effectiveness. However, we know from research that there can be significant errors in grading students through external tests.

— Schools now have a vast amount of numerical data for tracking pupils' progress. Each NC level was originally conceived as a broad descriptor of approximately two years' worth of progress. Pressure on schools to account for progress more regularly has led them to invent criteria to describe progress within NC levels. There is little agreement regarding these 'sub-levels' and QCA and The National Strategies have been reluctant to support their use. Nevertheless a notion of progress (and teachers' performance) is predicated upon these somewhat specious concepts.

2. *Testing at Key Stages 1, 2 and 3*

— At Key Stage (KS) 1 the new assessment arrangements (from 2004) have encouraged a more flexible approach to combining externally set tests with teachers' professional judgements. National and Hampshire LA Evaluations of this approach have been very positive. In particular, stress on pupils has diminished and teachers have found the system to be more professionally rewarding. In Hampshire, the changes have triggered effective procedures for year 2 and 3 teachers to work together systematically and regularly to discuss standards and moderate their judgements. National Evaluation findings were also encouraging in respect of workload issues.

— At the end of KS2, tests can serve a useful function in contributing to the pupil information for secondary schools; but at present these tests drive the whole assessment system. The majority of children sit at least one externally devised set of tests each year during KS2. Practice tests for the so-called "optional" QCA year 3,4 and 5 tests are common. The system is underpinned by a target setting culture for schools, teachers and children and this has become increasingly bureaucratic.

— Value added measurement from KS2 to KS3 is insecure since local decisions about when pupils take the "end of KS3" test (Year 8 or 9) are leading to incompatible measurements of progress.

— For many schools the threshold indicator of level 5 presents little challenge as many pupils are already at this point when they join the secondary school at year 7.

— The vast majority of secondary schools see the five year measure of progress as one of the most useful indicators since, unlike KS2 to KS3, it does measure the totality of performance at KS4 compared with core subject performance at KS2.

— Any assessment at or during KS3 should therefore be for formative purposes and as a professional tool for teachers rather than as an accountability and performance measure.

3. *Impact on teaching, learning and the curriculum*

— LA monitoring has identified widespread teaching to the test and practising of tests especially, but not exclusively, in the final years of each KS.

— This seriously detracts from time spent exploring more imaginative and creative aspects of the curriculum and skews teaching.The emphasis is on short-term commitment to memory and 'test tactics' rather than deeper learning and understanding.

— The perceived status of those areas of the curriculum not formally tested (eg: the arts and humanities at KS1–3) is diminished as a consequence.

— It is misleading to claim (as the DfES Making Good Progress pilot does) that tests can support personalised learning. So much of individual pupils' experiences are narrowed because of the tests and tests can reflect only a small part of a pupil's skills, abilities and understanding.

— It is disingenuous to argue that teacher assessment (TA) provides a counterbalance to test results. TA appears to have very little status in the world of the School Improvement Partner (SIP), Ofsted or the reporting of end of KS results to parents.

— External tests detract from the development of a strong, well-informed professional teaching force. The implication is that external "objective" test markers know better or that teachers cannot be trusted to make and agree (moderate) their own judgements.

— A vast body of research into assessment suggests that students make best progress when assessment information can be used by their teachers and by the students themselves in a formative way ie: through response to feedback which is specific and close to the point of learning. Externally marked tests do not serve this purpose and we should not pretend that they make a significant contribution to the progress of individuals.

4. *The effects on the people involved*

— The high stakes involved have an observable effect on the behaviours of teachers, children and parents. There are ambivalent attitudes here since all involved have a natural desire to do well in a competitive business. In addition there are many students who find tests an interesting and enjoyable challenge. Equally though, many students fail to do their best under test conditions and suffer considerably at examination time. It is widely claimed that English pupils are more tested than any pupils in other nations in Europe. It is probably no coincidence that a recent survey also found they are the least happy!

5. *The financial cost of testing*

— Tests are now part of a huge and very expensive industry including:
  — commercially produced practice tests,
  — external markers, reviewers and checkers for statutory tests,
  — LA monitoring of tests at KS 1, 2 and 3,
  — monitoring of special arrangements to deal with applications eg: for additional time for pupils with special educational needs (SEN),
  — exam board bureaucracy on a grand scale, and
  — National Assessment Agency (NAA) and LA test maladministration investigations.

Much of the funding and energy involved might be better directed at further improving the quality of day-to-day teaching and learning.

Tests have a place and value in schools. For pupils and teachers they can provide evaluative information about what students can achieve independently under restricted conditions. However their increasingly extensive use for the purposes of accountability has now become a distraction for teachers, headteachers and governing bodies in their core purpose of educating pupils.

The Future

— If tests did not exist, schools would feel they had to invent some, at least for internal use. The materials produced and extensively trialled over the years by QCA, the National Foundation for Educational Research (NFER) and others are of high quality. Schools would welcome banks of such tests they could use to support their internal monitoring of progress and the processes of self-evaluation.

— This is different from the shorter "when ready" testing proposed in the current DfES pilot (Making Good Progress). The latter is likely to lead to more tests whose results are even less reliable and valid as an overall picture of a child's progress than at present.

— The QCA (KS2) Monitoring Children's Progress pilot (and a similar pilot for KS3 English) offers better opportunity for the formative use of a wider range of information about children's skills and abilities. It promotes good professional development for teachers and more immediate feedback for their planning.

— Investment in the development and trialling of rigorous moderation of teacher assessment processes at KS 2 and 3 is long overdue. A variety of approaches might be examined. Key Stage 1 assessment arrangements provide one model. A number of LAs (eg: Oxfordshire and Birmingham working in partnership) have examined other approaches at KS 2.

— It might then be possible to develop a system of national teacher assessment at end of K S 1, 2 and 3, supported by a method of whole school or pupil sampling of national externally set tests. The APU (Assessment of Performance Unit) model, for instance, provided useful national information without a hugely expensive bureaucracy.

Schools readily acknowledge the need to monitor pupil progress, provide regular information to parents and use assessment information evaluatively for school improvement. The key issue now is how to balance the need for accountability with the urgent need to develop a fairer and more humane assessment system that genuinely supports good learning and teaching.

May 2007

Printed in the United Kingdom by The Stationery Office Limited
5/2008 388381 19585