



Qualifications and
Curriculum Authority

Report on Comparability between GCE and International Baccalaureate Examinations

December 2003

Background

QCA has an extensive remit, requiring it to ‘create a clear, coherent and well regulated framework of national qualifications’ covering the whole range of qualification types. To achieve this the organisation must be able to demonstrate how far equivalence, similarity and difference exist between qualifications.

Clearly it is not feasible, or desirable, simply to compare all qualifications with each other, so it is necessary to consider areas where issues of comparability have been raised, or where possible issues can be foreseen. Such work is important in helping meet another of QCA’s key aims, of securing a rigorous, consistent and fair system of assessment.

One particular qualification about which it is important that QCA can speak with some authority is the International Baccalaureate Diploma. There are several reasons for this. At the simplest level, it has recently been accredited as a general qualification at level 3 and it is important that the decision is followed up at a fairly early stage. It is also a qualification with a number of staunch supporters in the UK who are vocal in declaring its qualities. In addition, with the Tomlinson Task Force reviewing the whole structure of post 14 qualifications in England and seeming to favour some sort of diploma, it is timely to consider the main example currently in use in the UK.

It was therefore decided that QCA should mount a comparability exercise looking at A level examinations and the IB diploma.

There is a further critical factor that served to make the work particularly timely. The DfES tasked QCA with devising a mechanism whereby all approved qualifications can be included in school and college performance indicators. The first phase of this work proposed a mechanism and associated scale for the process, together with values for most qualifications used in school pre-16. Phase 2 of the work has concentrated on further evaluation of the proposed values and the inclusion of other qualifications taken by students mainly post-16, including relevant level 3 qualifications. Given that the IB Diploma is unusual among qualifications which students take in this phase in being the whole of a full-time educational programme rather than something additional to the familiar set of level 3 qualifications, it is particularly important that it can be included in the work.

The Qualifications

The International Baccalaureate Diploma

The International Baccalaureate Diploma Programme was created in 1968. The International Baccalaureate Organisation (IBO) describes the Programme as a demanding pre-university course of study that leads to examinations. It is designed for highly motivated secondary school students aged 16 to 19.

The programme is a comprehensive two-year international curriculum, available in English, French and Spanish. It generally allows students to fulfil the requirements of their national or state education systems. The Diploma Programme incorporates the best elements of national systems, without being based on any one.

The programme was born of efforts to establish a common curriculum and university entry credential for students moving from one country to another. International educators were motivated by practical considerations but also by an idealistic vision. They believed that students should share an academic experience that would emphasize critical thinking, intercultural understanding and exposure to a variety of points of view.

The programme has the features of a traditional and broad curriculum, but with three important additional aspects. These are the theory of knowledge (TOK), Creativity, action and service (CAS), and an extended essay.

The TOK is an interdisciplinary requirement intended to stimulate critical reflection on the knowledge and experience gained inside and outside the classroom. The course challenges students to question the bases of knowledge, to be aware of subjective and ideological biases and to develop the ability to analyse evidence that is expressed in rational argument. The course is unique to the IBO, which recommends at least 100 hours of teaching time spanning the programme's two years.

The IBO's goal is to educate the whole person and foster responsible, compassionate citizens. The CAS programme encourages students to share their energy and special talents with others. Students may, for example, participate in theatre or musical productions, sports and community service activities.

Each student has the opportunity to investigate a topic of special interest. The extended essay (EE) requirement acquaints diploma candidates with the kind of independent research and writing skills expected by universities.

The IBO recommends that a student devote a total of about 40 hours of private study and writing time to the essay. The essay permits students to *deepen* their programmes of study, for example by selecting a topic in one of their higher level (HL) courses. Or they might add *breadth* to their academic experience by electing to write in a subject not included in their programme choices.

Students for the Diploma also study six subjects each drawn from a different curriculum area. At least three and not more than four are taken at higher level (HL), the others at standard level (SL); HL courses represent a recommended minimum of 240 teaching hours, SL courses cover 150 hours. Active citizenship and global perspectives are encouraged in each area of the curriculum. The six groups are:

Group 1 - language A1

More than 80 languages have been offered for examination as part of the IBO's policy of encouraging students to maintain strong ties to their own cultures. Students ideally develop strong written and oral skills, respect for the literary heritage of their first languages, and an international perspective.

Group 2 – second language

All diploma candidates are examined in a second language. Several options accommodate bilingual students with a very high level of fluency, genuine second language learners with previous experience learning the language, and beginners. The principal aim for the subjects in group 2 is to enable students to use the language in a range of contexts and for many purposes; the courses focus on written and spoken communication.

Latin and classical Greek are also available in group 2.

Group 3 - individuals and societies

Subjects included in this group are: business and management, economics, geography, history, Islamic history, information technology in a global society (at SL only), philosophy, psychology, and social and cultural anthropology.

Group 4 – experimental sciences

The subjects available in group 4 are: biology, chemistry, physics, environmental systems (SL), design technology. Practical laboratory skills are developed and collaborative learning is encouraged through an interdisciplinary group project. Students develop an awareness of moral and ethical issues and a sense of social responsibility is fostered by examining local and global issues.

Group 5 – mathematics and computer science

All candidates for a diploma are required to complete a mathematics course, and four options are available to cater for different abilities and levels of student interest. These are: mathematics HL, further mathematics SL, mathematical methods, mathematical studies. Each course aims to deepen a student's understanding of mathematics as a discipline and to promote confidence and facility in the use of mathematical language. Computer science is an elective subject in group 5; it is not compulsory.

Group 6 – the arts

This group includes visual arts, music and theatre arts, with emphasis placed on practical production by the student and exploration of a range of creative work in a global context.

Instead of a group 6 subject, a candidate may select an additional subject from groups 1 to 4, or further mathematics SL, or further mathematics and computer science from group 5.

Grading

Each subject, whether HL or SL is graded on a seven point scale, with 7 representing the highest attainment. In addition, performance in TOK and the extended essay together are allocated up to three further points, making a total scale of 45 points. There is no credit awarded for CAS, other than the requirement that it is satisfactorily completed. The Diploma is awarded for candidates gaining 24 points or more. No differentiation between HL and SL performance is made in the award of the Diploma, although there is an extensive range of hurdles with the main aim of ensuring balanced achievement across the programme.¹

GCE A Levels

Unlike the IB Diploma, A levels do not call for students to follow an overall programme. Each subject is offered as a discrete qualification, with little real guidance even for each subject as to the programme of study to be followed. There is, for example, no recommendation as to the number of hours of study expected. This is partly because the qualification is available for a wide range of possible candidates, including mature students perhaps taking a single subject out of general interest. It is also probably the case that even the institutions which offer A levels as full time programmes for students between 16 and 18 need to cater for a more heterogeneous student body than those which offer the IB in the UK.

¹ These range from requiring all candidates to have a final grade in all six subjects, the EE and TOK and to have completed the CAS requirement (this is a key characteristic of a baccalaureate system and an important difference from the A level system), to proscribing very poor performance in any one element or relatively poor performance in several. There are specific further restrictions for candidates with 24-27 points overall and others for candidates with 28 or more. Many of the restrictions are designed to ensure that the candidate had performed adequately in both their HL and SL subjects.

There are, however, features of A levels characteristic of all the examinations. This is particularly the case with the qualification post 2000 when it was extensively revised. Every A level is now divided into six units of roughly equal size. Three of these units are set at AS level, defined as being the standard expected of a candidate at the end of the first year of a two-year course. The remaining three units are A2 units, designed to be taken during the second year of a two-year course. It is a requirement that the six units taken together maintain the standard of the previous A level examinations. There are limitations imposed for each subject on the maximum weighting allowed for internally assessed work.

Because each examination is modular, there are also rules to govern this aspect. For example, at least 20% of the assessment in each subject must be synoptic in nature, ie assessment which tests candidates' understanding of the connections between the different elements of the subject. This is normally externally assessed. In addition, candidates are allowed one re-sit in each unit, with the better result counting towards the final grade.

An important further aspect of the A level system is that there are several independent awarding bodies offering examinations in a full range of A level subjects. Sometimes, in fact, a single awarding body may offer more than one A level in a particular subject. To try and ensure fairness in the system, it is important that all the examinations in a subject are seen as comparable by all users of the qualification. To this end, there is a regulatory authority for England, the Qualifications and Curriculum Authority (QCA), which has the task of supervising the system.² This takes the form of drawing up criteria for the qualifications to meet, accrediting specifications against the criteria, and establishing codes of practice outlining the principles and processes through which the awarding bodies should conduct their business. The aspect of external regulation was felt to have had some impact on the two qualifications, and this is discussed further below.

Grading

Each separate A level subject reports on a five-point letter scale, A-E, with A representing the highest attainment. Candidates taking the full A level also receive an AS grade based on the aggregate outcomes of the three AS units, also on an A-E scale, and a separate standardised mark on each unit. It is this information which allows candidates and their teachers to make informed decisions about whether or not to re-sit a particular unit.

Introduction To The Study

All exercises in examination comparability present formidable problems. The very existence of differences means that one is never comparing like with like. In the case of this exercise, the difficulties are magnified. Indeed, such are the differences that some of the questions are almost impossible to answer. This is reflected in the approach taken to the work on this occasion, which confined itself to the less intractable aspects.

² There are parallel bodies in Wales and Northern Ireland who work with the QCA to regulate the system of qualifications across the three countries.

There are two key differences between the qualifications which militate against producing effective comparisons. The first of these derives from the IB Diploma being a single full-time course of study for 16-18 year olds (with some level of optionality but precise rules of combination) leading to an overarching qualification while A levels are single subject qualifications which can be taken by people of any age and in combination with any other qualifications, or with none.³ This makes some aspects of comparison almost impossible. For example, how does one compare a very specific, designed programme of study (including non-assessed elements such as service to the community) with a clear philosophy to individual examinations having no overall philosophy beyond the aims and rationale in each subject? The issue is even more intractable in that the changes to A level introduced in 2000 have not yet settled in sufficiently to allow for any assumptions to be made about a typical A level programme for 16-18 year old students.

For the purposes of this study, for which there was some imperative to be able to report in autumn 2003, the decision was taken to concentrate comparisons at the level of individual subjects. Four subjects were chosen with the aim of covering a range of traditional subjects in the curriculum, ensuring that each one came from within each of four of the IB curriculum groupings. The subjects were chemistry, English literature, history and mathematics. Clearly, as long as findings were reasonably consistent across subjects, this would permit some consideration of how the whole Diploma programme might relate to A levels.

The second difficulty could not be ducked: the quite different grading scales the two qualifications use. It is important to note that there are no pre-set points of equivalence, partly because the international nature of the IB diploma means that it cannot peg its standards too closely to those established in any one country.

Importantly for the purposes of the design of this study, the relationship between HL and SL seems to be neither that between the current AS/A2 model (i.e. the one at an explicitly lower level than the other) nor that of the old model where the AS was of the same standard as the full A level, but with less breadth. Rather it seems to be something of a hybrid of the two.

Significantly, HE offers to students taking the IB often take the form of requiring a particular number of points, with an additional requirement of a particular performance in a given subject at HL. This pattern is very similar to that seen for A level candidates who may be asked for a particular overall performance with a specified grade in a given subject. For this reason, the decision was taken to compare the HL IB examinations with full A levels. For the A levels, the specification with the highest entry was used as the comparator.⁴

Taking such a decision does not, however, resolve key questions about the scope of the exercise. In this respect, the changes to the A level examination introduced from 2000 raised important issues about how best to structure the exercise. There were two key matters to consider. In order to understand these it is useful to know the basic intended structure of the study. For each subject, the exercise aimed to compare two aspects of the assessment:

- the demands of syllabuses and their assessment instruments (eg question papers, mark schemes); and
- the level of performance required of candidates at key grade boundaries.⁵

³ In fact it is possible to gain certification for individual IB subjects – a fact which helped to justify the approach taken in this study – but it is only the IB Diploma which has been accredited for use in the UK.

⁴ This meant that AQA specifications were used for Chemistry and English Literature and Edexcel ones for History and Mathematics.

⁵ The term syllabus was used throughout this exercise as being one which all those involved felt familiar with and usefully neutral. A levels now have “specifications”; the IB uses the term “subject guide”.

The AS/A2 structure of A levels creates an immediate question when addressing both these aspects. For the first, the question is at least relatively easy to answer. The IB is a two-year course, followed by students from 16 to 18. The syllabuses provide a description of all the subject material a student has to cover in both years, and information about all the assessment they will undergo. For any given A level, a candidate has to take all six units of assessment, both AS and A2. To evaluate the demands of the two examinations means that the full range of A level content, both AS and A2, needed to be considered. So too did all the examinations taken, in considering the overall assessment burden on a candidate in a given subject. (Even this still presents considerable problems for evaluation since it necessitates balancing the added volume of assessment provided by the AS with the fact that it is, by design, less demanding than that in the A2, but expected to be taken by students in the first year of the course.)

When comparing levels of performance in the two qualifications, the issue of what materials to use to make the comparison is much harder to resolve. Performance in the AS is not intended to be at the full A level standard. It is not easy to see what useful comparisons could be made between this new standard of performance and the IB grading scale, especially given the quite different nature of the scales. For the purposes of this study, it was decided to compare performance in the A2 units with that in the IB examinations. The main rationale for this is that both examinations are taken by students at the end of their course and graded accordingly.

There were other aspects of the examinations which raised issues about the exercise, but these are discussed in the main body of the report.

Methodology

Personnel

The exercise treated each subject discretely and made use of subject experts to provide judgements about each of the two key aspects of the examinations mentioned above. There were nine experts used for each subject making a total of 36 in all.

Part of the methodology required participants to transfer grading standards across qualifications. Where the qualifications differ significantly, this is extremely difficult and calls for a close knowledge of the grading standards used in the “home” qualification. This meant that some of the experts needed to be drawn from the communities of examiners working on each qualification. The IB was asked to nominate, for each subject, three senior members of their examining team who had knowledge and experience of the standards expected for grading purposes. The A level awarding body with the highest entry specification was also asked to nominate three members of their senior examining teams with equivalent knowledge and experience. At the same time, it was felt that a level of independent evaluation of the qualifications was also desirable. To this end QCA recruited, for each subject, three teachers in higher education who had knowledge and experience of the kind of standards required for admission to their institutions (ideally admissions tutors).

The work in each subject was managed and facilitated by QCA staff, but they played no part in the judgements made.

Materials

The IBO and the relevant A level awarding bodies were asked to supply original examination materials for the study and to nominate suitable personnel to participate. For the review of syllabus demand, the materials comprised copies of the syllabus, of all examination papers for the summer 2002 series together with associated mark schemes, any examiners' reports, details of the marking criteria associated with any internally assessed element in the examination, and any other materials routinely supplied to centres, such as book lists, formula sheets, guidance on the setting and marking of assignments etc.

For the review of candidate performance, discrete materials were supplied for each component of the IB examination and each A2 unit of the A level. Where there were coursework and non-coursework alternatives, the non-coursework alternative was used – for logistic reasons only. For each component/unit scripts from three candidates were supplied, on marks at regular intervals running from a level of performance well above the highest grade awarded (grade A at A level, 7 points for the IB) to well below the lowest “pass” grade awarded (grade E at A level, 4 points for the IB). The choice of boundaries was dictated in part by those where judgemental grading decisions are made. The rationale for the range of scripts was that, because there was no agreed or expected equivalence between any of the grades and points, it was necessary to supply performance covering a very wide range.⁶

Approach

For a wide range of reasons – logistical, resource constraints and a desire to avoid any misleading precision that may arise from quantitative analysis – the basic approach was qualitative, with the aim of obtaining limited quantities of data-rich information which would be sufficient to establish broad areas of congruence and difference. Participants were asked to complete a number of structured protocols to provide a reasonably consistent analysis of the cognitive demands of the two sets of examination materials. On several occasions, as part of this process, they were asked to provide numerical ratings of an identified aspect. In such cases, it was made clear throughout that such ratings were not going to be analysed; rather the objective was to provoke qualitative comment to explain the ratings, especially if they differed.

The aim of this structured approach was to ensure a reasonable level of consistency in the aspects covered, without attempting to eliminate critical differences in perspective which might prove vital to understanding similarities and differences and their effects. Copies of the protocols used are provided in Appendix A, but the basic method was to describe, then analyse, then evaluate. The question papers and mark schemes, which in many ways hold the key to the demands of an examination syllabus, were subjected to a focused piece of analysis using the key features of examination questions through which the cognitive demand can be altered.

Participants carried out this part of the work in their own time, completing exactly the same set of forms in a given subject. The forms were tailored, however, to ensure they suited the nature of the subject. For example, the forms for English literature and History contained no reference to the mathematical demand of the papers; those for Chemistry did.

⁶ In fact, the pilot exercise had suggested that such a wide range was not necessary for plotting points of comparison across the key judgemental grades. However, the desire not to pre-judge the main exercise, especially in subjects which had not been piloted, meant that the wide range of performance was still felt desirable. Moreover, the way this part of the exercise was carried out for HE reviewers meant that no assumptions could be made about the levels of performance expected.

For the review of candidate performance, the tasks carried out by the reviewers were conceptually the same, but differed to accommodate the different perspectives and experience that the three groups of experts brought to the exercise. The group of senior examiners from the A level examination, working independently, were asked to look at the work of the IB candidates, component by component, and decide where they felt they would set their own judgemental grade boundaries (i.e. minimum A and minimum E). To assist with the task, again working individually, they first looked at the units from their own qualification and decided appropriate A/B and E/U boundaries. This served both to refresh their memories to the kinds of performance they had seen when awarding the units operationally and to provide data to help in calibrating their judgements. (The operational boundaries were known and if, in the circumstances of the project, a reviewer were to be consistently harsh or severe compared with the real boundaries, this could be built into the evaluation of the results.) The IB examiners carried out a parallel task, refreshing their memories of what they looked for in minimum 7 point and 4 point performance on the components of their own examination and then trying to apply those standards to the A2 units.

The HE reviewers had a slightly different task. They had no direct experience of grading either examination and, usually, little or no knowledge of the details of the syllabuses. For each individual component/unit, they were asked to review the candidates' scripts methodically and determine what level of performance represented the minimum level needed for a student to be able to cope with a degree course in the subject at their institution. Because HE institutions vary, there was no expectation that the HE reviewers would have the same requirements of their prospective students; indeed, ideally reviewers were sought from a range of institutions so that a number of different levels of performance might be identified. All that was required was that each reviewer should be reasonably consistent in the expectations they brought to bear when forming their judgements.

The review of examination materials was carried out in reviewers' own time. However, they came together on several occasions, although the HE reviewers were kept apart from the examining teams. First, the two groups of examiners and, separately, the HE reviewers met at QCA to be briefed about the work and introduced to the specific tasks to be carried out. At this initial meeting, reviewers were given the background to the work, and specifically briefed on the nature of the two qualifications. (The IB was unfamiliar to most of the non-IB participants, except in the most general terms, while the reforms of A levels in 2000 meant that neither the IB examiners nor the HE reviewers were necessarily fully aware of the changes. In particular, of course, they were not familiar with the specification being used for the work.) The forms to be completed were then described and discussed to ensure that all participants were happy that they suited the nature of the subject concerned and that there was a shared understanding of the various terms used and judgements to be made. For example, in one of the forms, reviewers were asked to rate various elements on a five point scale ranging from 1 = very undemanding to 5 = very demanding with 3 as "just right". It was made clear that 3 meant "just right for students following a two-year course in the subject between the ages of 16 and 18". There was no requirement, however, that precisely what this meant should be identical for all participants: indeed, differences in this kind of perception were felt likely to be quite illuminating.

After completing their forms, participants came together, again in their two separate groupings, to explore their initial findings and to review candidate work. This took the form of a two-day residential meeting, although it was stressed that judgements should continue to be independent. (In fact, bringing participants together for this part of the exercise was purely for logistic reasons: sending scripts through the post for review by nine separate individuals is both high-risk and very time consuming. Moreover, the residential meeting allowed QCA staff to be on hand to respond to any queries about the task.)

There was a final meeting where reviewers were given a short briefing on the general nature of the findings in their subject and in particular to any key issues which had been raised. Any differences of opinion were explored to try and ensure that they were real and could be properly reflected in any report, while areas where there was agreement were evaluated to determine their effect. Reviewers were also invited to comment upon the methodology.

Findings and discussion

Overview

The findings reported here are those where there was sufficient commonality across all four subjects to suggest they are a feature of the two qualifications, although examples will be drawn from individual subjects. More detailed findings for each subject are provided at Annexes A-D.

Overall, the reviewers felt that the task, although presenting considerable problems generally and some specific difficulties in particular subjects, had been feasible, so that there could be reasonable confidence in their findings.

These findings, broadly, were that A level and IB examinations were comparable in the demands they placed on the candidate at the level of the individual subject. There were some interesting subject-specific differences either in breadth or depth, compensated in each case by parallel differences in depth or breadth. In terms of candidate performance, there was general consistency in suggesting that the kind of performance meriting an E was similar to that required for 4 points and that that needed for an A was comparable to that needed for 7 points. The HE reviewers in general tended to confirm these findings, in that the range of equivalences implied by their range of judgements tended to be those that should occur were A and E comparable to 7 and 4 respectively.

Where there were exceptions to this, they tended to occur on components where there was least comparability. For example, the IB Chemistry examination assesses practical skills by means of an internally-assessed component, calling for the assessment against prescribed criteria of candidates' reports on a series of experiments. In the A level, the assessment of practical skills used in the study took the form of an external examination. These very dissimilar methods of assessing the same range of knowledge and skills proved hard for all the reviewers to compare.

Examination demand

Within the general background of comparability of demand, there were some consistent differences. These are interesting to consider, partly because sometimes the same underlying cause has radically different effects in different subjects. There appeared to be two principal sources of difference between the two qualifications. The first of these derives from the extent to which they are externally regulated; the other comes from the nature of the candidates who are expected to take the examinations, where the situation is more or less the mirror image of that arising from regulation.

To take the second of these first. The IB is as much a course of study as it is a qualification. It is designed to be a full-time course for highly-committed and internationally-minded students, between the ages of 16 and 18. Its nature is highlighted by the fact that it includes a compulsory element of what in the A level world are called enrichment activities. This element, is non-assessed, but fairly closely defined and controlled and satisfactory performance is a pre-requisite for the Diploma to be awarded. As already described, the subjects studied are also fairly tightly prescribed. In fact, many of the students take all their examinations in what is effectively their second language (usually English) and are assessed in a further language as their foreign language. The whole structure of the course – including the examination on the theory of knowledge – and to a great extent its assessment are traditionally rooted in a concept of the educated person with which Matthew Arnold would probably have felt comfortable.

The A level is designed as a set of single subject examinations which can be taken in combination with any other subjects or with none. In addition to this level of flexibility, it is also a part of the design that the examinations can be taken by candidates of any age following more or less any kind of course – or, again, none.

The international dimension of the IB colours much of the subject matter studied. The range of possible texts for the English literature examination is vast and highly international, and there is a requirement that some of the literature is studied in translation. Cross-cultural themes abound. The A level examination in this subject is much more focused on literature in English, with literature in translation only permitted, at most, a subsidiary role. There are dimensions of the study of literature from other English speaking cultures, but these do not really create a powerfully international dimension to the examination.

Similarly, the IB history syllabus is structured around world history, requiring a much broader sweep of study than the A level. To accommodate this greater scope, the topics are very modern (starting essentially with the 19th Century). The A level offers much more tightly focussed topics – a handful of years in some cases – and has a compulsory element of British history. Some of the areas that the IB includes – the Asias, for example – are absent or almost wholly so in the A level.

The other two subjects are much more universal in character anyway. The general content of mathematics and chemistry do not vary greatly from country to country. However, there is a key difference in the mathematics syllabuses. The IB examination concentrates almost entirely on pure mathematics, with a small optional element, which can itself be more pure mathematics. The A level, reflecting its need to cater for a much more diverse potential client group, has a highly complex and flexible structure. The compulsory core is much smaller – indeed, if the full range of possible mathematics A levels is included, almost non-existent. A particular difference is the complete absence of any mechanics in the IB. The basic rationale for this seems to be the need to cater for the core mathematics requirements in a wide range of countries, almost always defined in terms of pure mathematics.⁷ These, and the range of applied mathematics options that are offered are deemed sufficiently demanding of themselves.

Conversely, the A level attempts to provide a mathematics course which can be customised to meet the needs of a wide range of different users. Two of the key features of the specification are that all units are equally weighted, allowing many different combinations of units and greater flexibility and that there are routes leading to AS qualifications in Mathematics, Further Mathematics, Pure Mathematics, Statistics, Mechanics and Applied Mathematics and A levels in Mathematics, Further Mathematics, Pure Mathematics and Statistics.

Not only do IB students need to be highly committed and internationally minded, the institutions they attend do too. It is here that regulation becomes relevant. Although the IB is now an accredited UK qualification, and thus subject to some of the aspects of regulation, the IBO has far more freedom of manoeuvre than awarding bodies offering A levels, although it is clearly constrained by its need to sustain its appeal to an international clientele. In part, this relative freedom stems from the fact that there are several different awarding bodies offering A levels. There is an obvious need in such a system for some guarantee of comparability between the various specifications in a subject. (There are interesting parallel but distinct expectations for comparability in a baccalaureate system, but these are much more complex.)

⁷ The IBO also feels that there would be too much overlap with Physics, and its substantial element of mechanics, if it also included mechanics in its mathematics course.

To achieve such comparability, there is a system of regulation used in England, Wales and Northern Ireland, with bodies appointed specifically for that task.⁸ The structure of regulation requires each specification to comply with a number of external constraints, reducing the possibility of unreasonable variation across specifications. There are general criteria with which all awarding bodies, including the IBO, have to comply. There are qualification-specific criteria, which in the case of A levels lay down the general principles for the construction of an A level: AS/A2 equally weighted, 6 units, weighted between 15 and 20% each, the need for at least 20% of the assessment to be synoptic in nature etc: to which all A levels must conform. Then there are subject specific criteria, which set out required assessment objectives – usually with specific weightings attached – prescribed content, limits on internal assessment, grade descriptions etc. For any specification in a given subject to be accredited for use in the UK it must conform to all these sets of criteria.

What is more, this conformity must be evident in the assessment. The accreditation process includes the evaluation of sample assessment materials and allows the regulatory authorities responsible for accreditation to ensure that candidates taking the examination will actually be tested on what is specified. In some subjects, this is relatively straightforward: each unit assesses some part of the subject content and it is relatively easy to confirm that any question paper provides sufficient coverage of that content and doesn't stray beyond what is specified. In others, conformity with requirements is harder to ensure. The balance of assessment objectives tested by an essay question in English, for example, is much harder to control. It was interesting to hear the A level examiners in both English and history speaking, with some enthusiasm, about the extent to which they could be confident that assessment was at some level consistent for all candidates, whenever they took their papers and whichever combination of papers they studied and questions they answered. (To be fair, the historians were also acutely aware of what had been lost in the assessment of their subject to achieve this. It is also important to record that the HE representatives in each subject did not find this a particularly important feature in the examination, happily accepting the uncertainties of the IB.)

By contrast, the IB makes clear what it expects of its centres in the subject guides and associated materials, but it makes much less effort to police these requirements. The scale of assessment across six separate subjects is inevitably rather less than at A level, typically involving something over an hour less assessment per subject. The approach was therefore rather more one which sampled the range of content studied. In the case of the English literature examination, the subject guide specified that candidates should study at least 15 books from the various booklists, according to some clear rules of combination. However, it was not at all clear that there was any way of checking whether this number of books had, in fact, been studied. In other subjects, the way the examination was structured meant that there was a risk of question and topic spotting (by student or teacher) leading to an actual course of study significantly narrower than prescribed. The IBO is extremely active in the support and guidance it gives to its centres, to make sure that they understand the spirit of the qualification, but it cannot necessarily control the activities of these centres.

One example from mathematics illustrates neatly both the differences in the way in which regulation operates in the two qualifications and the underlying assumptions this implies about the client group. It concerns the approach taken to calculators. In the IB, graphical calculators are essential, and the papers are designed to include the use of such calculators in the overall assessment. The IBO has to specify the minimum requirements needed for a candidate to be able to access the papers, while the paper setters need to ensure on the one hand that all questions can be answered within this specification and that those calculators which exceed it do not allow an unfair advantage to candidates using them.

⁸ These are QCA for England, ACCAC for Wales and CCEA for Northern Ireland, who adopt a common approach except where there is a need to accommodate differing national requirements.

Conversely, in the A level, although calculators are allowed, the maximum permitted specification is what is regulated – and relatively restricted – and the use of graphic calculators is explicitly forbidden. This is a central regulation applying to all mathematics specifications, and the regulators have to ensure there is an up-to-date list of permitted calculators available to all centres entering candidates for the examination.⁹ The philosophy behind such an approach is that no candidate should be disadvantaged by lack of access to resources and that means that, without banning their use altogether, calculators must be restricted to the lowest common factor.

Examination performance

There were several interesting features underlying the generally consistent set of outcomes of this part of the study. The first was that, although the participants for one or other of the qualifications were usually in agreement with each other about where the boundaries for both their own and the other qualification should lie, there were occasions where one of them was a little out of line with the others. However, here the tendency was for the quality to be seen consistently, so that it did not call into question the overall finding.

A second was the extent to which the HE reviewers did, as expected, have somewhat different expectations of potential students. This ranged from a participant from a highly selective university, whose expectations were consistently above the highest grade available, to those who were able to see potential in performances meriting only very poor actual grades. Here, too, the differences did not call the overall findings into question, because the pattern tended to be consistent for an individual reviewer.

Limitations of the study

There are several aspects of the study which mean that any outcomes need to be treated with a degree of caution.

The first of these has been acknowledged throughout the body of this report: it is the failure to address the nature of the IB Diploma as an umbrella qualification awarded for successfully completing a programme of study as compared to A levels which are single-subject qualifications awarded for demonstrating sufficient attainment against defined objectives. One important issue here is that all participants were made aware of the difference from the outset. A full briefing about the nature of the IB diploma was provided, although it was stressed that the focus of their work was on their particular subject. Participants reported that they found it difficult wholly to ignore their knowledge of the overall demands made on IB candidates, although they agreed that it was hard to find a way formally to include the knowledge. It is possible, therefore, that estimates of the demand made by the IB were affected by this awareness.

A further obvious limitation on this work is its scope. It looked only at the HL examinations in four subjects, albeit covering four cells of the IB hexagon. It is dangerous to extrapolate from any findings here to other subjects and still more so to SL. However, the relative stability of the findings across subjects is reassuring that at least at HL there can be sufficient confidence that there is a degree of comparability between the qualifications, both in the nature of the demands and levels of performance required for particular rewards, at least at the level of the individual subject.

⁹ For the purposes of this report, the term centres is used in the technical sense of an institution permitted to enter candidates for examinations. It covers schools and colleges, but also a wide range of other institutions such as hospitals and prisons.

The consistency is reassuring, but in all cases it must be remembered that there were significant differences between the syllabuses and the associated assessment across the two qualifications. This meant that for most subjects there were particular aspects where there could be less confidence. The case of the assessment of practical skills in chemistry has already been mentioned. In that subject, also, the IB examination allowed some degree of optionality and sampled the subject content while the A level comprised compulsory units, each of which assessed the full content specified using compulsory questions.¹⁰ The effects of choice of topic and sampling are very difficult to estimate.

The approaches the two qualifications take to assessment in both history and English have also already been alluded to as being a significant difference. The need to ensure that any given examination paper assesses given assessment objectives in a given ratio has led to A levels in these subjects using more, but shorter, questions with fewer of the more familiar essay questions, although these remain. IB examinations in these subjects, where the prime form of assessment remains the open-ended essay question, would have been easier to compare to legacy A levels. These changes perhaps made it harder for IB participants (and those from HE who were not yet familiar with the revised A levels) to evaluate the demands that the A level papers presented to the candidates. One would guess that for the A level examiners, all of whom had substantial experience on legacy syllabuses, would have found it easier to review the IB papers.

In mathematics, too, the form of assessment differed markedly between the two qualifications. For example, in the IB, there is a compulsory internally assessed component; in the A level there is virtually no internal assessment and none in the units that the work focussed on. Perhaps of more impact, the question papers themselves also differed markedly, as well as the difference inherent in a series of shorter module-based papers against two terminal examinations. The IB has two papers, the first of which is a series of relatively short questions covering a wide range of the subject content. The second comprises much fewer much longer questions sampling a proportion of the content. The A level papers are all much more homogeneous in nature. Each unit, whether AS or A2, comprises a number of medium length questions covering the subject content specified for the unit. The exact implications for candidates of these quite major differences in approach are clearly hard to evaluate.

A further problematic aspect of the study derives from the structure of A levels post 2000. Comparing a modular examination with a linear one is inevitably problematic: how much easier is it for a candidate to be assessed on much more focussed content and much closer in time to when it has been taught? Conversely, how much does a candidate gain in knowledge and understanding of any topic from the opportunity to see it in the context of the full content?

These imponderables apply to any modular/non-modular comparisons. There are particular aspects of the revised A levels that participants felt made their judgements difficult. First, there is the fact that there is no prescribed time or order for taking the units. Although the qualifications was designed so that the AS could and, perhaps, should be taken during the first year of a two-year course, there is no requirement for this: some centres choose to enter candidates for all their units at the end of the course, effectively making it a linear examination.¹¹

The availability of re-sits is a further complicating factor. There is evidence that the influence of re-sits is much smaller than may be assumed, and that any improvement for most candidates typically only takes them from underperformance to their predicted outcome. However, the possibility of re-sitting clearly played some role in participants' perceptions of the demands of the A level system.

¹⁰ Other A level specifications would have been more like the IB in this respect, in that some offer a choice of optional units. However, the policy was to use the largest entry specification not the most similar.

¹¹ This refers to two-year candidates. The lack of compulsion derives in part from the need for the qualification to be available to other types of student, such as those pursuing one-year courses.

The division of the A level into AS and A2 units at intentionally differing standards also presented difficulties to the design of the study. These have already been mentioned and their resolutions identified in the account of the methodology, but it is important to re-state the issues here. On the whole, in terms of subject content, the matter is relatively unproblematic. The total content of the AS and A2 are the parallel of the content in an IB HL subject guide – or at least it was appropriate that this is what the study should seek to establish.

What needs to be compared when considering the assessment burden is probably equally easy to agree. It is, though, much harder to evaluate effectively. It is inevitable that one should look at all the assessments that a candidate in a particular subject takes to get a feel for the overall assessment burden. However, there were real and predictable problems for participants in deciding what value to place on AS examinations that were not intended for candidates at the same point of progress as candidates taking their IB papers, still less to build that in to an overview of the overall assessment burden for an A level candidate.

The effect of the new structure on the study was most significant when it came to reviewing standards of performance. Ideally, one would consider the whole work of candidates achieving particular scores. However, the logistical demands of trying to meet the ideal were quite intractable. (It must be remembered that there are no pre-set or agreed points where the two grading scales are expected to coincide, meaning that a wide range of performance needed to be available.) For the A level, with its modular structure, in particular, it is almost impossible to put together all of a candidate's work, without requiring the awarding bodies to hold on to all scripts (and possibly coursework) for at least a two-year period. Instead a pragmatic view was taken that the study should focus on performance on individual units/components from each examination. There is some theoretical justification for adopting this approach: grading decisions are originally made at this level in both qualifications.

A consequence of this decision is that there could be no assumptions about the final grade obtained by of the candidates whose work was used for the study.¹² In turn, this made it easier to decide to use only A2 units. Only these units are graded on the underlying assumption that the candidates are at the end of a two-year course, ie in an equivalent position to IB candidates when they take their examinations. However, it is clearly the case that the impact of the AS on overall grading standards rather than those expected by awarders at the end of a course is unknown. It is also true that differences in how marks are aggregated and overall subject grade boundaries determined may also affect the level of comparability at the subject level. All that can really be said is that the levels of exactness that the study aspired to mean that the outcomes remain useful insights not precise statements.

¹² For IB candidates, of course, their final result in the Diploma is a further stage removed.