

# Five year review of standards

A level psychology

# Introduction

Every summer, the publication of GCSE and A level examination results prompts public interest in the standards of those examinations.

In 1996, Lord Dearing in his *Review of Qualifications for 16–19 Year Olds* made several recommendations to ensure that ‘there is a basis and accepted procedure ... for monitoring and safeguarding standards over time’. In the same year, SCAA (one of QCA’s predecessors) and the Office for Standards in Education jointly investigated standards in English, mathematics and science (chemistry) in 16+ and 18+ public examinations over time.<sup>1</sup>

The outcomes of this work were published in *Standards in Public Examinations 1975 to 1995*. One of the recommendations was that there should be:

‘... a rolling programme of reviews on a five-year cycle to ensure examination demands and grade standards are being maintained in all major subjects. Physics, history, French and German should be included in the programme at an early stage.’

The five-yearly review of standards programme is a response to these recommendations. It is run by QCA in collaboration with the regulatory authorities for Wales and Northern Ireland, ACCAC and CCEA, and is designed to investigate the standards in A level and GCSE examinations. It aims to find out if:

the demand of syllabuses and their assessment instruments has changed over the last 20 years (examination demand);

the level of performance required of candidates at grade boundaries has changed over the last 20 years (grade standard).

Organised to run in five-year cycles, the programme was structured to cover every major subject during its first cycle. Each year, up to 100 independent specialists review around 2,000 exam scripts, drawn from all the awarding bodies, together with their associated syllabuses, question papers and mark schemes.<sup>2</sup>

---

<sup>1</sup> 16+ examinations cover GCE O level and Certificate of Secondary Education (up to 1987), and GCSE (from 1988).

<sup>2</sup> For the purposes of this report, the general term *awarding bodies* is used to cover both the A level examination boards and the GCSE examining groups.

# Methodology

Each study was organised in two stages:

- stage one – investigating changes in examination demand;
- stage two – investigating changes in standards of performance.

Each covered four sample years: the year of the study and its predecessors from five years, 10 years and 20 years earlier.

## **Stage one: examination demand**

### ***Aim***

The aim of this review was to establish whether the demand of syllabuses and their assessment instruments changed over the period of the review.

### ***Evidence base***

The awarding bodies were asked to supply, for each subject, copies of one major syllabus from the most recent year and its predecessors for the other three years in the study. They were also asked to provide the related question papers, mark schemes, examiners' reports, and details of the procedures in operation at the time of each examination.

In general, syllabuses and question papers were available from all awarding bodies for all years in a study. Unfortunately, prior to 1988, few mark schemes and few documented details about awarding procedures had been retained.

### ***The process***

A coordinator and three reviewers – independent experts from a variety of backgrounds – were appointed for each subject. Each coordinator was given a framework and asked to use it to describe the main differences between the syllabuses from the different years. This description was given to the reviewers, who were asked to study the syllabuses, question papers and mark schemes and independently judge whether the differences between years affected the demand of the examination. After the material had been reviewed, the team for each subject area met and discussed any issues. The coordinator then reported on the findings and identified any conclusions.

## **Stage two: standards of performance**

### ***Aim***

The aim of the second stage was to find out if the level of performance required of candidates at grade boundaries has changed over the period of the study. The review focused on the performance of candidates at grades A and E at A level, and grades A, C and, sometimes, F for 16+ examinations.

### ***Evidence base***

The awarding bodies were asked to provide 15 examples of candidates' work at the defined boundaries for each syllabus studied in stage one. They were asked to submit the complete examination work of candidates, including all examination papers, coursework and any oral examinations.

On the whole, the samples provided for the most recent year of each study were complete. However, the coursework was sometimes missing and work from modular syllabuses presented a problem, in that it was seldom possible to provide the entire work of individual candidates. Usually, several modules from one candidate were provided, supplemented by modules from other candidates to produce the appropriate overall result.

Samples of work from earlier years were much less complete. The awarding bodies could rarely provide work from enough candidates or did not have the complete work of candidates – coursework and orals were usually missing and the work consisted of individual components. No work from the earliest year of the reviews was available.

### ***The process***

A team of up to 12 reviewers was recruited for each subject. The reviewers came from a variety of backgrounds, including universities, selective and non-selective schools, maintained and independent schools, and further education institutions (including sixth form colleges). Some of them had backgrounds working for the various awarding bodies.

The coordinator from stage one was used again in this stage and the syllabus reviewers normally participated.

The review took place over two days. Before the meeting, each coordinator produced a general description of the standards expected for the grade boundaries in the study. Where these were available, published grade descriptions normally formed the basis of the performance descriptors. The coordinators were asked to take into account the fact that they would be looking at borderline performance rather than that comfortably in grade which is the intention of grade descriptions. The performance descriptors were discussed and agreed by the team at the start of the meeting.

Reviewers were each given a batch of scripts for a particular year, grade and awarding body. Working independently, they were asked to judge if the scripts matched the agreed grade description. They could categorise the work as:

- above the expected standard;
- slightly above the expected standard;
- at the expected standard;
- slightly below the expected standard;
- below the expected standard.

They were then given another batch of scripts of the same grade, either from another awarding body or of a different year from the same awarding body. They categorised these scripts and compared them with the first batch to identify any significant differences between candidates' performance. A sampling framework ensured adequate coverage of the sample. A copy of part of one framework is provided on page 5.

At the end of the two days, a plenary session was held and the reviewers discussed their findings and any significant issues. As with stage one, the coordinator reported on the findings and conclusions.

### **Limitations of the study**

Comparing examination standards over time is a complex task, heavily dependent on the evidence available and the ability of reviewers to make valid judgements on it. When considering the findings and conclusions, several limitations need to be kept in mind.

### **Changes in syllabus and examination content**

In some subject areas, syllabuses and examination papers changed radically over the period of the review. For example, in assessing modern foreign languages the relative importance of the skills of reading, writing, speaking and listening has changed considerably. Fundamental changes make it difficult for reviewers to make valid judgements about relative standards because they are not comparing like with like.

### **Individual opinion**

Each individual places different values on each part of a subject. Agreed definitions of standards and frameworks show reviewers the standards they should work to, but it is difficult for them to avoid applying their own values. This can lead to differences in opinion about the same syllabus or piece of candidate's work.

### **Lack of evidence**

While reviewers had syllabuses and examination papers (although not always mark schemes) for all the years in the study, they did not have all the evidence they needed to analyse standards of performance. The archiving practices of the awarding bodies vary, each keeping different amounts of evidence for any year. This applies particularly to examination scripts. What tended to be available from earlier years is work for separate components of the examination rather than the whole work of candidates. Coursework and any oral examinations were usually missing.

A national archive of essential evidence on examination standards has been established by the regulatory authorities. This should ensure that difficulties in this area are reduced in future studies.

**Table 1: Sampling framework for part of a typical A level study**

**DAY 1**

8:30	BOARD A, GRADE A	BOARD A, GRADE E	BOARD F, GRADE A	BOARD F, GRADE E	BOARD C, GRADE A	BOARD C, GRADE E
10:00	1996 1-7	1996 1-7	1996 1-7	1996 7-1	1996 1-7	1996 15-8
10:10	BOARD A, GRADE A	BOARD A, GRADE E	BOARD F, GRADE E	BOARD F, GRADE A	BOARD C, GRADE A	BOARD C, GRADE E
11:30	1991 1-3	1991 1-3	1996 8-15	1996 7-1	1991 1-7	1991 15-8
11:50	BOARD A, GRADE E	BOARD A, GRADE A	BOARD C, GRADE E	BOARD C, GRADE A	BOARD E, GRADE A	BOARD D, GRADE A
1:05	1996 1-7	1996 15-8	1996 1-7	1996 8-15	1996 1-7	1996 15-8
2:15	BOARD A, GRADE E	BOARD A, GRADE A	BOARD A, GRADE E	BOARD B, GRADE E	BOARD E, GRADE E	BOARD D, GRADE E
3:30	1991 1-3	1991 3-1	1996 15-8	1996 15-8	1996 1-7	1996 15-8
3:30	BOARD B, GRADE A	BOARD D, GRADE E	BOARD B, GRADE A	BOARD D, GRADE E	BOARD D, GRADE A	BOARD E, GRADE A
4:45	1996 1-7	1996 1-7	1996 15-8	1991 4-1	1996 7-1	1996 8-15
5:05	BOARD B, GRADE E	BOARD D, GRADE E	BOARD B, GRADE E	BOARD D, GRADE E	BOARD D, GRADE E	BOARD E, GRADE A
6:20	1996 1-7	1991 1-4	1996 8-15	1986 4-1	1996 8-15	1991 1-3

**DAY 2**

8:30	BOARD C, GRADE E	BOARD E, GRADE E	BOARD E, GRADE A	EDEC , GRADE A 1996	BOARD F, GRADE A	BOARD A, GRADE E
9:45	1996 7-1	1996 15-8	1996 1-7	7-1	1996 8-15	1996 15-8
9:45	BOARD C, GRADE E	BOARD E, GRADE E	BOARD E, GRADE A	BOARD B, GRADE E	BOARD F, GRADE E	BOARD A, GRADE E
11:00	1991 1-7	1991 3-1	1991 3-1	1996 8-15	1996 8-15	1986 7-1
11:20	BOARD C, GRADE A	BOARD E, GRADE A	BOARD E, GRADE E	BOARD E, GRADE A	BOARD C, GRADE A	BOARD A, GRADE A
12:35	1996 7-1	1996 7-1	1996 8-15	1996 8-15	1996 15-8	1996 1-7
1:45	BOARD C, GRADE A	BOARD E, GRADE A	BOARD E, GRADE E	BOARD E, GRADE A	BOARD C, GRADE A	BOARD A, GRADE A
3:00	1991 7-1	1991 1-3	1991 1-3	1991 3-1	1991 15-8	1991 3-1

# A level psychology: review of standards 1977–97

## **Introduction**

Substantial changes have been made to A level Psychology since its inception in 1970, some as a result of the subject's very success. In 1977, there was only one syllabus in the subject. In 1987 and 1992, there were two and in 1997, there were three: more if the choice between modular and linear routes is taken into account. The initial examination entry in 1972 was 275 candidates. By 1997, the number of candidates was 28,000.

In 1997, 80 per cent of candidates gained grade E or above, with 11 per cent gaining grade A.

## **Examination demands**

### ***Materials available***

The reviewers considered a range of syllabuses, question papers and mark schemes from 1977, 1987, 1992 and 1997. Full details of the materials used in the review are given in Annex A.

Between them, the syllabuses represented in this study accounted for all the 28,000 candidates for A level Psychology examinations in 1997.

### ***Aims***

AQA/A was for many years the only awarding body to offer A level Psychology. The philosophy of the examination, as well as its assessment techniques, derived from those who devised the syllabus. In 1997, three awarding bodies offered syllabuses in the subject and clear differences in philosophy existed. AQA/NEAB included the subject in its suite of science syllabuses, while AQA/A and OCR classed it as a social science.

### ***Focus***

In 1977, the syllabus and examinations tended to focus on knowledge and understanding. More recent syllabuses, questions and mark schemes stated explicitly that evaluation and critical analysis should be demonstrated by candidates.

### ***Assessment objectives***

Since the introduction of Psychology at A level, there has been a clear trend to greater explicitness in assessment objectives and their weightings. Very few assessment objectives were stated in the earlier syllabuses, and even fewer weightings were given. In 1987, 1992 and 1997, AQA/A had six objectives with no weightings indicated. In 1997, AQA/NEAB divided its objectives into three main categories: knowledge and understanding, application and analysis, and

investigation. These were subdivided into four, five and six sub-objectives respectively. The weightings given to the main objectives were 50 per cent, 30 per cent and 20 per cent, with these weightings reflected in the mark schemes. In 1997, the OCR assessment objectives and weightings were knowledge 30 per cent, understanding 20 per cent, analysis 10 per cent, evaluation 10 per cent, application 10 per cent, research skills and methodological awareness 20 per cent. The differences were judged to be more presentational than substantive, but they made it harder to be sure that the syllabuses were truly comparable.

### ***Assessment pattern***

In 1977, there were two essay-based examination papers. The examination also contained a coursework element. In 1997, both OCR and AQA/NEAB offered modular assessment in the subject. The modular syllabuses thus offer opportunities for assessment throughout the course rather than at its completion. This was not considered to affect the level of demand.

### ***Assessment time***

There is general comparability between awarding bodies in the approach to assessment, but the external examining time varies considerably. The AQA/NEAB modular scheme takes nine hours; its end-of-course syllabus six hours; OCR takes seven and a half hours and AQA/A six hours.

### ***Question types***

The structure of questions also changed. The 1977 papers consisted solely of essay questions. In 1987, these were supplemented by a range of question types that had two or more distinct subsections, each with its own mark allocation. Structured questions have two distinct effects: they enable candidates to apportion their efforts according to the marks available, and they help to make clear the full demands of the question.

There was also a growing tendency between 1977 and 1997 to provide stimulus material on which to base the structured questions. This material was used to require candidates to apply knowledge and understanding to an unfamiliar situation. This helped to avoid stereotyped questions and formulaic responses, but involved a considerable increase in reading time and cross-referencing. OCR was the main exponent of questions using stimulus material.

In the OCR syllabus, however, the potential for creativity such stimulus material offered tended to be negated by providing candidates with hints for each question as to how the question might be answered. In addition, some of the papers were seen by the candidates one week in advance of the examination in order that material could be prepared. Since neither of these two potentially helpful initiatives was extended to candidates by the other awarding bodies, there was judged to be variation in demand between syllabuses.



### ***Mark schemes***

In the early days of A level Psychology, the mark schemes were very limited in scope. Over the period of the review, their level of detail and complexity increased significantly. In part, the inclusion of extra detail coincided with their publication, which made them important source materials for both teachers and learners.

As a result, candidates in 1997 had more opportunity to be clear about what was needed than their predecessors. However, the AQA/A mark schemes were highly complex and their comprehensiveness and length made them inaccessible. These mark schemes, though over-complicated at first glance, were logical and flexible in their application to various different structures of questions, adaptable in dealing with responses that give partial answers, and were judged to be effective operationally.

All the 1997 mark schemes contained reference to content and skills. However, while both OCR and AQA/A had marks set against assessment objectives for each question, this was not evident in the AQA/NEAB scheme.

### ***Coursework and practical assessment***

Although the weighting for practical assessment between awarding bodies and across time was consistently at or around 20 per cent, the nature of the assessment changed significantly. Demand across the awarding bodies was brought into line, and the nature of the coursework assessment was made commensurate with the weight it carried. Thus AQA/A progressively reduced its requirement for practical activity and statistical and methodological understanding, but increased its demands for detailed report writing. In 1977, candidates had to prepare 12 reports. An external examiner then conducted a half-hour oral examination of candidates' understanding of statistics and methodology using these reports as the basis for the questions. The removal of the oral examination led to a dramatic change in the nature and demand of practical assessment. The number of reports required dropped in 1992 and again in 1997. By 1997, just two were required, although these were expected to be fuller. The reports were internally assessed by teachers and externally moderated. There was no oral assessment.

In 1997, AQA/A and OCR had a similar approach to coursework, requiring the writing of reports to a standard format. Both attempted to assess skills by inference whereas AQA/NEAB assessed them by observation.

### ***Optional routes***

In some cases in 1997, papers were so constructed that candidates could choose to answer structured questions only or essay questions only, or a combination of the two. This made it hard for awarding bodies to ensure parity of demand across the possible combinations of questions. This difficulty is also shown by the fact that OCR, where little choice was possible, required even the lower attainers to demonstrate skills of critical analysis to obtain a grade, whereas the other awarding bodies allowed candidates to obtain up to 64 per cent of their marks on knowledge and understanding.

Moreover, AQA/A, and to a lesser extent AQA/NEAB, offered a substantial number of options on written papers. In some options, the questions set tended to be somewhat predictable and formulaic. This did not apply to OCR, where there was

also very little question choice. This is another source of variation of demand between syllabuses. AQA/NEAB and AQA/A tended to have many options in their syllabuses but candidates were able to select options from a narrow range; OCR required candidates to cover practically all of the syllabus but not necessarily in such depth.

These differences in approach mean that, overall, OCR was considered less demanding on knowledge but more demanding on skills than the other two awarding bodies.

### ***Quality of language***

There is one other difference between the 1997 syllabuses: the approach to the assessment of quality of language. AQA/A awarded this element separately, allocating up to five per cent of the total component mark; the others assessed it within the marking criteria for each question. It was hard to be sure whether the two approaches produced the same outcome.

## **Summary**

Substantial changes occurred in A level Psychology syllabuses over the period. These were reflected in the associated question papers and mark schemes, where both content and question types have evolved to make the demands of the examination more explicit to teachers and candidates. This did not affect the level of these demands.

The requirements of coursework had been reduced from 1977, but in 1997 more fairly reflected the weight attached to the component.

There were significant differences between the 1997 syllabuses. These were found in the content that had to be learned; the options that were offered; the duration of the written examinations; and the way that questions were presented. The complex nature of these differences made it hard to judge their overall effect.

## **Standards of performance at grade A and grade E**

### ***Materials available***

The script reviewers had a wide range of scripts available from 1997, with a much more limited selection from 1992/3 and 1987/8. Full details are given in Annex A.

There was some doubt whether all the scripts had been selected according to the same specification, making the task of comparison even more difficult.

The descriptions of expected performance used in this exercise were developed from existing grade descriptions, adjusted to take into account the fact that the work was from borderline candidates.

## **Standards expected at grade A**

Candidates demonstrate and communicate clearly relevant, accurate and detailed knowledge and critical understanding of a range of psychological concepts, theories,

studies, research and applications. They show effective analysis and evaluation when considering psychological concepts, theories, studies, research and applications, and take an eclectic approach to the subject. If required, they demonstrate that they are able to use attributes by applying them appropriately to unfamiliar situations. Candidates demonstrate an ability to plan and conduct psychological investigations in an effective manner to arrive at a balanced conclusion as the result of well-constructed discussion.

## **Performance at grade A**

Reviewers considered that the standard of performance in grade A scripts was consistent across the years and between awarding bodies, but the way that standard was obtained varied. AQA/A scripts in 1988 were regarded as being of a very high level in their demonstration of detailed knowledge and critical understanding. Evaluation was less evident, partly because the questions did not always explicitly ask for evaluative skills. In 1993, more questions had a specific requirement for evaluation or critical analysis, but it was still possible for candidates to obtain a grade A with a predominantly descriptive approach. In 1997, candidates obtained their A grades by demonstrating the additional skills of evaluation and critical analysis. There was some evidence, however, that the evaluation tended to be second hand and not the candidates' own.

OCR scripts at grade A generally showed good knowledge and understanding both of a wide range of studies and of some theories and concepts that were applied to problems based on the stimulus material provided. The skills of application, analysis and evaluation were demonstrated at a high standard; very good essay writing skills were displayed; the writing was comprehensive. However, answers tended to be formulaic, containing material that had been insufficiently assimilated from set texts and similar ideas and points of evaluation, analysis and evidence. The occurrence of these features may have been exacerbated by the hints for possible content that were provided at the end of each question. The effect of these hints may thus have been to lower performance rather than improve it.

AQA/NEAB candidates had many opportunities to display problem-solving skills if they selected those questions that related psychological theories to real-life situations. Questions that placed candidates in an unfamiliar situation had the merit of lessening, though not altogether preventing, routine answers, but such questions also meant that some scripts showed a lack depth of knowledge and understanding. Candidates of this awarding body were less successful in using psychological evidence to support their responses than those from AQA/A or OCR.

The reviewers felt that marking schemes had affected performance. In particular, the relatively rigid assessment used by AQA/NEAB allowed little credit for partially correct answers. AQA/NEAB and AQA/A tended to use a similar assessment format for all their written examinations, whereas OCR had three different forms of assessment. The OCR approach was considered more challenging, but fairer.

The formulaic answers aroused much comment among the reviewers, although the practice was not universally condemned. There was perceived to be a tension

between the need to have teaching materials and guidance available and the tendency for teachers and learners to become over-dependent on a single source.

### **Standards expected at grade E**

Candidates demonstrate and communicate some knowledge and understanding of appropriate psychological concepts, theories, studies, research and applications. Alternatively, they may analyse or evaluate appropriate psychological concepts, theories, studies, research and applications. Or they may demonstrate a limited combination of the above knowledge and understanding, and analysis or evaluation. They make some appropriate use of psychological terminology. Candidates demonstrate that they have conducted psychological investigations and have come to simple conclusions.

### **Performance at grade E**

There was less agreement amongst reviewers about the scripts at grade E. There were several aspects to this, not least of which were concerns about whether all the work at this grade conformed to the specification. Differences between the assessment techniques were considered to have had more impact at this boundary. The inflexible mark scheme used by the AQA/NEAB almost certainly counted more against borderline grade E candidates than against those who obtained higher marks. The need for OCR candidates to cover the whole of the syllabus was more demanding, but the hints supplied at the end of each question and the use of pre-released materials were likely to assist candidates at grade E. Even the different examining times were considered likely to have been a factor where marginal grade E candidates were concerned. Added to these problems was the tendency of candidates at this level to perform very inconsistently, performing weakly in three essays out of four, but scoring highly on the fourth.

Overall, an interaction between examination paper demand and candidate performance was identified at this grade. Demand was higher for AQA/NEAB and OCR candidates. Consequently, some OCR candidates produced work that was of a very poor standard, while the AQA/NEAB scripts were considered to be weakest of the three. Even AQA/A scripts, where the demands were lowest, contained in all years too much description at the expense of evaluation.

### **Summary of performance**

Reviewers were agreed that scripts at grade A were of a comparable standard between the years and across the awarding bodies.

In performances at grade E, there were differences identified between the awarding bodies though not over time. There was little work that met the expected standard. However, more than at grade A performance at this grade was affected by the different syllabus demands, which made judgements, especially comparative ones, especially difficult.

## Annex A: Materials used in the review.

Table A1 shows the materials available for the review of examination demand.

Awarding body	AQA/A	AQA/NEAB	OCR
<b>1997</b>			
Syllabus	✓	✓	✓
Question papers	✓	✓	✓
Mark scheme	✓	✓	✓
<b>1992</b>			
Syllabus	✓	✓	
Question papers	✓	✓	
Mark scheme	✓		
<b>1987</b>			
Syllabus	✓	✓	
Question papers	✓	✓	
Mark scheme	✓		
<b>1977</b>			
Syllabus	✓		
Question papers	✓		
Mark scheme			

Table A1: materials available for the syllabus review

Table A2 shows the materials available for the script review.

Year	Grade	AQA/A	AQA/NEAB	OCR
<b>1997</b>	Grade A	12	12	13
	Grade E	12	12	13
<b>1993</b>	Grade A	5		
	Grade E	5		
<b>1988</b>	Grade A	5		
	Grade E	5		
<b>1977</b>	Grade A			
	Grade E			

Table A2: Numbers of sets of candidates' work available for the script review

### Notes:

The AQA/A scripts were from 1993 and 1988 rather than 1992 and 1987 respectively.

The AQA/NEAB scripts covered several modules.

# Key to the awarding bodies

During the period of the reviews, the number of awarding bodies operating fell. There are currently five: AQA, CCEA, Edexcel, OCR and WJEC. However, the three English awarding bodies came together through a number of mergers and a government requirement for unitary awarding bodies which could offer the range of GCSE, A level and GNVQ/VCE qualifications. This means that the qualifications used in the reviews came from a number of earlier examination boards and examining groups.

For the purposes of the reports the following abbreviations will be used:

AQA/A, AQA/N, CCEA, Edexcel, OCR and WJEC.

AQA/A covers AQA legacy A level syllabuses offered by AEB; legacy GCSE syllabuses offered by SEG; and O level syllabuses offered by AEB.

AQA/N covers AQA legacy A level syllabuses offered by NEAB, NEA and JMB; legacy GCSE syllabuses offered by NEAB and NEA; and O level syllabuses offered by JMB.

CCEA covers A level and GCSE syllabuses offered by CCEA, NISEAC and NISEC; and O level syllabuses offered by NISEC and NIGCEEB.

Edexcel covers A level and GCSE syllabuses offered by Edexcel, ULEAC and ULSEB; GCSE syllabuses offered by Edexcel, ULEAC and LEAG; and O level syllabuses offered by ULSEB.

OCR covers A level syllabuses offered by OCEAC, OCSEB, UCLES and UODLE; GCSE syllabuses offered by MEG; and O level syllabuses offered by OCSEB, UCLES and UODLE.

WJEC has retained the same name throughout the period.

© **Qualifications and Curriculum Authority (QCA) 2001**

83 Piccadilly

London W1J 8QA

**[www.qca.org.uk/](http://www.qca.org.uk/)**

***Order ref:*** QCA/01/763