

## **Comparing Public Examination Standards over Time**

Paper presented at BERA conference, September 1996 (ext/bera96/hp/8.96)

Helen Patrick

University of Cambridge Local Examinations Syndicate

**Note:** The views expressed in this paper are those of the author and are not to be taken as those of the University of Cambridge Local Examinations Syndicate.

### **Introduction**

For those of you who are not familiar with school examining in the UK, I will begin by providing a few explanatory notes.

The University of Cambridge Local Examinations Syndicate is a department of the University of Cambridge, one of whose functions is to provide external school examinations for 16 year olds and 18 year olds, in the form of GCSE and A level, or GCE, respectively. We have two UK arms: the Midland Examining Group (MEG), which provides GCSE, and the Oxford and Cambridge Examinations and Assessment Council (OCEAC), which provides A level. We are one of six organisations providing public examinations in England, Wales and Northern Ireland.

School examining is regulated by the government through the School Curriculum and Assessment Authority (SCAA), the Office for Standards in Education (OfSTED), the Independent Appeals Authority for School Examinations (IAASE), the Department for Education and Employment (DfEE) and, ultimately, by Parliament. Our licence to provide school examinations depends on satisfying the rules and regulations enforced by these bodies, and we also answer to a Syndicate of the University of Cambridge.

### **The background**

A level pass rates began rising in the mid 1980s. Although the movement from year to year was almost imperceptible, the cumulative effect gradually became apparent.

Concerns about standards over time were by no means new, as you will see from some of the literature referred to in this paper. Nor are pass rates much of a basis for judging what is happening to standards. But such statistics are part of the public face of the Boards' work and it looked as if we were seeing the start of a new trend.

The Boards' research staff therefore advised the Secretaries (the heads of the Boards) that we should start an archive of candidates' work, with accompanying syllabuses, question papers, mark schemes, etc, so that, in due course, we could make comparisons over time. The nine A level Boards which existed at that time already had individual archives of varying size and quality which had been kept for a range of purposes, but our trawls of these had revealed that material of a comparable nature from Board to Board did not exist. So we agreed jointly to keep comparable samples in the same subjects on a five year rolling programme, starting in 1989, to provide a basis for making comparisons over time at some future date.

Now I want to jump forward a few years. In the spring of 1995 SCAA and OfSTED announced that they were jointly going to carry out a study of examination standards over the past twenty years. The results are reportedly going to be published this autumn. The Boards contributed materials for the study, as far as they could, and a literature review, but have had little direct involvement.

The next development was that earlier this year the Secretary of State for Education asked the Boards to set up a national script archive, starting with the summer 1996 examinations, so that standards could better be monitored over time.

It was against this background that I agreed to give this paper at BERA.

### **How meaningful is it to compare standards over time?**

I am not going to regale you with anecdotes about the doom-mongers of history, who can be found in every age and every country, and who tell us, at regular intervals, that things are not what they used to be and that standards are falling. For an entertaining look at this approach, see Benjamin (1939). Nor am I going to attempt to define the term 'standards'. Instead what I propose to do is to look at the changing context within which public examination standards reside in an attempt to show that it is not meaningful to make comparisons of standards over any length of time.

The appendix to this paper includes extracts from English examination papers from 1906, exactly 90 years ago, and from 1951, mid way between 1906 and 1996 and the first year of GCE O level and A level. I chose English composition for comparison because it is easy to understand and, as a task, fairly similar in nature to what candidates might be asked to do today. The earlier examples have a relatively long time allowance because they include other items about writing English, such as punctuating and correcting sentences and explaining meanings.

These examples illustrate the futility of trying to make comparisons over long periods of time. Even the twenty years of the SCAA/OfSTED study has severe problems because the context within which these examinations are set, taken, marked and graded is so very different. Consider how much has changed.

I don't think that I need to spell out the social, demographic, cultural and technological changes which have taken place in the last twenty years. We have become increasingly highly educated and middle class, we've had the Falklands War and the Gulf War, we've had immigration and emigration, the basis of employment has moved away from manufacturing towards service industries, we have closer union with Europe, the Social Democrat Party has come and gone, our public utilities have been privatised, we have microwave ovens, PCs, electronic games, CD-ROMs, scientific calculators, the M25 and the National Lottery. We have the BERA programme on the Internet. This rather miscellaneous list illustrates how some aspects of our lives have changed.

Just think about education in 1976. We had a Labour government and Shirley Williams replaced Fred Mulley as Secretary of State for Education. The Prime Minister, James Callaghan, made his famous speech at Ruskin College, in which he complained that education failed to equip students with the skills they needed for the world of work and

the Assessment of Performance Unit was set up. Some of you will remember the suspicion with which that was regarded. Debate was raging over progressive versus traditional teaching methods, fuelled by Neville Bennett's book, *Teaching Styles and Pupil Progress*, and the William Tyndale enquiry. The Plowden Report on primary education was not yet ten years old. The school leaving age had recently been raised from 15 to 16, the Houghton pay award and the Bullock Report on language across the curriculum were only a year old. The Sex Discrimination Act had just come into force and the Race Relations Act was about to be passed.

Since then we've had substantial changes in education. We've seen the growth of sixth form colleges and FE colleges; the spread of comprehensive schools; the introduction of grant maintained status, local management of schools and increased powers for school governors; a decrease in the powers of local education authorities; new vocational qualifications - GNVQs and NVQs; and the National Curriculum. We've had about twenty Acts of Parliament relating to education (not counting those for Scotland). There has been a host of reports on education, including Warnock on children with special educational needs, Cockroft on Mathematics teaching, Swann on the education of West Indian children, Kingman on the teaching of English, Elton on discipline in schools.

There have also been several reports with a direct bearing on examinations and assessment - Waddell (1978) on school examinations, which heralded GCSE, Keohane (1979) on the Certificate of Extended Education, a post 16 qualification, the Task Group on Assessment and Testing (TGAT) (1987), which heralded the National Curriculum, Higginson (1988) on A levels, Dearing (1993) on revising the National Curriculum and, in 1996, Dearing again, this time on 16-19 education.

Public examinations are a small part of the education system. They change in order to reflect changes in education, which in turn reflect changes in society more generally. Changes in public examinations result from pressure from schools and, increasingly in recent years, from the government and its regulatory bodies (Murphy, 1993).

Twenty years ago school examinations were under the wing of the Schools Council, students took O level, aimed at the top 20% of the ability range, or CSE, aimed at the next 40% of the ability range, leaving 40% of 16 year olds for whom there was no specific national school leaving qualification. Less than 20% of the cohort stayed at school beyond 16, about 15% took A levels, and less than 10% went to university - there were only 46 universities in the UK. In England, Wales and Northern Ireland there were nine Boards offering GCE A levels and O levels and over 20 offering CSE.

The first joint 16+ examinations, combining O level and CSE, in which my Board was involved were taken in 1977. By 1986 substantial numbers of candidates took MEG examinations. MEG was one of the six new GCSE Groups set up in the 1980s in England, Wales and Northern Ireland to replace the CSE and O level Boards. In 1988 full blown National Criteria GCSE examinations started. In 1994 we had the first National Curriculum Key Stage 4 GCSEs. This month candidates have started preparing for KS4 Mark II, or GCSE Mark III, which they will take in 1998.

Statutory change at A level has not kept pace with what has happened at 16+, but there have been major structural changes. In the 1980s the Boards agreed a set of common

cores for A level subjects. These were revised in the 1990s and the first of the new syllabuses were examined this summer. No sooner had this process of revision got under way than the Dearing Review came along and now all A level syllabuses will be revised for the year 2000. Advanced Supplementary (AS), half an A level, was introduced in 1989 to try to broaden the sixth form curriculum and it too will be restructured for the year 2000 as a qualification intermediate between GCSE and A level. The number of A level Boards in England, Wales and Northern Ireland has been reduced from nine to six and the government is reported to have plans to reduce the number further.

Today GCSEs are taken by over 90% of 16 year olds and A levels by over 30% of 18 year olds. Over 30% of the age cohort go on to higher education and we have over 100 universities. The post 16 female education participation rate has overtaken the male and the male/female balance has changed markedly in some of the mainstream subjects. By the time 16 year olds reach GCSE, they have had National Curriculum tests at 14, and soon those tested at 7 and 11 will be coming up to GCSE.

Alongside these changes and in some cases directly related to them have come substantial curriculum changes. Statutory changes such as the introduction of the National Curriculum have had obvious effects, but even in 1988 GCSE was seen not just as a change to the 16+ examination system, but as a major curriculum development. Examinations have also reflected curriculum changes in a variety of subjects. Obvious examples are the 'new' Mathematics, the influence of the Cockcroft Report and of projects such as the School Mathematics Project and the increasing emphasis on data handling and statistics; the 'new' History, with its emphasis on the use of evidence; the growing importance of oral skills in languages, including English; the increasing emphasis on practical work in Science, influenced by projects such as Nuffield and Salters Science. Today all sixteen year olds must study Science and most do so in the form of Double Science. We teach subjects which were virtually unheard of twenty years ago - Business Studies, Media Studies, Computer Studies. Even where subjects are not new, some have changed out of all recognition. Woodwork, Metalwork and Technical Drawing have been replaced by Technology.

The nature of assessment has changed accordingly. We have coursework in most subjects at both GCSE and A level, multiple choice has almost disappeared, the assessment of practical work, including oral skills, scientific experimentation, designing and making artefacts for Technology, is commonplace. A level modular syllabuses have overtaken linear syllabuses in some subjects. Examinations have to be designed to cater for a much greater ability range than ever before.

Examination Boards have become more open and accountable. We offer teachers a range of help that was unheard of twenty years ago. Syllabuses are no longer lists of set books or content headings, but include assessment objectives and criteria, along with detailed information about skills, concepts and knowledge to be tested. We publish mark schemes and examples of marked work with explanations of why it gained the marks it did, we provide more information about marks, for example, course work grade boundaries, component and module grades and marks. We run in-service courses for teachers. Our examination procedures are subject to Codes of Practice and we are

regulated by SCAA, OfSTED and the IAASE. Many of our senior examining personnel work in schools and colleges and see our work at first hand at grading meetings.

There is no need to prolong this catalogue to demonstrate that the proposition that we can compare standards in 1996 with those pertaining in 1976 has to be meaningless. We're not assessing the same thing, we're not assessing it by the same methods and even if we were we would be doing so in quite a different context, which of itself would change the very thing we were assessing. To run a mile in four minutes in 1996 is to meet the same criterion as it was when Bannister first did it in 1954. But now that we have improved training, better running shoes, faster tracks, and more knowledge of the best kind of diet for athletes, what can we say about running standards? In one sense the standard is exactly the same. There has been no change to the length of a minute or to the length of a mile. But the context has changed enormously. Have standards risen or fallen? As soon as a criterion of this kind is established, people will do everything they can to achieve it. If more of them succeed, are standards rising or falling?

Why should we expect to be able to answer such a question? As Mortimore (1996) suggests, 'crime statistics and economic indicators frequently puzzle experts and make valid overall conclusions impossible, why should educational trends be easier to interpret?'. Goldstein (1983) makes a similar point, using the analogy of the retail price index. The weightings of the commodities which make up the index are determined by typical consumer spending patterns. But as spending patterns change, so does the index. How do we know whether the new index is measuring the same thing as the old one?

### **Methodologies for comparing public examination standards over time**

Perhaps because of the inherent difficulty of making sense of what is happening, there have not been many studies in this country which have attempted to make longitudinal comparisons of examination standards.

A major study of longitudinal examination standards was published by Christie and Forrest in 1980. They asked examiners to mark and grade scripts from 1963 and 1973 in Mathematics, English Literature and Chemistry, under different experimental conditions, and compared the outcomes. In Mathematics the syllabus content had changed to some extent but the types of questions had not, in English Literature both the content and the types of question had remained similar, though the mark scheme for 1973 reflected perceived changes in the teaching of the subject over that period, while in Chemistry in 1973 both the syllabus and the nature of the questions differed in significant respects from those of 1963.

The study illustrates some of the practical and philosophical difficulties of making comparisons of examination standards. One of the weak points of the study was the paucity of scripts from 1963. Even with the A level archive set up in 1989 by the Boards and the new national archive to be started this year, this is likely to be a problem, though the advent of electronic scanning may make it possible to store large numbers of examples of candidates' work and, in effect, replicate for previous years the material available for any given current year.

The main philosophical difficulty with the study, which is also in a sense a practical difficulty, is that the examiners who took part were inevitably from the later period. Even if examiners from 1963 had been available, they would not have been able fully to ignore whatever experiences they had undergone in the intervening ten year period. So the work from 1963 inevitably had to be looked at through 1973 eyes. I would take issue with Christie and Forrest's claim that those examiners who marked the 1963 scripts according to the 1963 mark scheme could, given enough time and enough of the right materials, have immersed themselves in the 1963 work to the point where they could fairly have devised a 1963 style mark scheme to apply to the 1973 work.

The Christie and Forrest study comes closest to replicating what examiners do in the live situation when they award grades. Two other studies carried out in 1994 (Fowles, 1995; Quinlan, 1995) on A level Physics and Mathematics also involved examiners making judgments about candidates' work. These studies used a cross-moderation ratification technique, where examiners from a number of Boards considered whether candidates' work from each other's Boards was or was not in a given grade boundary region. The studies included work from 1989 and 1994, in the case of Mathematics, and from 1990 and 1994 in the case of Physics, and all syllabuses were treated in the same way, whichever year they came from. Using syllabuses with relatively few years between them did not seem to cause any problems for the examiners. But cross-moderation methodology is loaded with assumptions and, as the authors concluded, cannot show conclusively whether there are statistically significant differences in grading standards. In any case, even with only a four or five year gap, the validity of making comparisons over time was rendered more doubtful by the introduction, during the intervening period, of the A level Code of Practice, which forced some Boards to alter their aggregation procedures in a way that had a direct effect on their results distributions. One of the main positive outcomes of the studies was the opportunity afforded to examiners to work on a range of scripts from a variety of syllabuses.

An earlier study by Willmott (1977) made comparisons of grading standards using a reference test. Test 100 was administered to CSE and GCE candidates in 1968 and 1973 and changes in examination results were compared with changes in Test 100 results. The main difficulty with this approach in the longitudinal context is that it assumes that nothing else has changed in the interim, that candidates' performances both on the examinations and on the reference test have been neither advantaged nor disadvantaged by anything that has happened in the intervening period.

Given the impossibility of being certain that all other things were equal, the variations in results which Willmott found from subject to subject and for male and female candidates, the differences in correlations between the reference test and the examination results between 1968 and 1973 and the difficulty of making sense of results which are highly significant statistically but which in terms of fractions of grades are very small in substantive terms, it is difficult to know what to make of the findings. It is true that most of the findings are in the same direction, to a greater extent than chance alone would decree, but, as Willmott himself points out, there may be explanations other than changes in grading standards. How can we know where the 'true' explanation lies? As Newbould and Massey (1979) suggest, 'Common tests would seem to look more attractive monitors of standards than they really are'.

One of the difficulties faced by those who work in this field is that their results are frequently expected to bear a weight of interpretation far beyond what they can reasonably support (Wood and Power, 1984). In a recent report on aspects of candidates' writing in English examinations between 1980 and 1994 (Massey and Elliott, 1996) two of my colleagues start off by identifying the weaknesses of both their data and their methodology and plead, '...we ask readers to appreciate at the outset that we are using the only data available, to try to make what comparisons we can'. Of course this did not stop their work being used as the basis for all sorts of unwarranted conclusions.

What they did was to take samples of candidates' writing in English examinations from 1980 (O level), 1993 and 1994 (GCSE) and compared the vocabulary, spelling, punctuation, sentence structure and use of non-standard English by grade and sex for each year. This admittedly limited kind of comparison provides interesting descriptive information on the nature of candidates' writing but, as Massey and Elliott themselves agree, does not allow us to draw any conclusions about what has happened to grading standards in English between 1980 and 1994.

The only other studies which I've been able to find which attempt to make longitudinal comparisons of examination standards as such are two exploratory studies carried out at the Test Development and Research Unit in Cambridge in the 1970s (Massey, 1978; Massey and Newbould, 1978). These used a form of the delta index and subject pairs analyses to attempt statistical comparisons over time, though some of the assumptions which had to be made were unlikely to hold, and the work was not taken further. Similar subject pairs work is reported in Christie and Forrest (1981).

I set out in this paper to look at the issue of comparing public examination standards over time. It is of interest in itself that the studies just described exhaust the publicly available work done on public examinations in England, Wales and Northern Ireland. For those of you who want to look beyond public examinations to other forms of assessment, I would recommend Nuttall's (1986) account of the problems inherent in the measurement of change, whatever methodology is used, and Brooks et al (1995).

### **Alternative approaches**

If it is really so difficult to make comparisons over time, is there any point in trying? The answer must be *Yes*, for at least four reasons.

One reason is that it is not unreasonable to attempt comparisons over relatively short periods of time. This is the approach, in the main, taken by the Boards themselves. The Physics and Mathematics comparability studies already cited, for example, made comparisons over four and five year periods (Fowles, 1995, Quinlan, 1995). Such studies cannot be free of the criticism that the context within which the examinations operate has changed, but the changes in such a time span are much less comprehensive than in the twenty year period essayed by SCAA and OfSTED.

This is presumably why the NCE Briefing Paper on standards in literacy and numeracy (Brooks et al, 1995) reports comparisons between any given test and its immediate predecessor, rather than attempting to make comparisons over long periods of time.

Similarly, Nuttall (1982) presents a quote which points out that one of the aims of the National Assessment of Educational Progress (NAEP) programme in the United States was to identify changes in student achievement levels over periods of time, 'characteristically four-year or five-year periods'.

The Boards' grading practices, as set down in the Codes of Practice (SCAA, 1995 and 1996), rely heavily on making comparisons with the previous year and/or with other recent years. These comparisons are both judgmental and statistical. A list is given in the Appendix to this paper. Of course these kinds of comparisons can be very difficult indeed when there is a substantial discontinuity in the system, as has been caused by the introduction of the two versions of GCSE in 1988 and 1994, and by the new A level syllabuses currently coming on stream. Once new systems have settled down, however, short term comparisons are not inappropriate.

The second reason for making comparisons over time is that, however imperfect our methodologies, they may alert us to potential problem areas which merit monitoring and further investigation. In Willmott's (1977) study, for example, he found that the Test 100 results suggested that the calibre of the Technical Drawing candidates had worsened between 1968 and 1973, to about the same extent as the calibre of candidates generally, but the Technical Drawing examination results were much worse than results in other subjects, suggesting that something was happening in Technical Drawing that was not happening in most other subjects. Willmott speculated that there might have been a change in the Boards' policy with regard to this particular subject, or some change in the preparation and/or motivation of the candidates. There would certainly seem to have been a case for further exploration of such a finding.

The third reason for making comparisons over time is that relative shifts can be detected. These might be differential trends in different subjects, as in the Technical Drawing case just cited. Another obvious example is the performance of male and female candidates. Although the Boards monitor sex differences, for operational purposes all candidates are treated as one group, so any divergences between the sexes, or any other sub-groups, are not caused by differential adjustments. As Elwood and Comber (1996) point out, the percentage of candidates achieving grades A to C at GCSE (since 1994 A\* to C) has risen, but it has risen considerably more for female candidates than for male and the gap between the sexes increases each year. The same kind of pattern is beginning to emerge at A level. Similarly, comparisons over time can tell us about the relative performance of schools and how their effectiveness changes (Gray et al, 1996). These kinds of comparisons tell us nothing about the absolute standard of what schools are achieving from year to year, but can identify schools which lie outside the general trend.

The fourth reason is that comparisons over time, although they may prove little or nothing about what is happening to standards, do provide insights into what is happening to the curriculum and to the means by which it is assessed. The question becomes not whether standards are rising or falling, but whether they are appropriate (Wood and Power, 1984). Longitudinal studies, of course, are only one way of tackling this question, but they can provide insights into what has and has not changed and hence provide a basis for considering what should and should not have changed.

Sutherland and Pozzi (1995), for example, in researching the changing mathematical background of undergraduate engineers, made an exploratory study of the content of A level Mathematics textbooks, syllabuses and examination papers in recent years and, although they concluded that a more detailed study was needed, they provided some interesting insights into developments in A level Mathematics.

As some of you will be aware, there has been a running controversy in this country for some time now, with one lobby claiming that A level Mathematics is too hard in comparison with other A level subjects, and another lobby claiming that A level Mathematics is not hard enough because holders of the qualification are ill-equipped for what is required of them in higher education. These kinds of claims and counter claims do not get us anywhere. Work like that of Sutherland and Pozzi, however, can provide the basis for a more sensible debate. What is A level Mathematics for? Who is it for? What would it be appropriate for candidates with A level Mathematics to be able to do in the 1990s? If we could tackle such fundamental questions, and reach agreement on the answers, we would be in a position to design syllabuses and examinations accordingly. I'm certainly not suggesting that this is easy but it seems a more positive way forward than the current sterile argument about standards in A level Mathematics.

### **Conclusion**

It is a very simple question - are public examination standards rising or falling? But it is also one of those tantalising questions which we've all encountered in educational research which are not just hard, but probably impossible to answer. My solution is to change the question.

### **REFERENCES**

BENJAMIN, H (1939) The saber-tooth curriculum, reprinted in Hooper, R (ed) (1971) The Curriculum: Context, Design and Development, Edinburgh, Oliver and Boyd

BROOKS, G, FOXMAN, D and GORMAN, T (1995) Standards in Literacy and Numeracy: 1948-1994, London, National Commission on Education, NCE Briefing, New Series 7

CHRISTIE, T and FORREST, GM (1980) Standards at GCE A-level: 1963 and 1973, London, Macmillan and Schools Council

CHRISTIE, T and FORREST, GM (1981) Defining Public Examination Standards, London, Macmillan and Schools Council

ELWOOD, J and COMBER, C (1996) Gender Differences in Examinations at 18+, London, Institute of Education

FOWLES, DE (1995) A Comparability Study in Advanced Level Physics, Manchester, NEAB on behalf of the GCE Boards

GOLDSTEIN, H (1983) Measuring changes in educational attainment over time: problems and possibilities, Journal of Educational Measurement, 20, 4, pp 369-377

GRAY, J, GOLDSTEIN, H and JESSON, D (1996) Changes and improvements in schools' effectiveness: trends over five years, Research Papers in Education, 11, 1, pp 35-51

MACKINNON, D and STATHAM, J, with HALES, M (1995) Education in the UK: Facts and Figures, London, Hodder and Stoughton, with the Open University

MASSEY, AJ (1978) A model for screening against standards drift between years using information concerning pass rates in different types of centre in the maintained school sector: O level Mathematics and English Language Examinations set by the Oxford Delegacy of Local Examinations and the University of Cambridge Local Examinations Syndicate between 1969 and 1975, Cambridge, TDRU

MASSEY, AJ and NEWBOULD, CA (1978) Standards Drift: A Screening Analysis, UCLES Advanced Level Subjects 1969-1976, Cambridge, TDRU

MASSEY, AJ and ELLIOTT, GL (1996) Changes in writing in 16+ English examinations between 1980 and 1994, Cambridge, UCLES Occasional Research Report

MORTIMORE, P (1996) Self-raising power, Times Educational Supplement, 14.6, p 19

MURPHY, R (1993) GCSE - a positive indication of success? British Journal of Curriculum and Assessment, 4, 1, pp 14-15

NEWBOULD, CA and MASSEY, AJ (1979) Comparability Using a Common Element, Cambridge, TDRU

NUTTALL, DL (1986) Problems in the measurement of change, in Nuttall, D (ed) Assessing Educational Achievement, Lewes, Falmer, pp 153-167

QUINLAN, M (1995) A Comparability Study in Advanced Level Mathematics, London, ULEAC on behalf of the GCE Boards

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY (1995) Mandatory Code of Practice for the GCSE, London, SCAA

SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY (1996) Code of Practice for GCE A and AS Examinations, London, SCAA

SUTHERLAND, R and POZZI, S (1995) The Changing Mathematical Background of Undergraduate Engineers: A review of the issues, London, The Engineering Council.

WILLMOTT, AS (1977) CSE and GCE Grading Standards: the 1973 Comparability Study, Basingstoke, Macmillan

WOOD, R and POWER, C (1984) Have national assessments made us any wiser about 'standards'?, Comparative Education, 20, 3, pp 307-321 (reprinted in Wood, R (ed) Measurement and Assessment in Education and Psychology, London, Falmer, 1987)

## **APPENDIX**

### **Boards' procedures for making comparisons from year to year**

These may not all be carried out by all Boards. The list may be incomplete for some Boards.

#### **Setting the Papers**

- 1 Setting comparable assessment tasks from year to year based on the same syllabus and specification grid and with equivalent standards applied in the mark schemes and in the assessment criteria. Where multiple choice items are pre-tested, it is possible to construct components consisting of items with similar facility values.

#### **Grading**

- 2 Comparing threshold marks at component and syllabus option level with those of the previous year.
- 3 Comparing mean marks and standard deviations for this year and last year, at component and syllabus option level.
- 4 Comparing grade distributions at component and syllabus option level with those from the previous year's award meeting.
- 5 Making comparisons of the kind described in 2, 3 and 4 above, but based on each of the previous 3 or 5 years, or on the averages of the previous 3 or 5 years.
- 6 Comparing scripts and course work on last year's thresholds with scripts and course work on this year's thresholds, at component level. Work from earlier years may be used, if appropriate.
- 7 Applying the same grade descriptions from year to year.
- 8 Comparing this year's candidature with last year's, on the basis of size, centre type and sex. A set of benchmark centres, usually defined by relatively large and stable entries, may be used as a proxy for checking the effect of awarding decisions, particularly if the candidature is not stable.
- 9 Comparing this year's forecast grade distributions with last year's, at syllabus and syllabus option level.
- 10 Comparing performance from year to year on multiple choice components, where item facility values are available.
- 11 Reports from Chief Examiners on the relative difficulty of the papers and performance of the candidates compared to the previous year.
- 12 Comparisons with last year's results from other Boards, making allowance for variations in entry by Board.
- 13 Some of the above comparisons might be made with the results from previous years for similar syllabuses. Syllabus pairs data might be used in this context.

#### **Grade Review**

- 14 Year on year centre comparisons, which identify centres whose results after grading are significantly different from those of the previous year.

Other factors which contribute to comparability include continuity of personnel i.e. Board staff and Chief Examiners, continuity of syllabuses and stability in candidature.

WEDNESDAY, 12 December 1906. 3½ to 5 or 4 to 5½.

## English Composition

[Candidates are recommended to give not more than three quarters of an hour to the essay. Regard will be paid to excellence of matter and style, rather than to length.]

1. Write an essay on **one** of the following subjects:
  - (a) Russia.
  - (b) "Back to the Land!"
  - (c) Electric Trams.
  - (d) William the Conqueror.
  - (e) Oliver Cromwell.
  - (f) Napoleon.
  - (g) The Olympic Games revived.
  - (h) A discussion of the value of the historical novel.

1  
ORD. LEVEL  
ENGLISH  
LANGUAGE I  
*Tuesday*  
1 JULY 1951  
1½ hours  
A

UNIVERSITY OF CAMBRIDGE  
LOCAL EXAMINATIONS SYNDICATE  
GENERAL CERTIFICATE OF EDUCATION

### ENGLISH LANGUAGE

ORDINARY LEVEL

PAPER I

(One hour and a half)

*Your composition should be of reasonable length; marks will be given particularly for style, subject-matter, and arrangement. Indicate by using the appropriate number at the top your paper which subject you have chosen.*

Write a composition on **one** of the following subjects:

1. The fascination of shop windows.
2. Monday.
3. You are left one day to look after a house where you are staying as a visitor. Recount your experiences.
4. Crowds.
5. A pageant **or** a procession through a town or village.
6. Castles in the air.
7. You have recently attended an auction sale as a spectator. Give an account of what you saw and heard while you were there.
8. During your holidays you consented to make a selection of books for the bookshelf of a younger brother or sister, or of a young visitor staying in the house. Describe how you set about your task, and how you succeeded in explaining your selection and in persuading the child to read the books.
9. In what ways is your locality contributing to the Festival of Britain? Do you approve of what is being done (or has been done) and, if so, why? What suggestions would you have made if you had been consulted?

WEDNESDAY, 12 December 1906. 3½ to 4¾ or 4 to 5¼.

## English Composition

Candidates are recommended to give about 40 minutes to the Essay. Regard will be paid to excellence of matter and style rather than to length.

1. Write an essay on **one** of the following subjects:
  - (a) Moths.
  - (b) A River.
  - (c) Elephants.
  - (d) Signs of Spring.
  - (e) One of the following characters: Sydney Carton, Quentin Durward, Amyas Leigh, Dr Primrose, John Ridd, John Silver, Diana Vernon.
  - (f) Travel in the Arctic or Antarctic Regions.
  - (g) Any of the English counties or of the British colonies.