

A BRIEF HISTORY OF POLICIES, PRACTICES AND ISSUES RELATING TO COMPARABILITY

Kathleen Tattersall

Abstract

This chapter which describes developments in education and assessment in England from the mid 19th century to the present day is in three parts:

1. the emergence of a 'national system' (the 1850s to the end of the Second World War)
2. the development and expansion of the system (1945 to the 1990s)
3. the emergence of a regulated system from the 1990s onwards.

The role of comparability at each stage of these developments is highlighted.

1 An emerging national system of education and examinations

1.1 Overview

Comparability is a fundamental requirement of England's assessment system: between providers of the same product, qualifications, subjects, the demands of tests and tasks, between assessor judgements of students' work and across time. Comparability is also an essential element of England's accountability methodology and is expected to guarantee to governments, the general public, students and selectors that assessment is fair, equitable and reliable. Without assurances of equivalence between different instruments and outcomes that are used to make selections for jobs or a university place or to aid future learning, the basis of decision-making is flawed. The more intense the competition for work, places on courses or financial reward, the more intense is the pressure on the assessment system to demonstrate the steps taken to assure comparability.

From the outset, comparability was a major driver in the English examination system, emerging as an issue in the mid 19th century, growing in importance as new examining boards came into being and the examination system emerged in the early 20th century; it continues to shape educational and assessment policies in the 21st century.

England's modern educational and examination systems are rooted in the 19th century. Competitive examinations for university entrance were a consequence of the establishment in 1836 of the University of London, which enabled two existing

colleges, University and King's, to award degrees, the first outside Oxford and Cambridge (linked in the public mind as 'Oxbridge') universities' awards. In 1838 the University of London set a matriculation examination to facilitate objective selection for entry to the two colleges in order to avoid the privilege inherent in the Oxbridge systems. An entrance examination for the civil service came into being in the 1860s. Other landmark developments, including the 1870 'Foster' Education Act, the 1902 'Balfour' Education Act and the foundation of new 'red-brick', civic universities in cities such as Manchester, Leeds and Liverpool, opened education to a wider group of learners. Improved administration by a Board of Education (1899) and Local Education Authorities (LEAs, a consequence of the 1902 Balfour Act) facilitated the growth of the education system. These developments marked the end of education geared almost exclusively to privilege, social class and males, and the beginning of today's national system of education and examinations.

As the 19th century progressed, grant-aided and public schools began to look to universities for guidance on standards in order that they could better prepare their students for university or a profession. London's 1838 matriculation examination and, later (1851), its intermediate examination provided such guidance on the range of subjects across the arts, classics and mathematics with which the university would expect well-educated prospective entrants to be familiar. Oxford (1857), Cambridge, London and Durham (1858) also responded to the demands of schools by providing syllabuses and examinations that candidates could take locally in their own schools. The requirements were specific to the needs of particular universities and were part of the hotchpotch of examinations which characterised the period; a far cry from today's national, closely regulated examination system, which measures individual and institutional achievement and is central to national accountability.

1.2 The emergence of comparability as an issue

The existence of several providers of examinations – London and Oxford and Cambridge – raised the issue of comparability of demand. As early as 1849 London granted admission to students who had satisfied the matriculation requirements of Oxford or Cambridge, thus implicitly accepting equivalence across the requirements. However, no reciprocal recognition seems to have been forthcoming from the two ancient universities. In later years, following the introduction of the School and Higher School Certificates in 1917, London would take a more active and questioning stance of the demands of other boards' examinations.

From the outset, it was important to Oxford and Cambridge that they should act together on important issues. Thus, in 1857, the Cambridge Syndics, debating what title to confer on those who succeeded in their new 'Local' examinations, considered:

...that it is of great importance to the success of the proposed system of Examinations that the Universities of Oxford and Cambridge should act together harmoniously on this matter.¹

However, disagreements emerged in 1859 when the two universities could not agree a common timetable as requested by 52 prominent headmasters. This opened the

way in due course to questions regarding the respective standards of their examinations: in May 1872, a headmaster wrote in the following terms to *The Times*:

Oxford pitches her standard, if not too high, higher than her Sister University. No junior can pass at Oxford without satisfying the Examiners, in addition to the preliminary subjects, as to his knowledge in a foreign language, mathematics or chemistry; while at Cambridge the subjects of an English education only may suffice. And again, in the case of seniors, Oxford sets no special books to be read in Latin, Greek, French, or German; whereas Cambridge does. As the public value of the Oxford and Cambridge certificates is much the same, masters, parents, and pupils naturally prefer entering for Cambridge.²

The Times correspondence about standards took on a personal flavour when an individual who had been failed by Oxford but passed by Cambridge claimed that his experience was evidence of the universities' disparate standards. The Secretary to the Syndicate, George Forest Browne, entered into protracted correspondence in *The Times*. Mr Browne's final letter, citing maturation as a possible reason for the different outcomes of the student's experience, was not sent as by then the correspondence had grown stale:

RB's experience as a Candidate, when fully stated, does not... prove his point. He failed in the summer at Oxford, and after that salutary lesson and the experience which one examination gives he passed with considerable credit with Cambridge six months later, a most important six months at his time of life. That the latter was a correct estimate of his ability is shown by the fact which slips out inadvertently – that Oxford afterwards gave him the very high honour of a First Class in the Senior Examination.³

Public scrutiny of question papers and observations on their standards put the spotlight on comparability within a university's own school examinations, across time and across providers. In 1893, the standard of arithmetic in the Junior paper of Cambridge was considered by one critic of the Syndicate to be as difficult as that of its Senior paper of 1883.

'Is it supposed', the critic asked, 'that boys' brains are better than they were ten years ago, or are we to work longer hours?'⁴

The examinations of the University of London were also drawn into the debate in a letter to the *Journal of Education* dated 1 September 1893:

Sir, You would much oblige by telling me what is the comparative value of the London Matriculation certificate with that of the Cambridge Higher Local. I was astonished to hear a lady the other day state that the two examinations are of similar value. So far as my experience goes, the Matriculation certificate is not equal to that of the Cambridge Senior Local, and to compare it with the Higher Local approaches the absurd... The late Secretary to the University of Cambridge... considered a pass in the Cambridge Higher Local equal to a B.A. pass of Cambridge University, and a Cambridge Higher Local Honours certificate equal to a B.A. in Honours.⁵

The letter provoked a response from a headmistress who appreciated the complexities and different demands of syllabuses and examinations and realised how

misleading superficial comparisons could be, a view echoed a century later by those charged with considering the vexed question of standards over time:

It is difficult to compare the two examinations, as they are so different in their natures. To matriculate, a candidate must pass in all seven subjects at the same time, and has little choice of subjects; while to obtain a Higher Local certificate, except that she must take mathematics or languages, she can choose what subjects she likes, and take them when she likes. As a pupil of mine, who knew no mathematics, learnt enough in a year to pass the Higher Local, besides working at other subjects, I think the difference in the standard of the two examinations cannot be so very great.⁶

Concern was also voiced that having several examining boards would result in unhealthy competition, which would lower standards. A letter to the *Pall Mall Gazette* in December 1894, under the sub-heading 'Downward Competition of Examining Boards', did not mince its words:

It must not be imagined for a moment that its (College of Preceptors) examinations – or the Locals – are really independent exterior tests of a school's efficiency. These boards are competing boards; they exist upon the recognition afforded them by schoolmasters and schoolmistresses... Examiners are 'satisfied' by a rudimentary knowledge of arithmetic, reading aloud, and the answering of papers in religious knowledge and English. Such complacency would find in a glass of water and a crust, nectar and ambrosia. Even at that the 'Local' examiners reject a quarter of the candidates, and their customers the teachers grumble. As a result a process of downward competition seems to be setting in.⁷

A century later similar allegations would be made by a television programme, *Dispatches*, which, unlike the *Pall Mall Gazette*, reached millions of homes (this is discussed later).

The examination scene was complicated further by the arrival of the new civic universities, most notably those in the north. The Northern Universities' Joint Matriculation Board (NUJMB) was established in 1903 by the Universities of Manchester, Leeds and Liverpool (formerly the single federated Victoria University) on receiving their separate charters. In its later years, reflecting its 1978 decision to operate nationally rather than in the self-determined region of the north and the midlands, the board described itself simply as the Joint Matriculation Board (JMB), dropping the northern reference.

The board's prime purpose was to ensure comparable entry requirements across the three founding universities, which were joined by Sheffield in 1905, with Birmingham completing the board's membership in 1916. Schools in the north and the midlands increasingly looked to the JMB for guidance on subject content and standards as they prepared students for university entrance.

1.3 The codification of school examination provision

By the early years of the 20th century, about one hundred separate examinations existed, with each major profession, as well as the universities, setting its own specific conditions of entry, 'without,' according to the President of the Board of

Education in 1917, 'there being much regard to the general educational convenience of the country'. Many considered this unregulated plethora of examinations to be detrimental to the education of able students who wanted to keep open their options of a professional career or university entrance. The Board of Education's Consultative Committee observed in 1911:

More than four-fifths of those who left grant-aided secondary schools after reaching the age of 14 did so without sitting for a public examination; on the other hand, of the twenty-nine candidates who entered for Bedford College scholarships in 1909, eight had already taken three public examinations, eight had taken four, three had taken five, two had taken six and one had taken seven examinations.

Petch (1953, p. 65)

Worries about the effect of multiple examinations on their students, scepticism about their comparability and the limited currency of certificates, led headteachers to call for a codification of examinations. The JMB received a resolution in February 1904 to this effect from the Incorporated Association of Headmasters (Petch, 1953, p. 53). They argued for an examination *system* in which the certificates of any approved examining board would have widespread acceptance. For such a system to operate, authoritative assurances would be needed that the certificates were broadly equivalent in respect of the range of subjects tested, their standard and format. In 1903, the Board of Education's Consultative Committee had circulated plans for a scheme of school-leaving certificates but the proposals had been shelved. In 1909, the Committee returned to the issue, but it was not until 1911 that a system of school examinations at two levels was formally proposed. The Committee's recommendations for a School (leaving) Certificate (SC) for 16-year-olds and a Higher School Certificate (HSC) designed to allow entry into any university finally came into being in 1917.

The Consultative Committee proposed that certificates would record attainment across an agreed range of subjects, some 'main' others 'subsidiary', with Pass and Credit standards, providing some consistency in the subjects studied and the standard reached. The subjects were arranged in Groups and, in order to achieve a certificate, students had to demonstrate their attainment across subjects in different Groups. The eight university-based boards (see Table 1) became 'approved' boards and were invited to set, mark and certificate the examinations under the jurisdiction and authority of a central coordinating authority, the Board of Education. As the certificates would follow the same format and encompass the same Groups of subjects, it was intended that attainment at agreed standards would obviate the need for specific university matriculation examinations; in the event the HSC was accepted as the basic standard for matriculation but universities continued to specify their own additional entry requirements, a practice that continues today. Most professions too were willing to accept that the certificates were equivalent and sufficient for their entry purposes provided that specific subject requirements were met. The NUJMB's 1926 Syllabuses and Regulations typically contained several pages listing the different entry and matriculation conditions of universities and leading professional organisations.

The Secondary Schools Examinations Council (SSEC) minutes of 14 February 1931 record approval of the JMB's proposal to offer, as optional subjects for study beyond the basic requirements for a certificate, vocational subjects such as navigation, spinning and weaving, and mechanical engineering. This meant that these subjects came to be examined by the university examining boards. There was, however, no attempt made to bring within the system any of the vocational examinations (e.g. those of City and Guilds) that had proliferated in the 19th century especially in the years after the Great Exhibition of 1851. This exclusion was entirely in keeping with the prevailing views that 'trade' was not relevant to education. However, the exclusion of vocational examinations from the national system at this early stage would have a detrimental effect on generations of learners and result in vocational examinations being accorded a lower status than their academic counterparts. Attempts to address this issue and demonstrate comparability would drive national assessment policy and shape examining board structures in the last decades of the 20th century.

The Board of Education exercised its coordinating responsibilities for the School Certificate system through an advisory committee, the SSEC, on which all eight examining boards were represented until the Committee's reconstitution in 1946 when, as a result of their opposition to the teacher assessment recommendations of the 1943 Norwood Report, they were dropped from the Committee. The SSEC remained in existence until 1964 when it was replaced by the Schools Council.

1.4 The School and Higher School Certificates

The SSEC's remit in relation to comparability was to secure reasonable equivalence in the demands of the Group requirements and the examination standards of the examining boards. 'Reasonable' was not defined, nor did the SSEC have powers to enforce an interpretation or regulate the activities of the examining boards, which remained under the control of the autonomous universities. The Committee suggested improvements to the examining boards and made recommendations to the Board of Education. It had no powers to approve either syllabus content or the range of subjects that a university could demand in the Higher Certificate for its specific matriculation requirements, which undermined the comparability of the system. The modest expectations of the SSEC of the examining boards were expressed in a later report:

The independence of the Examining Boards (is) subject only to the condition that certificates should be reasonably equivalent and should be awarded in accordance with an agreed scheme of requirements... It is all to the good that they should have their own characteristics and their own ways of doing things; yet it remains true that they are engaged on common work. They possess indeed unrivalled opportunities for studying the techniques of examining and for trying out experiments in regard to the type of examination papers which are most suitable in School and Higher Certificate Examinations. It is the more necessary that knowledge gained should be brought into the common stock, and that all possible steps should be taken to promote a frank interchange of ideas and experiences.

SSEC (1932, paragraph 20, p. 20)

If not satisfied with the standards of an examining board, the Board of Education could withdraw its approval but this sanction was never applied; had it been, no state-funded secondary school would have been permitted to enter the examinations of the board in question – not unlike the situation after the 1988 Education Act when state-funded schools were not allowed to offer courses that did not appear on the statutory list of approved qualifications. Given the ramifications, particularly where an offending examining board served a large number of schools – as did the JMB, London and the Oxford Delegacy – it was in the interests of all parties to agree a way forward.

In the absence of formal structures, it was not easy for the examining boards to work together on assessment issues or speak with a single voice on policy matters. The minutes of the Cambridge Syndicate are revealing. The minutes of its Joint Committee for Examinations (which included Syndics and representatives of the leading teacher associations) refer to a long-running difference of opinion with London, which, perhaps understandably, was unwilling to passively accept the standards of other providers, as it had been in the mid 19th century, without clear evidence of equivalence. London demanded ‘a higher standard’ for matriculation but did not make clear to either Cambridge or the SSEC exactly what its requirements were. The minutes of the Syndicate’s General Purposes Committee meeting of 15 March 1918 record:

It appears that pressure will be brought to bear on London by the Board of Education, HMIs etc. Meanwhile, what ought we to do? Not to raise our own standard to try to fit a standard which we cannot gauge. We have not the necessary information before us.⁸

The lack of agreement on what constituted the required standard of attainment was a worry to teachers. In October the Joint Committee for Examinations urged the Syndicate to:

... put before the London Matriculation Board the difficulty experienced by candidates in attaining ‘good’ in each of the three English papers and in each of the Mathematics papers and to urge the acceptance of one aggregate ‘good’ standard in each of these two subjects.⁹

The row rumbled on well into the 1920s and resurfaced in a more serious form in 1931 when the SSEC conducted one of its intermittent investigations into the School Certificate, which was also investigated in 1918 and 1924. Similar investigations were made of the Higher School Certificate in 1920, 1936 and 1937.

Unusually, the 1931 findings were published, ‘in view of the growing importance of the examination... and of the criticisms of the examination, some of them useful but not all of them well informed, which have appeared in the press and elsewhere’ (SSEC, 1932, p. 5). The investigators’ report referred to London’s refusal to accept the central tenet of the School Certificate: that the syllabuses, examinations and marking of each examining board were of the same standard and made equal demands on candidates. This dissent was a clear challenge to the authority of the Board of

Education over the universities and, therefore, over the examinations they ran. London demanded the right to satisfy itself first hand of the quality of the students' work in the examinations of the smaller examining boards, Durham and Bristol, by checking the standard of their syllabuses, questions and marking of scripts. This was unacceptable to Durham, Bristol and the investigators who commented in the 1931 report:

This [subjects] the Examining Boards to a kind of dual control, the control of the SSEC... and the control of a powerful University... It is of vital importance to [the Examining Boards] that their certificates should be accepted for matriculation purposes by the University of London... When the School Examinations were reorganised ... it was assumed that the Universities would one and all be willing to leave this question of standards to the SSEC and that no University would seek to interfere with the immediate responsibility of an Examining Body for the conduct of its own School Certificate Examination.

SSEC (1932, p. 51)

Differences of view between examining boards meant that the SSEC was not in a position to give a ringing endorsement to the comparability of the standards and procedures of the examining boards. Its limited remit – not extending to syllabus approval – also meant that it could not express confidence that there was comparability between different syllabuses in the same subject and across subjects, even though equality of standards was an implicit requirement of a Group award. In reality, there was considerable variation between the syllabuses of different examining boards and between alternative syllabuses of individual examining boards – it would be another half century before syllabuses would have to conform to Subject Cores or National Criteria. Frequent alterations to Group requirements, changes of subject content across years and differences between the demands of subjects within Groups further obfuscated comparability between subjects and from year to year.

1.5 Reliability of markers

As more students entered for examinations, the unreliability of markers' judgements emerged as a further concern. This problem had been exposed many years before the introduction of the School Certificate system by Professor F.Y. Edgeworth of Oxford University in his statistical analysis of university and civil service examinations (Edgeworth, 1888). In the case of the civil service examinations, Edgeworth had observed that one-third of scripts marked by different examiners received a different mark and, further, that in a re-examination of scripts by the same examiner one-seventh received a different mark (Roach, 1971, p. 284). Similar findings would be observed in cross-moderation studies conducted by the GCE examining boards in the 1950s and 1960s (Bardell *et al.*, 1978).

Placing candidates accurately in grades (several, or pass/fail) was, therefore, fraught with difficulties. This unreliability would be offset, Edgeworth suggested, by multiple marking and designing examinations with a number of components: the more individuals marked a piece of work, the more likely it was that a 'true value'

would emerge; the more components were aggregated, the more likely that individual marker errors would cancel out. Edgeworth's conclusions and his suggested remedies continue to influence examining practices in the 21st century.

How to improve markers' reliability was a hot topic at an international conference on examinations in Eastbourne in May 1931, leading to considerable research in Europe and the United States (Montgomery, 1965, p. 260). In 1936, Hartog, Rhodes and Burt summarised the context of the research:

Professor F.Y. Edgeworth, many years ago, found that the marks allotted independently by twenty-eight different examiners to a single piece of Latin prose varied from 45 to 100 per cent, and made a number of other investigations on variability in marking. In the United States, Messrs. Starch and Elliot, and in France, M. Laugier and Mlle. Weinberg, have found similar results, but no systematic comparison has hitherto been published of the marks allotted by a number of different examiners and by different boards of examiners, all experienced and qualified for their task, to sets of scripts actually written at public examinations. Both the English and French Committees have attacked this subject... The results are similar in the two countries and equally disquieting.

Hartog *et al.* (1936, Preface, paragraph vi)

The research confirmed Edgeworth's findings of half a century earlier: in England the marks given by 14 examiners to 15 School Certificate history scripts varied considerably. The same individuals marking the same scripts some 12–19 months later produced an even greater level of variation with 92 cases out of 210 receiving different pass, fail or distinction classifications. How to address the problem? A sole marker was thought to be the ideal solution (though this assumed intra-marker consistency) – but proved unrealistic as entries for the examinations increased. Double marking was a possibility – provided that there was an adequate supply of reliable examiners; this practice was adopted for English examinations by some boards until burgeoning entries for the subject in the early 1980s made it impossible to recruit sufficient examiners.

Another possible solution was more detailed marking schemes to provide tighter parameters for examiners' judgements. However, the SSEC's investigators were not enthusiastic: their comments on the JMB's Higher School Certificate history examination in the 1937 investigation pre-date a later debate about the relative merits of holistic or atomised marking:

The application of a detailed marking scheme to each answer might possibly be defended on certain grounds... [But] it is apt in time to become merely mechanical... The Investigators have seen nothing in the system as applied here to suggest that this examination has received any benefit from it, or that it is in any way superior to the practice of assessing the value of each individual answer as a whole. At times the Investigators were at a loss to understand on what principle the marks were being awarded... and the indication of errors in the scripts was definitely inconsistent.

SSEC (1939, p. 16)

Marker reliability was not the only problem; other factors were known to affect the reliability of examinations:

There is also the element of chance due to the variability of conditions for individual candidates, arising from illness or accident, which it is difficult to estimate statistically. It may be reduced, in a rough and ready way, when examining boards take into account school-records in border-line cases. Then there is the element of chance due to variability in the difficulty of the papers set. Some examining boards dealing with large numbers of candidates attempt to reduce marker variability by correcting the marks assigned by the examiners to candidates, so as to make them conform [in accordance with a suggestion of Edgeworth] to a curve regarded from experience as being suitable for the particular examination. This last expedient helps to avoid violent fluctuations in the proportion of those who pass or fail, or are awarded marks of credit and distinction...

Hartog *et al.* (1936, Preface, paragraph ix)

School records, attempts to ensure consistency in the demands of papers within and across years, adjustments to distributions and individual examiners' marks were all part of the armoury that the School Certificate examining boards used to underpin comparability. Little has changed, although researchers now have the benefit of sophisticated computing and technological tools to analyse the variables that have a bearing on assessment outcomes.

1.6 Standards

It was one thing for all markers to be reliable, another for questions and the expected responses to be set at the appropriate standard. Chief Examiners were key to the process – they set the papers and their marking, assumed to be 'correct', was the absolute standard against which individual markers were measured. These assumptions remain at the heart of the examining process. As an early publication of the JMB put it:

We have, in fact, imposed the standard of the chief examiner on the whole panel, but what guarantee have we that this standard is the correct one? What units have the chief examiners to measure by? Actually none whatever. They are very experienced examiners, but we have only their opinion to go upon... What can the chief examiner do to maintain the same standard as at an examination held twelve months before? It has been claimed that it cannot be done... Experienced and conscientious examiners vary one from another, they cannot all have got the correct and absolute standard – one or more may [by accident] have found it, but which? There is no means of telling.

Crofts & Caradog Jones (1928, p. 44)

The problem, therefore, was not the marking as such but the determination of the credit standard of a certificate made up of attainment in several subjects:

Taking the subjects all round, one may say that good examiners can and do place the candidates in their correct order of merit. The difficulty is the fixing of the absolute standard; having obtained the order of merit of the candidates, should they all pass, or should they all fail, or where ought the line to be drawn?

Crofts & Caradog Jones (1928, p. 47)

It was particularly important to the JMB and other examining boards with large entries for School and Higher School Certificate to maintain consistency from year to year: the award of scholarships and other financial rewards as well as entry to university courses were determined by success in the examinations. There were, however, no joint arrangements in the first instance for the examining boards to agree a common interpretation of the Pass standard; the JMB's solution to the problem was a statistical one:

Where large numbers of candidates are being dealt with, the variation of standard among them in the mass from year to year is small, or, at any rate, small compared with the variations we know take place in the standard of examiners. The candidates are not like a fruit crop, which may suffer blight and produce poor results in any one year; in normal times variations in standard are small, and we should err very little if we kept the percentage of passes in the important subjects fairly constant from year to year.

The Chief Examiners should therefore be asked... whether their percentage of passes differs much from previous years, and if so to state what in their opinion is the cause of the difference, and whether there is any reason why the figures should not be altered so that the final figure might be brought into line with the previous one.

Crofts & Caradog Jones (1928, pp. 45–46)

The use of statistics to moderate outcomes and provide year-on-year comparability, which continues to operate in the English examining system, had been adopted by the JMB as early as 1918 when its governing body resolved:

That the attention of the examiners for the Higher School Certificate should be drawn to the disparity between the different subjects in the results of the examination in Higher Alternative papers, and to the irregularities from year to year in the same subjects in that examination, and that the Board should suggest that a percentage of from 10 to 15 candidates obtaining 260 marks out of 400 should be roughly the norm to follow.

Quoted in Petch (1953, p. 137)

It is not clear whether this imperfect solution was adopted universally but by 1931 the SSEC investigators could refer with some confidence to the 'two recognised standards' in each subject, the 'norm-referenced' – and, therefore, 'clearly understood' – Credit standard and the 'less clearly defined Pass standard':

The Examining Boards have been reasonably successful in maintaining a steady credit standard in each main subject. In any given examination the credit mark may vary from subject to subject and it is not necessarily the same for a given subject from year to year. The credit standard in a subject in an examination is... best described not in terms of the credit mark but broadly speaking as the standard which a given percentage of the candidates offering the subject in question in that examination will reach... Generally speaking credit is obtained by about half the candidates, but it would be quite untrue, even as a general statement, to say that the credit standard is mechanically fixed to give this result. There is a fairly definite relationship between the credit and the pass marks in a subject; thus in one large examination the pass mark is 4/5ths and in another 7/9ths of the credit mark: and of late the Examining Boards have been endeavouring to keep the pass standard reasonably uniform from year to year.

SSEC (1932, paragraph 30, p. 31)

Some 75 years after the publication of the 1931 report, the investigators' pithy acknowledgement of the difficulty of setting standards in any given year and maintaining comparable standards across years resonates powerfully:

The practice of keeping the percentage of credits and passes nearly constant from year to year is reasonable, if the general performance of candidates is from year to year practically the same; but if there is a real and continuous improvement in the general quality of work in a subject and the percentages of passes shows no corresponding rise, it follows that the standard of the examination is unintentionally being raised. Alternatively, an influx of pupils from weak forms might result in candidates being passed who in former years would have failed. All that can be said on this matter is that these possibilities should be kept in mind, the situation watched, and alterations made when these seem justified. If for example it was found that in a given subject and on comparable papers the performance of candidates on the papers set in 1930 and 1931 was distinctly better than that of their predecessors, say five years earlier, there would be good grounds for increasing the percentage of passes and credits.

SSEC (1932, paragraph 35, p. 41)

The reference to comparable standards across years made an impression on at least one examining board Secretary, Mr Crofts of the JMB; his personal copy of the report, to which the author had access, underlines the words 'five years earlier' and a handwritten note reads, 'Would there? Why not 50 or 100 years earlier?'

Norm-referencing may have helped examining boards with large entries to maintain some consistency across their subjects and years, but was less helpful to those whose entries were very low. Entries were distributed unevenly across the eight providers and little reliance could be placed on statistical comparisons of the awards of different boards. In 1930, London (20.4%) and the NUJMB (28.0%) accounted for almost half of all the 63,117 School Certificate entries while at the other end of the spectrum Bristol (0.8%) and Durham (1.9%) accommodated less than 3%. Table 1 shows the changed pattern of entries (expressed as a percentage) between 1918 (the first year of the examination) and 1930.

Table 1 Percentage distribution of candidates among the several examinations

	1918	1930
Oxford and CJB	8.8	11.9
Oxford Local	33.4	16.9
Cambridge Local	17.9	11.0
NUJMB	14.8	28.0
London	12.4	20.4
Durham	1.9	1.9
Bristol	0.3	0.8
Wales	10.5	9.1
Total no. of candidates	22,873	63,117

The full table can be found in SSEC, 1932, paragraph 19, p. 19.

All of the issues of comparability, which emerged during the period of the School Certificate, remain relevant in the 21st century. The first examining boards were conscious of their responsibility to demonstrate comparability within the marking of particular subjects, to set consistent standards for their subjects and certificates, and to ensure that there was widespread public acceptance of, and confidence in, their work. They developed new techniques of assessment, sought more reliable ways to ask questions and improve marking, credit answers and standardise the increasing number of examiners they employed. Individually they inched their way to solutions, which today are commonplace, but in the 1920s and 1930s were quite revolutionary. The JMB, for example, introduced new-style papers in geography – ‘answer books with the questions displayed in such a way, and with such detail, as to indicate not only precisely what the examiners were asking about but also approximately how much length they were asking for’ (Petch, 1953, p. 94) – in an attempt to achieve better reliability in the marking. The inevitable charges of lowering standards accompanied such changes.

2 Development and expansion of a national system of education and assessment

By the end of the 1930s a national examination system, loosely coordinated by the Board of Education, was well established. The School and Higher School Certificates provided schools with a reasonably stable benchmark and students with worthwhile targets. Along with matriculation requirements, the examinations largely determined the curricula of public and state-funded secondary schools. Judged by 21st century levels, examination entries remained low. In 1947, 107,356 candidates took the School Certificate examination and 26,322 were entered for the Higher School Certificate (Gosden, 1983, p. 75). The first GCE O level examinations in 1951 were taken by around 134,000 candidates, while A level attracted about 37,000 candidates (Bardell *et al.*, 1978, p. 13).

2.1 GCE O and A level examinations

Wartime provided the impetus to re-examine social policies which the 1945–1951 government enacted. Education underwent major changes. The 1944 Butler Act opened opportunities to a wider social group by raising the school-leaving age from 14 to 15 and establishing the principle of free secondary education within a tripartite system of secondary schools. A year earlier, the Norwood enquiry into the curriculum and examinations had proposed that future examinations should be single-subject based in order to unshackle schools from the diktat of the examining boards and encourage more students to take examinations (Board of Education, 1943). By being free to drop their weak subjects, students would be better able to demonstrate their achievements. This approach was not welcomed universally. Some, such as the headmaster of Liverpool Collegiate School, Mr A.L. Kneed, argued that there would continue to be a demand for attainment across a range of subjects. Writing in the *Liverpool Echo* in January 1952, he claimed that a certificate would only have currency if it contained five or six subjects, including English, mathematics and a foreign language (see endnote 10) – not so different in range from the former School Certificate and similar to later proposals for a baccalaureate-type Diploma. A system

of teacher assessment at 16 with external examinations at 18, primarily for university selection, was also proposed. However, both this recommendation and rearguard attempts to retain broader requirements for certification were unsuccessful.

The new General Certificate of Education examination at Ordinary, Advanced and Scholarship levels came into being in 1951, but selection at 11 for grammar, technical and modern secondary schools and a school-leaving age of 15 (16 from 1974) continued to limit the number of students taking GCE O level. Candidates for O level had to be at least 16 years of age on the 1 December preceding the examination, a requirement that provoked bitter opposition from the examining boards and schools. This minimum age requirement was relaxed for the 1953 series of examinations when headteachers had to certify that it was educationally desirable for younger students to take the examination and that they were likely to pass in the subjects they had studied.

The main concern about the new GCE O level was the linking of its pass standard from 1952 onwards (1951 was a transitional year) to the credit standard of the School Certificate which, it was considered by many headteachers, would deprive students, who would have met the former pass standard, of a certificate. The headteachers were also unhappy that the new examination would be pass/fail with no distinguishing credit or distinction 'grades'. Schools were encouraged to report to students their marks but it was not until 1975 that formal grades were recorded on O level certificates (although some boards provided unofficial grades).

In his same 1952 *Liverpool Echo* article, the outspoken Mr Kneed of Liverpool Collegiate School was blunt about what was perceived to be an unhelpful raising of the bar:

This raising of the standard can mean only one thing – not, I fear, that by a stroke of the pen the children in our schools get cleverer the more work is done, but that a much larger number of children will go out into life without any certificate at all, though they have completed five years at grammar school and not done too badly...

The idea of parity has gone so far as to remove incentives from the good worker and the able child because now no difference is made between the bare pass and the really high mark. The child who gets 51% and the child who gets 90% are both awarded just a pass – no credit, no distinction for the really good performer, no evidence as to what is a child's strongest suit, or promising line.¹⁰

A new comparability concern materialised with the first results: the different levels of attainment of boys and girls. Mr E.R. Wood of the High School for Boys, Hereford, wrote to the *Times Educational Supplement (TES)* on 30 November 1951:

The Cambridge Local Examinations Syndicate has recently issued a table of percentage results for the General Certificate in the summer of 1951. The results at ordinary level show the performances of boys and girls separately, and a comparison of them is startling. In English literature, for instance, 75.6% of the girls passed, but only 54.7% of the boys. The superiority of girls' results is marked in nearly every subject, including such science

papers as physics and general science, and if we divide the total number of passes by the total number of scripts in each sex, we find 62.6% of girls successful, as contrasted with only 54.4% of boys. The Syndicate's figures for passes at advanced level are not given according to sex, so that we have no evidence as to whether girls are also superior to boys in advanced work.

It would be interesting to know of teachers in mixed schools whether the apparent superiority of girls as examination candidates accords with their experience. If it is generally true, what is the explanation? Superior intelligence? Greater conscientiousness? Better teaching? And does it last?¹¹

In the years that followed, Mr Wood's questions would continue to provoke debate and research into gender (Elwood & Comber, 1996) and other equity issues that could have a bearing on attainment (Gipps & Murphy, 1994).

The new O level was intended to be used by grammar schools and not by technical and modern schools, which were discouraged by government from entering whole classes for the examination. Many schools ignored government advice:

One of the problems which is arising in connection with the new examination concerns its use by increasing numbers of pupils presented by technical colleges and modern secondary schools, whose needs are not identical with those of the grammar schools.¹²

Attempts to restrict the GCE examination to grammar school students resulted in a proliferation of local leaving certificates for technical and modern school students provided by LEAs and others. Such certificates had no national currency.

In 1953 a new board, the Associated Examining Board (AEB), was approved as the ninth (excluding Northern Ireland, but including Wales) GCE examining board, sponsored by City and Guilds, mainly to provide examinations for the growing numbers of students interested in technical and vocationally related subjects. (The number was reduced to eight in 1964 with the demise of Durham and then to seven in 1990 when the SUJB merged with Cambridge.) Appendix 1 represents the boards' histories. However, it was not until the 1960s and the advent of the Certificate of Secondary Education (CSE) which brought into public examinations a large new population and new comparability challenges, that the need for local certificates disappeared. Until then, the majority of students left school with no formal qualifications. In 1958 a mere 13% of school leavers left education with five or more O levels. In 1960, 18.8% of leavers attained one or more A levels and fewer than 5% went on to university. As late as 1972, 43% of students left school without even attempting a qualification.

2.2 Collaboration to achieve comparability across different providers' examinations

The increased demand for university places, which gradually followed the introduction of the GCE, made it important to demonstrate comparability across examination providers so that results could be used with confidence. The same was true in respect of employers, many of whom set basic entry requirements of a pass

standard in English and mathematics. As entries for O level and A level grew, more assistant examiners were required, increasing the pressures on comparability. More detailed mark schemes were required to ensure a consistent understanding by the examiners of what could be credited – the report of the chemistry cross-moderation study undertaken by the GCE boards in 1953 refers to London's:

...very detailed marking scheme, which was necessitated by the large number of examiners (13 on each paper).¹³

More systematic standardising arrangements were put in place by individual examining boards to ensure reliable marking.

The Secretaries to the GCE examining boards were conscious that they needed a formal structure to demonstrate that their examinations led to reliable and comparable outcomes for students. From the outset, therefore, they met regularly,

with a view to checking the co-ordination of pass standards at ordinary and advanced levels.¹⁴

These meetings became the backbone of inter-board co-operation and collaboration on standard setting, certification and administration. In March 1952 the Secretaries agreed to the JMB's suggestion to:

Arrange experimental enquiries regarding standards, beginning with Chemistry A level, a subject which might be less likely than others to raise controversial problems.¹⁵

The first enquiry (in March 1953 using the 1952 papers) established a format for cross-moderation studies, which the examining boards conducted each year in an agreed range of subjects at both O and A level: a conference of senior examiners who studied syllabuses, papers, mark schemes and scripts to determine whether the standard of the board under scrutiny was the same as that which they applied in their own board. The report of the 1953 cross-moderation GCE chemistry study refers to different interpretations of subjects and different lengths and design of papers but dismisses these as having 'no bearing upon the aim of the enquiry... differences could be ignored'; the examiners 'could detect no difference of standard as between the... examining boards'. The report went on:

It appeared therefore that the differing statistics of entries and percentage passes are not primarily due to differences in the general standards of the examinations, and that in view of the perennial difficulty of establishing a hard and fast line as between pass and fail there is general agreement... as to what represents the bare Advanced pass level in their subject. The differences in the statistics may be due to differences in the fields from which the boards draw their entries.¹⁶

The issues identified in this and subsequent cross-moderation studies led to research becoming part and parcel of the GCE examining boards' work. Research departments with high-profile directors were established: its former Secretary, Dr Petch, headed that of the JMB, created in 1964. A further indication of the value the GCE examining

boards attached to research came in 1970 when a Standing Committee of GCE Research Officers was founded to:

... deal with matters of common interest, including the comparability of standards.¹⁷

Prior to the establishment of a common timetable, it was possible for candidates to be entered for the same subject with different examining boards. These dual entries were a fruitful source of the boards' own research into comparability. Two studies in 1958 (when 1,311 O level and 41 A level candidates had taken, in the same series, the examinations of two different boards) and in 1966 (11,674 O level and 1,524 A level dual entries) were undertaken. The report of the 1966 study (UK GCE Examination boards, 1971) concluded that, as 70% of candidates at both levels were awarded the same grade (replicating the findings of the 1958 study and Edgeworth's 1888 work), there was no major difference in standard between the examining boards. Variations in awards to the remaining 30% were attributed to factors such as the preparation of candidates, their familiarity with one syllabus and not another, their choice of questions, 'which on one day can be disastrously ill-judged (but) can on another day be sensible and related to (the candidate's) capacity', motivation and chance. In later years a national system of assessment and accountability would be less tolerant of such variations.

The examining boards reported their individual and collaborative research in their annual reports and occasional publications, such as those of the JMB. The latter addressed specific issues, including comparative performance across subjects (subject pairs analyses), the topic of a 1982 publication (Forrest & Vickerman, 1982). The JMB's motive for publicising the work was simple:

For the public in general and for teachers, pupils, parents, employers and most users of examination results the question of comparability of standards in public examinations is probably seen as a simple matter: they expect a given grade to indicate the same standard of performance irrespective of the examining board which awarded it, the subject in which it was achieved and the year in which it was gained. It can certainly be argued very strongly that all users of public examination results are entitled to assume that the examining boards ensure that comparability of standards does exist, especially in examinations which claim to have national currency.

Forrest & Vickerman (1982)

The report concluded that:

Subject pairs analyses have... thrown new light on comparative standards in examinations in the same subject provided by different boards, both where inter board projects in the GCE are concerned and also in the case of the joint 16+ examinations provided by the consortia in which CSE boards work with the JMB. The general picture which emerges is that, as far as the performances of the JMB candidates offering these particular examinations are concerned, the examining boards, both GCE and CSE, are much more successful in achieving comparability between the different examinations they provide in the same subjects than is usually acknowledged by outside commentators on public examinations.

Forrest & Vickerman (1982)

Whether that conclusion was shared outside the examining community is a matter for speculation as the report was couched in dense and technical language, impenetrable for the general reader. How to explain specialised issues effectively to different audiences continues to challenge researchers and examining boards.

Increasingly, the government expressed its interest in all aspects of comparability. It was as concerned to have value for public money (from which most examination fees were paid) as to be assured that a national standard existed across the examining boards. The House of Commons Expenditure Committee considered the issue in 1976 (House of Commons Expenditure Committee, 1977).

This prompted the examining boards collectively to publicise their comparability research more widely. Two publications in 1978 (Bardell *et al.*, 1978) and 1985 (Forrest & Shoesmith, 1985) described the numerous studies that had been conducted between 1964 and 1985. The preface of the 1978 publication explained why the Secretaries were trying to reach a wider public:

In a climate of growing public interest in public examinations comparability of grading standards is a popular focus of attention; and of the various aspects of comparability – between subjects, between standards in a subject in different years, between modes of examining and between boards – the last usually generates the most earnest and heated debate.

Bardell *et al.* (1978)

Ironically, while the examining boards were working together increasingly to demonstrate comparable standards, matriculation requirements – a major driver for the 1917 School Certificate system – continued to diverge, as the JMB's 1956 annual report makes clear:

There are divergences of opinion between the Board and the other matriculating boards as to what university entrance requirements should be. In consequence automatic 'interavailability' comes to an end with the introduction of revised requirements on 1 November 1956. The achievement of interavailability for the period, 1951–1956, was an advance in educational procedure in this country, even though Oxford and Cambridge Universities have both stood apart from the agreement.¹⁸

While GCE A level was regarded as the basic general entry requirement, some universities refused to accept for entry purposes new subjects, including the AEB's sociology and the JMB's general studies. Use of English, on whose design the JMB, Cambridge and Oxford collaborated and which was designed to improve the standard of communication of entrants to higher education, also met with opposition, in spite of evidence that some undergraduates had a poor command of written English:

Among freshmen in general the level of ability to write English is disappointingly low. The suitability of the present GCE Examination in English Language at the Ordinary level is not here being criticised so far as it concerns the 16-year old candidate for whom it was designed, although opinion about this aspect of the examination is not wholly favourable.

It seems to be generally agreed however that the degree of ability to express oneself which might be accepted from the 16-year old candidate is not sufficient at university entry, that too often apparently such facility as may be present at 16 is not encouraged to develop *pari passu* with the development which goes on in the other aspects of the Sixth form curriculum. It may well be that if all the students were sufficiently 'literate' at entry, some of them might lapse into comparative 'illiteracy' while at the university unless care were taken to ensure that further development is actively encouraged and fostered within the university itself. That is a matter for the university authorities themselves; the duty of the Board is to ensure that at entry those who have been examined by the Board as potential university students have gone further than what is now accepted as O-level English Language.¹⁹

In the event, 'use of English' never became a universal requirement, being regarded by some universities as an additional obstacle to recruitment, and the examination disappeared in due course, demonstrating the importance of higher education's buy-in to new initiatives. This would be an important factor in the Curriculum 2000 reform of A level and the acceptance of vocational examinations as an alternative entry route.

2.3 The Schools Council and the CSE examination

In 1964 the Schools Council replaced the SSEC as the national overseeing body with a remit to co-ordinate the examining system. The Council worked in partnership with the education community and the examining bodies. Local Education Authorities (LEAs), teacher unions, subject associations and the examining boards were among the several organisations represented on its large, unwieldy governing council. The arrangement was in keeping with the freethinking of the 1960s, a far cry from later national overseeing bodies, which, with every new manifestation, accrued ever-greater powers of control over, and regulation of, the examining system. As its numerous publications demonstrate, the partnership was productive: the Council encouraged innovative approaches to teaching, learning and assessment, and sponsored individual and collaborative research projects by examining boards and university education departments (for example, Christie & Forrest, 1980, 1981). Bodies such as the National Foundation for Educational Research (NFER) carried out research into assessment techniques (Willmott, 1977), methods of moderation and comparability (for example, Skurnik, 1974). In an era when it was possible to make public statements that awards could err by +/- one grade without causing a general furore or destroying confidence in the integrity of the system, Willmott's study, based on 1973 data, was able to conclude that standards of the GCE examining boards were 'roughly' comparable (Willmott, 1977).

A new examination, the Certificate of Secondary Education (CSE), came into being in the 1960s to accommodate the ability range immediately below that taking the GCE:

Assuming that up to 20% of the total 16-year old age group may be expected to attempt GCE O-level in four or more subjects, we think the examination we propose might be taken in four or more subjects by candidates in the next 20% below these, and should be so designed that a substantial majority of pupils within this group would obtain passes in this range of subjects. We think that up to a further 20% of the age group might attempt individual subjects.

Ministry of Education (1960)

Fourteen new examining boards were established in England and Wales (a fifteenth, in Northern Ireland, came into being later) to administer the new examination on a regional basis. Appendix 2 shows the geographical distribution of the new CSE boards. With the addition of these new examining boards, the total number of CSE/GCE boards rose to 22, thus exacerbating the problem of comparability across providers. With the exception of the Associated Examining Board, brought into being in 1953 under the sponsorship of City and Guilds to provide a broader range of syllabuses and examinations, many of them vocational, and targeted at the growing further education population, the GCE examining boards were under university governance. By contrast, LEAs played a prominent role in the CSE boards' creation and their chief executive officers, rather than vice-chancellors, chaired their governing councils. Although teachers had always been involved in the policy and subject development committees of the GCE examining boards, they were the majority interest on the corresponding CSE committees. Teachers were also active as examiners and as moderators for coursework assessments, a feature of all CSE examinations.

3 Comparability issues relating to the CSE examination

3.1 Links with GCE O level

In order to benchmark the new examination against an established standard, the CSE's highest grade was linked with the GCE O level pass standard with the minimum point of CSE grade 1 (the highest of the examination's five grades) defined as equivalent to the minimum point of the O level pass standard. No formal mechanisms existed across the CSE boards to effect equivalence between awards made on different syllabuses and question papers, designed for different ability ranges and administered by separate organisations with their own decision-making processes. There were, however, several Schools Council-sponsored GCE/CSE cross-comparability exercises (such as Willmott, 1977) which, among other findings, highlighted the difficulties of comparing grades across regions of the CSE boards as well as between CSE and GCE O level. Not only were the sizes of the regions different but the nature of the schools within them also varied. Some comprised LEAs that had abolished selection and changed to comprehensive educational provision; in other LEAs selection continued to characterise secondary education. As Willmott observed:

School differences are in some cases enough to confuse the application of the chosen analytical method.

Willmott (1977)

The link with the GCE O level standard was a mixed blessing. On the one hand equivalence between CSE grade 1 and the GCE O level pass standard was essential to the public's acceptance of the examination. On the other, the newcomer, catering for a different range of ability, was inevitably seen as the poor relation of the established examination. A generation later, GNVQ and other vocational examinations would run into the same problem. The CSE's difficulties were compounded when formal and certificated grades of A–E, with D and E below the old pass standard, were introduced into GCE O level in 1975; a lower O level grade proved more attractive than a CSE certificate to many students and their parents. The GCE examining boards took steps to ensure a common interpretation of the new grades: the University of Cambridge Local Examinations Syndicate's (UCLES) Annual Report for 1976 records that:

...to ensure comparability of standards between the GCE Boards at grades D and E, a series of inter-board agreement trials in each of eight widely taken subjects was held in the early part of the year.²⁰

3.2 Question choice

Both GCE and CSE external examinations provided candidates with a choice of questions, a further complication for comparability within and between examinations and across the two systems. In Willmott & Nuttall's research into the reliability of 16+ examinations (Willmott & Nuttall, 1975), a GCE paper in their sample provided a choice of five out of ten questions while a CSE mathematics paper allowed a choice of five out of twenty-four questions. The possible combinations across the entry for the latter examination in particular were huge, raising questions about the internal consistency of the examinations. Research of this kind, together with more clearly defined objectives and content which would be examined, led in due course to a drastic reduction in the choice of questions within examination papers. In turn, more focused examining would result in a narrowing of what was taught and learned – assessment would increasingly come to drive the curriculum.

3.3 Curriculum freedom

Choice of questions within examination papers reflected the mood of the times. The CSE examination was intended to follow and support the curriculum; the 'curriculum comes first' was the mantra of the 1960s. The blossoming curriculum development movement of the period looked to the examining system to legitimise new projects, new subjects, new approaches to learning, with the support and encouragement of the Schools Council (Schools Council History, for example). Both CSE and GCE examining boards were keen to cooperate with projects instigated by foundations such as Nuffield, with new innovative alternatives offered to schools as alternatives to more traditional approaches. The practice of the GCE boards was for one of them on behalf of the others to provide the examination as long as it was in project form. All of the GCE examining boards reserved the right to provide separate examinations once the period of the project had ceased.

The Schools Council encouraged diversity, and the structure of the CSE examination was a vehicle for that. The examination embraced three approaches to assessment:

Mode 1 syllabuses and examinations were set and marked externally by the boards; in Mode 2 the board set and marked examinations on a school's own syllabus; Mode 3 provided for the school to determine its syllabus, set its examination and carry out the marking subject to approval and moderation by the board. Although Mode 3 was possible in the GCE system, it flourished in the CSE, with a huge growth in the number of schemes following the introduction of the examination.

The growth of Mode 3 schemes and the absence of criteria for subjects and awarding compounded the complexities of setting and applying a national standard consistently across examinations within and across CSE examining boards. Approval and moderation procedures for Mode 3 were varied. The Associated Lancashire Schools Examining Board (ALSEB), the smallest of the CSE boards whose first Secretary had been recruited from the JMB whose stringent procedures it adopted, exercised tight control over syllabuses and examinations through its subject panels. The panels were responsible for both Mode 1 and Mode 3 syllabus approval and awards, in order to apply common judgements to subject requirements and standards of award.

At the other extreme the West Yorkshire and Lindsey Regional Examination Board (TWYLREB) implemented the concept of curricula designed to meet local needs to a greater extent than any other of the CSE boards; its provision was almost entirely Mode 3 with only a handful of Mode 1 examinations. It handled Mode 3 quite separately from Mode 1 on the basis that imposing its view of what a subject should be would stifle innovation. Schools in TWYLREB's region were therefore afforded greater freedom to pursue their own ideas; it was said that at its height there were 10,000 Mode 3 schemes in Yorkshire alone. Whatever the benefits to teaching and learning, this liberal approach to subject content and the standards of awards left the examining board open to the criticism that standards within and across modes were not comparable.

This criticism was somewhat unfair as all the boards, including TWYLREB, opened their doors to research into their standards, particularly into the equivalence of grades awarded in Modes 1 and 3. This research, while detecting some tendency towards leniency in Mode 3, showed no startling differences (Nuttall, 1973). The leniency was considered to relate to the greater motivation of candidates and teachers alike on schemes that had been designed by teachers for specific groups of students:

It could be argued that apparently high performance in Mode 3 is only to be expected when a teacher decides not only what to teach, but what to assess and how and when it should be assessed.

Matthews (1985)

Mode 3 courses in all CSE examining boards increased in number and diversity after the raising of the school leaving age to 16, with innovative schemes designed to meet the needs of this new cohort coming on stream in 1974. Many of these – for example, 'preparation for living' – were criticised for their standard and for diluting the curriculum entitlement of students, some of whom dropped basic subjects like

mathematics and science at age 14 – a practice which the introduction in 1988 of the National Curriculum with its required range of subjects, including English, mathematics and science, brought to an end. These new courses added to concerns about the value of Mode 3 schemes and their comparability with externally provided syllabuses and assessments. The 1977 Annual Report of UCLES, the Cambridge board, expressed a commonly held view:

The Syndicate questions whether the use of many thousands of Mode 3 examinations, each set and marked by individual schools or small groups of schools, is consistent with the notion of a national standard and national qualifications.²¹

3.4 Coursework

The assessment of work that students undertook during the course of study was central to CSE examinations and also featured in some GCE schemes. The JMB, for example, had introduced an English O level in 1967 based entirely on coursework, which proved to be extremely popular and was the forerunner of the Northern Examinations and Assessment Board's (NEAB) joint 16+ (later GCSE) syllabus. The examination continued until 1994 when the rules governing coursework were tightened and 100% course-assessed examinations were proscribed. The increased use of coursework added to the difficulties of applying comparable standards across schools, regions and nationally.

While the assessment of work undertaken during the course of study widened the scope of what could be assessed and offset the limitations of end-of-course examinations, there was no agreement on what coursework constituted: was it normal, day-to-day class work (continuous assessment), or specific pieces of work designed to assess particular objectives – for example, a field study in geography or an historical enquiry? How to standardise teachers' assessments to achieve comparability between individual assessors and across schools was a major challenge – in essence the same problem as standardising examiners but on a greater scale and more geographically dispersed.

Methods of moderation included sampling work on site or by postal exchange, and moderation by consensus (that is, teachers meeting to assess work, discuss and agree standards). The efficacy of the system depended on the quality of the assessments of both teachers and moderators, which raised questions about the training they had undertaken in order to make valid and reliable assessments. The consensus approach was later to be favoured as educationally beneficial and effective in the report commissioned by the government, prior to the introduction of the National Curriculum, from the Task Group on Assessment and Testing (DES/WO, 1988a), but its recommendations were judged too costly to implement.

The provision of equal opportunities to all students to demonstrate their attainment through work that was sufficiently challenging was a further factor. Parental pressure on teachers, the potential for cheating and the possible influence of social class on attainment were also concerns that exercised researchers and sociologists (Broadfoot, 1979). The debate about these and other issues is as lively in the 21st century as it was

in the 1960s, with new factors such as access to the Internet adding to concerns about the reliability of coursework. The QCA's response in 2006 to those concerns was to restrict coursework to the classroom in order to prevent work being plagiarised from the Internet.

Regional provision, school-based examinations and coursework assessments militated against a national standard despite the best efforts of the CSE Secretaries who, like their GCE counterparts, worked together in the pursuit of comparable standards. Strong links were forged between the CSE examining boards and researchers working for other educational organisations such as the NFER. The appointment in 1976 of Desmond Nuttall, a prominent researcher, to the post of Secretary to the Middlesex Regional Examining Board further illustrated the value afforded to the contribution of research to examination development.

With the encouragement and support of the Schools Council, the CSE boards collaborated on cross-moderation exercises, developed methods of moderation, worked with teachers to raise awareness of standards, using marked examination scripts and other materials. However, regardless of these efforts, the complex structure of the examination meant that comparability was an elusive concept. Despite this weakness, however, the CSE examination provided long-overdue recognition to the achievements of a growing number of school leavers who otherwise would have had no formal externally validated record of their educational success. Entries grew rapidly in the first years of the examination from 230,977 in 1965 to 982,721 in 1969 (Gillan, 2003).

3.5 A single system of examining at 16+?

It seems to be the fate of all examinations, in England at least, to be criticised as soon as they have come into being as being irrelevant to prevailing educational opinion: it was true of the GCE in the 1950s, of the CSE in the 1960s, of the GCSE after 1988 and of the revised A levels in 2002. Such criticisms highlight the time taken to implement new concepts in systems where new courses, generally lasting for two years, need to have been approved at least 18 months prior to implementation. Turning educational and assessment systems around takes time and careful planning – and during that process other developments and changes can make 'new' thinking seem outmoded.

In the 1960s the two end-on systems of CSE and GCE O level theoretically catering for 40–60% of the school population (in practice a wider group) co-existed uneasily in an expanding system of comprehensive education. Comparability between the standards of the two examinations was open to question; employers and the general public were confused by two grading systems; two administrative structures duplicated costs, time and effort without adding value to students. In 1971, the Schools Council determined to investigate the feasibility of combining the GCE O level and CSE examinations and invited the examining boards to set up consortia to develop joint 16+ courses for pilot examinations involving a limited number of students (Schools Council, 1971); these feasibility studies came to fruition in 1974. Many of the GCE and CSE partnerships that were formed at that time stayed in

existence and continued to offer examinations leading to the award of both O level and CSE certificates. The JMB provided the GCE element for three consortia – Associated Lancashire Schools Examining Board (ALSEB)/North West Regional Examinations Board (NWREB); The West Yorkshire and Lindsey Regional Examining Board (TWYLRREB); The West Midlands Examination Board (TWMEB) – which were responsible for 15 of the 42 feasibility studies carried out nationally. The JMB and the four northern CSE boards (ALSEB, North Regional Examinations Board (NREB), NWREB, Yorkshire Regional Examinations Board (YREB)) under the umbrella of the Northern Examining Association (NEA) continued to offer the subjects of the original feasibility studies and additional subjects were developed as joint 16+ examinations.

The 16+ debate coincided with the government's growing interest in education and its relevance to the modern world. Prime Minister Jim Callaghan's speech in October 1976 at Ruskin College, Oxford, questioned whether education equipped school leavers with the tools required by industry and with the basic skills of literacy and numeracy, recurring themes over the next 30 years. The speech opened the way to more central control of what was taught and assessed in schools and, therefore, of examination syllabuses and their standards. Vocational provision, so long on the fringe of the national examining system, became a priority, with the launch of government-funded initiatives such as the Technical and Vocational Education Initiative (TVEI), intended to bridge the academic and vocational divide.

Against this background the debate about a single system of examining at 16+ was played out over several years. The government set up a committee, chaired by Sir James Waddell, which in 1978 recommended a single system, subject to the development of general and subject-specific national criteria in the main subjects (20 in all) with aims, objectives, content and skills and more specifically targeted examinations (DES, 1978). Their purpose was to provide a firmer basis for comparability across subjects and providers by tighter control of syllabuses, examinations and administration. The CSE and GCE examining boards established a Joint Council for the GCSE to develop these criteria and in 1984 the government agreed to the introduction of the General Certificate of Secondary Education (GCSE) for first examination in 1988. Four geographically based Groups in England (plus one in Wales and a sixth in Northern Ireland), each comprising at least one GCE and one CSE board, were approved to develop and administer the new examination in England and Wales. They were charged with carrying forward the O level/CSE standards with GCSE grades at C and above to be determined by awarders drawn from the GCE tradition and the lower grades to be determined by those drawn from the CSE tradition. In spite of these strictures, most awarding committees worked as one to determine the grades. The new Groups were voluntary partnerships; however, mergers followed and by the mid-1990s four merged examining boards offered the GCSE examination in England. Fewer syllabuses followed the reduction in providers, giving a better basis to achieve comparability across subjects.

Other changes to the examination system supported the tighter control of syllabus provision. The Waddell report had recommended that:

Arrangements for the central coordination of 16+ examinations should be strengthened and a central body should be responsible for securing agreement on criteria... and for coordinating further preparations.

Department of Education and Science (1978, paragraph 127 (iv))

Accordingly, the Secondary Examinations Council (SEC), which replaced the Schools Council in 1984, was given a stronger remit than any of its predecessors to approve syllabuses and co-ordinate examination provision. Responsibility for the school curriculum was given to a separate body, the School Curriculum Development Committee (SCDC). The examining groups saw the need for even closer collaboration to provide self-regulation and to agree and apply comparable standards. The mechanism for this collaborative work was the Joint Council for the GCSE, supported by a forum of GCSE Secretaries. The current co-ordinating body, continuing the role of the original Joint Council, is the Joint Council for Qualifications (JCQ), established in 2004.

None of these arrangements for the co-ordination of the GCSE affected the governance of GCE A level, which remained the responsibility of the independent GCE examining boards. Their Secretaries continued to meet to discuss A level issues until the emergence of new boards such as London Examinations and the NEAB whose responsibilities embraced both 16+ and post-16 examinations.

4 The GCSE examination

4.1 Tiered examinations

Before giving approval to the introduction of the GCSE examination the Secretary of State had to be satisfied that a single *system* was able to assess the full range of ability. Would GCSE be an umbrella system with different syllabuses or sub-sets and/or papers targeted at different ranges of ability? Or would it be a single *examination* with a single syllabus and assessment tasks for the whole ability range? From the outset, UCLES took the view that in an examination spanning the ability range accommodated by GCE O level and CSE:

The only way to prevent a substantial fall in standards is to create a tiered examination with different but related syllabuses and question papers for different ability groups, and with regulations which will ensure that the abler students take tests which will extend them to their full capacity.²²

In the debate that followed, the Schools Council commissioned further research into examining across the ability range (Tattersall, 1983) and a view emerged eventually that, while all GCSE examinations had to differentiate effectively across the ability range, the nature of subjects should be the main determinant of how this would be achieved. In broad terms subjects fell into two categories – those which set common tasks with a range of responses, that is, differentiation by outcome; and those which set papers designed for sub-sets of the ability range with a restricted number of grades available on the papers, that is, differentiation by task using tiered examinations. As time went by the rules for differentiation changed and tiered papers became the norm from which few exceptions were permitted. Tiered content

also characterised some GCSE subjects, requiring early choice of courses, not unlike the choice of CSE or O level course, which many had thought the single system would make unnecessary.

Overlapping grades on tiered papers were new challenges to comparability and fairness and prompted a raft of research (Good & Cresswell, 1988), which addressed questions such as: is a grade C, which is achieved on the higher tier – where the grade is the lowest available – of the same standard as a grade C attained on a lower tier where it is the highest available grade? Are the demands of different papers and questions, which target some or all of the same grades, comparable? Is it right to limit the aspirations of students to papers with a cap on the grades that can be awarded? Is it right to give no recognition to the attainments of students who fail to achieve the lowest grade on the higher tier? Do awarders apply consistent grading standards across differentiated papers? How to explain to the public the intricacies of a system that rewarded candidates highly for a seemingly limited performance, as exemplified by low marks for grade C on the higher tier?

4.2 Grade-related criteria

The system was spared the difficulties that would certainly have arisen had explicit criteria, as requested in 1984 by the then Secretary of State, Sir Keith Joseph, been adopted for the award of grades. His speech to the North of England Conference in January 1984 touched on clearer definitions of the objectives of examinations:

First, I can offer an account of what the minimum level to be attained at 16 by 80–90% of pupils would entail in a few areas of the curriculum, taken by way of example, and I repeat that I am talking about a minimum level. In English pupils would need to demonstrate that they are attentive listeners and confident speakers... that they can read straightforward written information and pass it on without loss of meaning, and that they can say clearly what their own views are....

Joseph (1984)

He went on to say:

It is clear that one cannot compare what pupils now achieve in the largely norm referenced 16+ examinations with the objective [of raising standards]... because that objective can be fulfilled only if the examinations become more criterion referenced... The more the examinations can measure in absolute terms... the easier it will be to motivate pupils to attain higher absolute standards by a proper acknowledgement of what they can do.

Joseph (1984)

Sir Keith's expectations were that 'grade-related criteria which will specify the knowledge, understanding and skills expected for the award of particular grades' would be developed to ensure a clearer meaning for pupils and clearer goals for teachers (DES, 1987). The SEC's Grade Criteria Working Parties' approach fell short of the Secretary of State's expectations but was complex, unwieldy and unlikely to achieve consistency within and across subjects, or be comprehensible to a wider audience:

... numerous and complex criteria – in the case of history, for example, ten sub elements across three domains, and criteria for four levels of performance within each sub-element, resulting in forty statements of performance for teachers and examiners to use in the assessment of candidates and for those interested in the performance of candidates to interpret.

Department of Education and Science (1987)

The initiative died a death, other than taking the much weaker form of grade descriptions, but the search for absolute criteria to define grades and levels of achievement resurfaced with the National Curriculum and its ill-fated numerous Statements of Attainment. The comments of an AEB research publication are pertinent to both failed initiatives:

It is linguistically naïve to believe that criteria... can ever be made sufficiently precise for their use not to involve subjective judgements of the type which they are intended to avoid... It is technically naïve to expect the use of complex aggregation rules to enable detailed descriptions of candidates' attainments to be inferred from summary measures like grades or that such rules, because they are explicit, necessarily operate in a way which is consistent with natural notions of fairness... It is philosophically naïve to assume that fair judgements can only be made if every candidate's script is judged by precisely the same set of criteria... It is psychologically naïve to assume... that performance is not profoundly affected by the context of the task being carried out.

Cresswell (2000)

4.3 Tighter curriculum controls

GCSE National Criteria carried the government into the secret garden of the curriculum, transforming its undergrowth into an orderly public park. Government recognised the potential of assessment for bringing about changes to the curriculum, influencing teaching and learning and monitoring educational standards. The 1988 Education Act brought about the National Curriculum and new central overseeing bodies, the School Examinations and Assessment Council (SEAC) and its sister organisation, the National Curriculum Council (NCC). They were given statutory powers over the National Curriculum and its assessment at four key stages (at ages 7, 11, 14 and 16), including GCSE syllabuses and examinations, a seismic shift from coordination to regulation of the system. The existing GCSE National Criteria were changed to accommodate the National Curriculum and tightened to enforce more explicit comparability across syllabuses in the same subject. New constraints on the numbers of syllabuses each examining board could develop were introduced, which led to a further cull of syllabuses, particularly Mode 3.

With every change of the rules, new syllabuses were required – more frequently than had ever previously been the case. This had compounded the difficulties of maintaining consistent standards from year to year, with no stable benchmark against which to measure curriculum requirements and examination standards.

Only those 16+ syllabuses on a list approved by the Secretary of State could be used in state-funded schools – reminiscent of the School Certificate. A narrow definition of 'General' excluded from the GCSE vocational-type subjects such as woodwork and

metalwork, even though these had been assessed by both GCE and CSE examinations. However, broader subjects such as business studies; craft, design and technology (CDT); electronics; and travel and tourism continued to provide a vocational route in GCSE whose standards were deemed by Ofsted in 1996 to be 'broadly in line with those in other GCSE courses' (Ofsted, 1996a). In later years the policy of excluding vocational subjects from GCSE would be reversed through the development of Vocational GCSEs.

From an early stage, the GCSE was thought by some to be an 'anachronism' (Nuttall, 1993), having been in gestation for 18 years and its introduction coinciding with the development of a statutory National Curriculum. A national system of assessment opened the possibility of abolishing the seven grades of GCSE and replacing them, in the National Curriculum subjects at least, with the ten-level scale, which reported attainment at the end of key stages. Some believed that the scale would provide a more positive report of attainment at age 16, that it would raise aspirations, open horizons and provide better differentiation. A single scale across the four key stages would also facilitate the tracking of students' progress from primary to the end of secondary education. Whatever its educational merits, the proposal for a single scale raised questions about the relationship of ten National Curriculum levels to seven GCSE grades and there were doubts as to whether a change to a single scale would maintain consistent standards:

[There are difficulties with a proposed] move from the lettered grades A–G, designating GCSE performance, to the 10-level numbered scale in which 10 and 9 differentiate within the present A grade, but 7 and 6 do not match the present C and D, and 4 will be the lowest grade to be reported by the GCSE. Here there are not only problems of achieving comparability between the two scales, but also of providing the more searching test looked for at the top end within an examination to be taken by the great majority of the age group.²³

Had the proposal been adopted solely in the ten subjects of the National Curriculum, a ten-point numbered scale would have run side by side with a seven-point lettered scale. That prospect was not attractive and neither was the option of changing to a numbered scale for all GCSE subjects. The proposal was quietly dropped.

5 An accountable and regulated national assessment system

Four interlocking themes dominated the post-1988 era, converging in the last years of the century to create an accountable and regulated national assessment system: regulation, accountability, standards over time, and post-16 education and assessment. These developments were both driven by and raised the stakes for comparability. Comparability was a key factor too in the drama that developed as the new system came on stream.

5.1 Regulation

The 1988 Education Act, together with GCSE National Criteria, marked the end of a long tradition of a largely decentralised and unregulated approach to teaching,

learning and assessment. Thereafter, the curriculum and assessment came under tighter central control; the remit of the central overseeing body (see Appendix 3 for an overview of the English advisory and regulatory boards governing examinations) was strengthened from one of loose coordination to tight regulation. Devolution in the late 20th and early 21st centuries to Wales and Northern Ireland of responsibility for education required separate regulatory organisations in each country and co-ordination across all three of them to ensure consistency in qualifications that crossed national boundaries.

Initially, the statutory regulatory powers over public examination syllabuses and assessments governed only the period of compulsory education to age 16; the control of post-16 examination provision did not come on stream until relatively late in the day. Control of syllabuses and assessment procedures became more stringent with every new manifestation of the central overseeing body: the SEAC (from 1988) exercised powers over the assessment of the statutory curriculum that the GCSE examined at Key Stage 4. The regulator oversaw revisions to National Curriculum assessments and to GCSE arising from the plethora of legislation, reports and enquiries which appeared in the 1990s. Arrangements for the GCSE were tightened to ensure better comparability; coursework, for example, was capped in favour of seemingly more reliable external, end-of-course examinations. A code of practice (initially for GCSE but extended later to other examinations) was developed with the aim of minimising differences in procedures that might have a bearing on comparability; awarding procedures, for example, were codified. Scrutinies and probes into examinations were carried out on a more frequent basis to expose and rectify any perceived weaknesses. The force of regulation waxed and waned over the years: self-regulation came and went in the mid-1980s; the heavy handedness that followed gave way briefly to a light touch in the early 1990s.

A not insignificant driver behind the pressure for increased regulation was a nagging fear that the examining boards might not simply be competing in terms of the quality of their syllabuses and support to teachers, but they might also be competing in terms of (lower) attainment standards (making it easier to get a grade with one examining board rather than another). There was nothing new about such fears, as the 1894 correspondence in the *Pall Mall Gazette*, quoted earlier in this chapter, demonstrates.

However, the medium of television was able to plant the notion of nefarious examining and awarding practices in the minds of a far larger audience than a newspaper could ever hope to influence: the introduction to the 1995 Channel 4 television programme *Dispatches* claimed:

On *Dispatches* tonight, an investigation that goes right to the heart of British education, casting the gravest doubt on any claim that standards are rising. Sarah Marris penetrates the closed world of the examination system. A story of papers deliberately made easier and grades deliberately made better; all in the name of competition and market forces. It's an exam system where money talks.²⁴

A complaint of unfair and unjust treatment, brought by a member of the AEB's staff who had featured in the programme, was upheld by the Broadcasting Standards Commission who ruled (June 1997) that 'the theme of the programme – deliberate and secretive grade-rigging by the examining boards – was at best an exaggeration and, accordingly, unfair'. This ruling offered a vindication of the integrity of those involved in the examining process. However, although the ruling made clear that there was no persuasive evidence that examining boards manipulated their standards to attract more candidates (merely speculation, typically based upon a naive understanding of the system and of the significance of differential pass rates), such allegations were damaging and helped to pave the way towards greater regulation of the examinations system.

As statutory powers over qualifications related initially only to the period of compulsory education, they did not extend to GCE A levels, which were taken, in the main, by post-16 students and were still governed largely by university-dominated councils. However, the GCE examining boards were under increasing pressure to conform to tighter controls: A level principles and subject cores were introduced to improve comparability and rationalise the number of syllabuses on offer. The cores were developed with the somewhat unwilling cooperation of the GCE boards, which were conscious of the restrictions on the freedom of the GCSE examining boards (that the National Criteria had brought) to develop the curriculum as they wished. Although the cores were a far cry from National Criteria, allowing considerable flexible interpretation, they nevertheless defined an agreed body of knowledge and skills, which enhanced, to some extent, consistency across A level syllabuses. In 1994, following a 1993 HMI inspection of the quality and standards of A/AS level examinations, a voluntary code of practice was drawn up for GCE, which was merged with the GCSE Code (mandatory from 1994) in 1997. The code laid down strict procedures for all stages of the examining process, from syllabus design to awarding, its intention being 'to promote quality and consistency across all examination boards and thereby minimise differences which hindered... (the achievement of) common and unchanging standards' (NEAB, 1996, p. 1). The 1994 GCE Code restricted the role of coursework (an upper limit of 20% in the majority of subjects) in A level. Conditions were also laid down by the regulator for the development of modular schemes, reflecting the concerns that Ofsted had expressed in the 1993 report about the comparability of standards of A level linear and modular schemes.

The universities' 'ownership' of the GCE examining boards protected them from undue interference from SSEC (the Secondary Schools Examination Council) and its successor organisations. However, the universities' stake in the system was weakened by the administrative separation of the GCSE and the GCE in 1988. The mergers of examining boards, which followed that change led most of the universities to withdraw from the governance of examining boards. By the mid-1990s, only Cambridge and the northern universities remained an integral part of the examining boards they had created in the 19th century. Durham had long disappeared from the scene; JMB's formal ties with its five founding members were loosened when it became part of the NEAB and then dismantled when the AQA was

created, although the new organisation sought higher education representation from Universities UK; Oxford sold its interests to UCLES and the AEB; London took a back seat when its examining arm merged with BTEC, bowing out entirely when Pearson acquired the business in 2002.

The dilution of university influence on the school examination system facilitated the greater involvement of the post-1988 statutory regulators in the examining boards' GCE operations; the autonomy which they had enjoyed for so long disappeared. Under SCAA (from 1993) and later QCA (from 1997), regulation of syllabuses (or 'specifications' of what was to be examined, as they came to be called) and assessment was extended to all examinations, including A level and those vocational examinations that claimed equivalence to either GCSE or A level – GNVQ and, later VCE. National Vocational Qualifications (NVQs) came into the framework too, but it proved more difficult to regulate vocational boards, which regarded such attempts as threats to their commercial existence. Some of the smaller boards were willing to take their chances in the marketplace with unregulated products.

By the end of the 20th century, the concept of accreditation had been introduced in England. This brought all aspects of examining board governance and procedures under the scrutiny of the QCA in its regulatory role. In particular, the examining boards had to demonstrate that their quality assurance procedures were sound and that they conformed in full to the QCA's requirements, including the code of practice. While these steps minimised the opportunities for different procedures which would impact on comparability of standards later, events would reveal the difficulty of controlling the interpretations and actions of human beings.

5.2 Accountability

Accountability became a central requirement of public service organisations in the last decades of the 20th century. In education, the comparability of assessment outcomes was thrown into sharp focus when performance tables were introduced in the mid-1990s to monitor the system. These tables, colloquially known as 'league tables', ranked schools according to the performance of their candidates in the GCSE and GCE A level examinations without, in the first instance at least, taking into account social and other factors that could influence examination performance. Schools and colleges were, therefore, held to account for the grades their students attained.

In most countries the assessment of individuals and the monitoring of the system are different functions, and they had been in England when the Assessment of Performance Unit (APU) operated from the mid-1970s until the advent of the National Curriculum. In reality, the examining system's disparate administration and complexities were ill-suited to a role wider than its prime function of assessing and reporting the attainment of individuals. The gradual inclusion in the performance tables of vocational as well as academic assessments compounded the difficulties of guaranteeing comparability across the data. By means of the tables, teachers, schools, LEAs and examining boards were held to account for the attainment of their students and, in due course, for the value they added to their educational progress. Powers of

inspection of education were strengthened. National Targets were set to raise the standards of education. A consequence of these developments was changes in the pattern of entries for individual examining boards as schools sought courses that would deliver the best outcomes for their students, fuelling claims, not unlike those made in the 1890s and by the *Dispatches* television programme, that awards were affected by market considerations and that the examining boards' standards were, therefore, inconsistent.

As greater precision and reliability were demanded of the examining system by individual candidates and their parents, and by schools anxious to demonstrate that their place in the 'league tables' was accurate, questions were raised not only about the accuracy of marking but about standards across the examining boards. The examining boards' reputation for fairness, integrity and probity came under increased scrutiny, leading to further demands for greater controls over the system. A single examining board was seen by many as the solution to comparability. However, the fundamental characteristics of the examining system – loose criteria, qualitative questions and marking – which made comparability a difficult issue to address, would continue to manifest themselves, regardless of whether there was a single or multiple providers.

A further consequence of the pressures on the system to deliver greater reliability was an increase in requests to the examining boards for special arrangements (extra time, for example, for students with special assessment needs) and consideration (where particular circumstances such as bereavement might have impaired an individual's performance). Challenges to published grades and National Curriculum levels also increased, a reflection of the concerns of schools that their published data were inaccurate. Greater degrees of accuracy, reliability, consistency and precision were sought from the system than examinations based on qualitative rather than quantitative judgements could deliver. Concerns about consistency across providers, between subjects and across years gathered pace and motivated successive regulators to control more tightly the activities of the examining boards.

5.3 Standards over time

The use of public examinations data for reporting attainment relied on there being a constant, unchanging standard, maintained over time, within and across subjects and providers. In reality this was not a straightforward matter, as acknowledged by SCAA/Ofsted (1996). Syllabus and examination changes, different styles of questions, new knowledge and society's expectations all combined to obscure comparability across time. Earlier reports on the issue (Backhouse, 1978) had arrived at similar conclusions in a period of much less change in syllabuses and examinations. Even though it was impossible to *prove* that standards had remained roughly comparable in a rapidly changing world, the examining boards' own researchers expended efforts to expose the issues and reassure the regulator and the public that all possible steps were being taken to underpin consistent standards (Massey, 1994; Newbould, 1994).

Concerns that standards had fallen were particularly acute at GCE A level, which had remained remarkably stable until the 1980s when both curricular and grading changes were made. On the curriculum front, subject cores led to a flurry of new syllabuses, which, while being more relevant to students and providing a greater degree of comparability across examining boards, unsettled what had been a largely unchanged examination. In terms of grading, the original 'Pass' and 'Distinction' categories had been replaced in 1963 by a new A–E grading system accompanied by SSEC guidelines for the proportions of candidates within each grade band: 10% A, 15% B, 10% C, 15% D and 20% E. Although the GCE boards themselves were critical of the grading scheme and did not follow the guidelines uniformly (Whittaker & Forrest, 1983), they remained the official guidelines for awarding until 1987.

In that year the examining boards were asked by SEC (still in the full flush of its search for absolute grade criteria) to award A level grades on the basis of examiners' judgement of the quality of work at three key boundaries, grades A, B and E, the remaining grades being arithmetically determined according to an agreed formula. The shift of emphasis from norm-referencing tempered by qualitative judgements to a weak form of criterion-referencing tempered by statistical data opened the way for a steady annual increase in the percentages of candidates succeeding in A level which, in turn, led to a greater demand for places in higher education. By 1995, 72% of students in England stayed in education beyond age 16, compared with 13% in 1955, while the proportion of 18-19-year-olds who advanced to higher education increased from 4% in 1955 to 31% in 1995.

Charges that the examining boards were lowering their standards, and complaints from universities that the examination provided them with insufficient differentiation to select their intake resulted in the 1996 SCAA/Ofsted investigation into standards over time. Although the investigation acknowledged that statistical shifts over time did not provide evidence of a change in standard, recommendations were made to monitor standards more frequently. A national library of scripts and a system for retention of scripts to assist future studies were established. Worryingly for the examining boards, many of which had already merged voluntarily, thus reducing the number of providers, the study urged the government to act on Sir Ron Dearing's recommendation for a further rationalisation in the number of examining boards to facilitate tighter control of the system.

5.4 Post-16 education and assessment

Successive governments had found it difficult to get a handle on the 16–19 phase of education. A levels were considered too specialised (a long-standing criticism) and, as more students entered further education, inappropriate in respect of both curriculum and standard for the new cohort. A variety of failed initiatives – proposals to extend the grading scale, an Advanced Ordinary level examination and a Certificate of Extended Education with a standard between O level grades A–C/CSE grade 1 and A level grade E – which were intended to address the narrowing effect of A level on students' education peppered the late 1970s and 1980s. In 1988 the Higginson Committee recommended five 'leaner' but 'tougher' A level

syllabuses of a slightly lower standard to broaden sixth-form studies, which would also have accommodated some learners for whom the A level standard was beyond reach (DES/WO, 1988b). The government rejected the recommendations and chose instead to retain the A level with an Advanced Supplementary examination, half the content but assessed at the A level standard which the Government guarded.

Far from broadening the appeal of A level, the new structure raised issues of comparability between the grades of a half or a full A level course, as did full and short course GCSEs in the 1990s. An Audit Commission report was critical of the 30% fall-out rate from A level courses: about one third of students who embarked on A level did not complete the course, and about one third of those who did complete were ungraded (Audit Commission/Ofsted, 1993). The report questioned whether an examination that failed to meet the needs of the majority of candidates represented value for public money.

The development of modular schemes of assessment in the 1990s was one answer to the problem, encouraging more students to continue their A level courses and attain grades that may have seemed beyond their reach had they not had the benefit of periodic feedback on their progress. By 1997 modular schemes accounted for 30% of all entries. Increases in the percentages of students attaining the 'pass' standard led to concerns that a modular approach provided an easier route to grades. Ofsted's 1996 enquiry was, however, confident that standards were comparable across modular and linear A level schemes:

There was no evidence in the 1996 examinations of any significant differences in the standards being set at the grade boundaries, between the modular and linear syllabuses inspected.

Ofsted (1996b)

Alongside the continuing focus on the quality and standard of A level, the Government began to feel its way increasingly to a vocational solution to a broader-based post-16 curriculum. Developments in the 1980s such as the Certificate of Pre-Vocational Education (CPVE) ran up against a similar problem to that which had dogged the CSE – how to ensure a link with an established standard (A level) while encouraging diversity and experimentation. Other attempts (TVEL, for example) to increase the take-up and status of vocational examinations administered in the main by separate vocational awarding boards also came to grief on the altar of comparability.

A different solution emerged (DES/DoE/WO 1991): the establishment of distinct learning routes for post-16 education: A levels (overseen by SEAC), General National Vocational Qualifications (GNVQ) and NVQ under the control of the National Council for Vocational Qualifications (NCVQ).

School and vocational examining boards were urged to work together to administer GNVQ alongside GCSE and A level provision. However, continuing references to the A level 'gold standard' made it nigh on impossible to establish vocational education

as worthwhile in its own right. As Sir Geoffrey Holland, former Permanent Secretary at the Department of Employment, observed in 1995:

A levels, far from being the gold standard that ministers and a lot of other people think they are, are in fact an altar on which have been sacrificed the enthusiasm and the hopes and, indeed, many of the capabilities of about half of our young people.²⁵

5.5 Convergence and resolution

The search for a solution to the problem of linking the various strands of post-16 education in such a way as to ensure comparability and parity of esteem continued, with two major reviews in 1995/1996: Gordon Beaumont's *Review of 100 NVQs and SVQs* (Beaumont, 1996) and Ron Dearing's *Review of qualifications for 16–19 year olds* (Dearing, 1996). Dearing was charged with the responsibility to make more coherent the fragmented post-16 education system. His report's numerous recommendations included creating a single national framework of academic and vocational qualifications; reducing the number of examining boards (from four to three, in line with the three vocational boards offering GNVQ: RSA, BTEC and CGLI) and giving them responsibility for all general qualifications, vocational and academic, thus reflecting the span of responsibilities of the newly merged Department for Education and Employment; introducing an entry-level standard for those students working at a level below GCSE; introducing an Advanced Subsidiary standard (that is, lower than A level); unitising the post-16 curriculum to facilitate broader courses and credit accumulation. New A/AS-level syllabuses would be needed to deliver these particular recommendations.

The government's 1997 response to the review, *Guaranteeing standards* (DfEE, 1997), set the scene for a fundamental overhaul of post-16 qualifications which brought A levels firmly into the National Qualifications Framework, controlled and regulated by the new QCA, which replaced SEAC and NCVQ in 1997. The new body was given powers to regulate qualifications, whatever their genesis. *Guaranteeing standards* envisaged the bringing together of academic and vocational qualifications into a single administrative structure through the creation of unitary examining boards. The clear intention was to address through such structural changes comparability across providers and parity of esteem between academic and vocational examinations. The education and employment communities generally welcomed the proposals.

Guaranteeing standards also pursued the question of how many examining boards a national system of assessment could sustain. A single board was dismissed on grounds of scale and three emerged as the favoured number. A new government in 1997 made clear at an early stage its support for three unitary awarding boards in a press release in June 1997:

We believe that there should be three awarding boards – each offering GCE A levels, GCSE and GNVQs. These boards will be best placed to take forward our commitment to support broader A levels and upgraded vocational qualifications – both underpinned by rigorous standards and key skills.

We also want a reduction in the number of GCE syllabuses... a single point of accountability and a single trading name.

... The new Qualifications and Curriculum Authority (QCA) [will] work with the awarding boards and the Department to develop a rigorous regulatory regime, building on self-regulation where that works best.

We... will expect to see far greater comparability in standards between similar examination syllabuses to avoid some papers being seen as 'easy'.²⁶

The reduction from four to three examining boards could only be achieved through a coming together or merger of two of the existing boards: in the event the AEB and NEAB joined forces to create the Assessment and Qualifications Alliance (AQA), at first a joint venture, but very quickly becoming a fully merged organisation in 2000. The City and Guilds of London Institute, which had been party to the initial discussions to create this new unitary awarding board, chose to remain an independent organisation.

5.6 The Curriculum 2000 A level system in operation

As the new millennium opened, the examination scene was very different from that which had existed throughout the late 19th and 20th centuries. Structural changes in government departments and the examining boards had been brought about by the demand for greater comparability across qualifications. A national framework of qualifications had been created comprising both academic and vocational qualifications. National Curriculum assessments at defined key stages together with the qualifications system provided the data to monitor and evaluate the education system. The newly found stability would be disrupted by the introduction of new A levels for first teaching in 2000, unitised, and with a new mid-way Advanced Subsidiary standard. Curriculum 2000 sparked a controversy which opened the way for further consideration of the 14–19 curriculum and assessment framework.

A new unitised AS/A2 system was fraught with comparability challenges, not least the requirement that the new A level awards in 2002 would be comparable with those of the old A level awards. To deflect criticism that modular, or 'unitised', assessment would be somehow easier than terminal assessment, all Curriculum 2000 A levels were required to include an element of 'synopticity' in the assessment of A2 units (often operationalised as a single synoptic unit). The idea of synoptic assessment was that it should require a student to draw upon ideas from across the entire syllabus, to demonstrate a depth of understanding and to identify links between core concepts.

The system comprised two qualifications, one of which was embedded in the other. Designing papers for an as yet unexemplified standard – the new AS – was far from easy. How to combine three units at AS standard, representing 50% of the total A level – but also leading to an AS award – with three units at the A2 standard (50%) – all six leading to the full A level award – presented a huge challenge to awarders. The robust technical and statistical information, which examining boards relied on to monitor their standards in more settled times, was not available, although, as ever,

the interpretation of statistics would be an issue: would an increase in the percentage of candidates who succeeded in A level represent a lower, the same or an improved standard? The very nature of a unitised structure, even with strict limitation on retaking units, suggested that students would be more likely to reach higher levels of attainment than in a single end-of-course examination – or to drop out before completion, if early module results were disappointing. Prior certificated attainment at AS raised a major question: should the A2 awards be used to keep steady the percentage pass rate? Would students be rewarded on the basis of their performance, or would statistical considerations moderate – or determine – the outcome?

As the first round of the new examinations unfolded it became clear that the statistical outcomes of the new system would look very different from those of the old. The new Advanced Subsidiary qualification had given students and their teachers quality feedback on progress and a prognosis of their likely overall A level award. For students who were on track to attain a good A level grade, the feedback encouraged them to stay the course; for others whose AS performance was more modest, there was the opportunity to drop out with an AS certificate as proof of their attainment. The likelihood was that there would be an enhanced pattern of A level awards, which would bear little comparison with the awards of previous years. There would be no way of knowing, other than reliance on examiners' expertise, whether the A level standard had been carried forward.

None of the issues raised by the introduction of a new qualification was new, much was at stake in 2002 and several reputations were on the line: of the government which had involved itself in the detail of the system; of the regulators, anxious to ensure consistent interpretation of their code of practice and bring the new examination safely home; of the awarding boards, anxious to ensure that the awards were fair and consistent, based on evidence, and would stand up to public scrutiny. All parties were conscious of their responsibility to guarantee fair and equitable treatment of all candidates and to enhance the standing of the new examination. The outcome has been well documented: charges of foul play by the examining boards; allegations of inconsistent practices and different interpretations of the code of practice by Accountable Officers who over-rode awarders' qualitative decisions in the interests of statistical comparability; claims of interference and unwarranted pressure on the examining boards to maintain levels of awards consistent with previous years by the regulator; assertions of a lack of control by the government. High profile resignations followed the publication of the 2002 results. Public trust in the system fell to a low level. An enquiry, conducted by the former Chief Inspector, Mike Tomlinson, was set up but resulted in very few changes of grades, suggesting that examiners and Accountable Officers had, in general, exercised sound judgements. The House of Commons Education and Skills Committee conducted an investigation (2003). Comparability in its many guises had come home to roost.

The A level 'crisis' and the consequent weakening of the examination in the eyes of the public opened the way for a more fundamental appraisal of the system. Mike Tomlinson was asked, in 2003, to undertake a review of the curriculum and qualifications across the 14–19 phase of education. Many believed that the curriculum

remained narrow and insufficiently focused on the needs of industry and that it failed to meet the needs of a large number of students who either disappeared from education or left compulsory education with a low level of basic skills and few other skills. The review, published in 2004, recommended a Diploma system, more akin to the baccalaureate system favoured by many European countries and reminiscent of the School Certificate with its Group requirements – and its accompanying comparability issues.

It also recommended (as had the Norwood report of 1943) that at 16 assessments should be largely in the hands of teachers, with appropriate moderation and their standards informed by benchmark tests and materials provided by the Regulator and the awarding boards. Greater use of the Internet and online marking was envisaged to standardise and make more reliable teacher assessments. Like previous attempts to place greater reliance on the judgements of teachers on work undertaken by students during the course of study, this particular recommendation was not accepted by the government. The reluctance to introduce this fundamental change to the system demonstrated how much remained to be done to raise the standard and status of assessment carried out by teachers. Regardless of the demonstrable inherent unreliability of all forms of assessment, public confidence in the reliability of externally provided assessment remained high, as did the belief that assessments carried out by teachers could not be relied on.

Tomlinson’s central recommendation – the replacement of the existing qualifications system by the Diploma – was also not accepted, with politicians from the Prime Minister down giving their support to GCE A levels which had been so highly criticised a mere two years earlier. However, a more selective Diploma system – Specialised Diplomas in defined applied areas of learning – was agreed, to come on stream in 2008. Existing qualifications will contribute to those Diplomas, as will teachers’ assessment of students’ work carried out during the course. An overall grade for the Diploma will be determined.

Whether the grading system will command the confidence of universities, employers and the wider public will make or break the Diplomas. The acceptability of the Diplomas will depend on comparability across components, across providers and across time. Nothing changes; comparability is *the* pre-requisite of public trust and confidence in examinations and qualifications.

6 Concluding remarks

As this historical survey shows, comparability has been ever-present in the English examination system. From the outset there were multiple providers, which were expected by schools and their students to provide syllabuses and examinations of a comparable standard that would open doors to higher education and the professions. The early examination bodies, all university based, explored ways of ensuring that standards remained constant from year to year, across providers and within individual subject areas. Marker reliability in particular was the focus of research in the 19th century and the findings remain relevant to the 21st century. As the

examinations became more competitive with ever increasing numbers of students, demands for consistency and comparability resulted in the codification of the system through School and Higher School Certificates; all examining boards continued to take seriously the need for comparability, creating research departments, undertaking joint studies and commissioning research from outside bodies. The government's increasing involvement in, and control over, the curriculum and assessment added new dimensions to the need for reliability and comparability: a National Curriculum and assessment system; the use of examinations as a means of evaluating the efficacy of the system as a whole and of individual schools and colleges; a belief that fewer examining boards would facilitate comparability; a loosely coordinated system gradually giving way to one of regulation underpinned by legislation. Although for most of the 20th century the emphasis was on reliable and comparable school-based examinations, as the century drew to a close, parity and comparability between academic and vocational examinations became an urgent issue, central to the raising of standards of learning and attainment and the development of skills appropriate to the 21st century. The modernisation of the examining system to meet these challenges and make use of new technologies is underway. All the indications are that as new approaches to assessment of the knowledge, skills and qualities required for the future are developed, comparability will continue to be a key and fundamental requirement of the English examination system.

Endnotes

- 1 Grace of 19 November 1857, in Graces Book 1857–1887, Cambridge Assessment Archive, EX/UC.
- 2 Letter to *The Times* by an unnamed headmaster, annotated by G.F. Browne as May 1872, in Graces Book 1857–1887, Cambridge Assessment Archive, EX/UC 1.
- 3 Manuscript letter to *The Times* by G.F. Browne (Secretary of the University of Cambridge Local Examinations Syndicate) (undated, but around 1872) which was not sent, 'correspondence having ceased', in Graces Book 1857–1887, Cambridge Assessment Archive, EX/UC 1.
- 4 Manuscript quote from a correspondent to the University of Cambridge Local Examinations Syndicate, December 1893, in Notes on Examination Papers, p. 5, Cambridge Assessment Archive, PP/JNK 2/1.
- 5 Letter from 'Brenda' to the editor of the *Journal of Education*, 1 September 1893, in Newspaper Cuttings file, Cambridge Assessment Archive, PP/JNK 2/2.
- 6 Letter from J. Chaplin, BA, dated 4 September 1893 to the editor of the *Journal of Education*, in Newspaper Cuttings file, Cambridge Assessment Archive, PP/JNK 2/2.
- 7 Article entitled 'The Sins of the Secondary Schoolmaster – II' in the *Pall Mall Gazette*, 8 December 1894, in Newspaper Cuttings file, Cambridge Assessment Archive, PP/JNK 2/2.

- 8 Minutes of meeting of University of Cambridge Local Examinations Syndicate's General Purposes Committee, 15 March 1918, in Assistant Secretary's Minutes Book, Cambridge Assessment Archive, C/CB 1/2.
- 9 Minutes of meeting of University of Cambridge Local Examinations Syndicate's Joint Committee for Examinations, 28 October 1918, Cambridge Assessment Archive, C/JCE 2/1.
- 10 *School Examinations*, article by Mr A.L. Kneed, Headmaster, Liverpool Collegiate School, *Liverpool Echo*, 17 January 1952, JMB Cuttings Scrapbook, AQA Archive, Manchester.
- 11 Letter from Mr E.R. Wood, High School for Boys, Hereford to the *Times Educational Supplement*, 30 November 1951, JMB Cuttings Scrapbook, AQA Archive, Manchester.
- 12 Ninety-third Annual Report to the University (1951), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.
- 13 One hundred and thirteenth Annual Report to the University (1971), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.
- 14 Comparison of the Standards of Different Examining Boards, Preliminary Report on the 1953 GCE Chemistry study, JMB, Cambridge Assessment Archive, PP/TSW 3/6.
- 15 Minutes of a meeting held in Bristol, March 17th and 18th 1952. Unpublished minutes, in *Secretaries of examining boards 1948–1960*. Cambridge Assessment Archive: PP/TSW 3/5.
- 16 Comparisons of the Standards of Different Examining Boards, Preliminary Report on the 1953 GCE chemistry study, JMB, Cambridge Assessment Archive, PP/TSW 3/6.
- 17 One hundred and thirteenth Annual Report to the University (1971), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.
- 18 Fifty-third Annual Report to the Universities constituting the Joint Matriculation Board (1956), being the report for the year ended 30 September 1956, published by the JMB, AQA Archive, Manchester.
- 19 Fifty-seventh Annual Report to the Universities constituting the Joint Matriculation Board (1960), being the report for the year ended 30 September 1960, published by the JMB, AQA Archive, Manchester.
- 20 One hundred and eighteenth Annual Report to the University (1976), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.

- 21 One hundred and nineteenth Annual Report to the University (1977), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.
- 22 One hundred and nineteenth Annual Report to the University (1977), University of Cambridge Local Examinations Syndicate, Cambridge Assessment Archive.
- 23 Annual Report of the Schools Examination Syndicate for the year ending 30 September 1991 (1991), Cambridge University Reporter, Annual Reports – Lent Term 1992, Special No. 15.
- 24 ‘Making the Grade’, *Dispatches*, Channel 4, 11 October 1995.
- 25 Sir Geoffrey Holland Permanent Secretary at the Department for Education and Employment from January 1993 to January 1994 quoted in the *Times Educational Supplement*, 17 March 1995.
- 26 Blackstone announces new structure of awarding bodies, press release 150/97, 19 June 1997, Department for Education and Employment.

References

Audit Commission/Office for Standards in Education. (1993). *Unfinished business: Full-time educational courses for 16–19 year olds*. London: Her Majesty’s Stationery Office.

Backhouse, J.K. (1978). *Comparability of grading standards in science subjects at GCE A level*. Schools Council Examinations Bulletin 39. London: Evans/Methuen.

Bardell, G.S., Forrest, G.M., & Shoesmith, D.J. (1978). *Comparability in GCE: A review of the boards’ studies, 1964–1977*. Manchester: Joint Matriculation Board on behalf of the GCE Examining Boards.

Beaumont, G. (1996). *Review of 100 NVQs and SVQs: A report submitted to the Department for Education and Employment*. London: National Council for Vocational Qualifications.

Board of Education. (1943). *Curriculum and examinations in secondary schools (Norwood Report)*. London: His Majesty’s Stationery Office.

Broadfoot, P. (1979). *Assessment, schools and society*. London: Methuen.

Christie, T., & Forrest, G.M. (1980). *Standards at GCE A-level: 1963 and 1973*. Schools Council Research Studies. London: Macmillan Education.

Christie, T., & Forrest, G.M. (1981). *Defining public examination standards*. Schools Council Research Studies. London: Macmillan Education.

Cresswell, M. (Ed.). (2000). *Research studies in public examining*. Guildford: Associated Examining Board.

Crofts, J.M., & Caradog Jones, D. (1928). *Secondary school examination statistics*. London: Longmans, Green & Co.

Dearing, R. (1996). *Review of qualifications for 16–19 year olds*. London: School Curriculum and Assessment Authority.

Department for Education and Employment. (1997). *Guaranteeing standards. Consultation paper on the structure of awarding bodies*. London: Department for Education and Employment.

Department of Education and Science. (1978). *School Examinations. Report of the steering committee established to consider proposals for replacing the General Certificate of Education Ordinary-level and Certificate of Secondary Education examinations by a common system of examining (Waddell Report)*. London: Her Majesty's Stationery Office.

Department of Education and Science. (1987). *Improving the basis for awarding GCSE grades*. Paper presented at the 1st Annual Conference of the Joint Council for the GCSE. London: Department of Education and Science.

Department of Education and Science/Department of Employment/Welsh Office. (1991). *Education and training for the 21st Century*. London: Her Majesty's Stationery Office

Department of Education and Science/Welsh Office. (1988a). *National curriculum task group on assessment and testing: A report*. London: Department of Education and Science and the Welsh Office.

Department of Education and Science/Welsh Office. (1988b). *Advancing A levels. Report of a committee appointed by the Secretary of State for Education and Science and the Secretary for Wales (Higginson Report)*. London: Her Majesty's Stationery Office.

Edgeworth, F.Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society, 1.1*, quoted in Hartog, Rhodes & Burt (1936).

Elwood, J., & Comber, C. (1996). *Gender differences in examinations at 18+*. London: University of London Institute of Education.

Forrest, G.M., & Shoesmith, D.J. (1985). *A second review of GCE comparability studies*. Manchester: Joint Matriculation Board on behalf of the GCE Examining Boards.

Forrest, G.M., & Vickerman, C. (1982). *Standards in GCE: Subject pairs comparisons, 1972–1980*. Occasional Publication 39. Manchester: Joint Matriculation Board.

Gillan, D. (2003). The northern CSE boards. In Assessment and Qualifications Alliance, *Setting the standard: A century of public examining by AQA and its parent boards* (pp. 95-112). Manchester: Assessment and Qualifications Alliance.

- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Good, F.J., & Cresswell, M.J. (1988). *Grading the GCSE*. London: Secondary Examinations Council.
- Gosden, P. (1983). *The education system since 1944*. Oxford: Martin Robertson.
- Hartog, P., Rhodes, E.C., & Burt, C. (1936). *The marks of examiners*. London: Macmillan & Co.
- House of Commons, Education and Skills Committee. (2003). *A level standards*. Third report of session 2002–03. HC 153. London: The Stationery Office Limited.
- House of Commons, Expenditure Committee. (1977). *Session 1976–1977: Tenth report: Attainments of the school leaver*. London: Her Majesty's Stationery Office.
- Joseph, K. (1984). Speech by Sir Keith Joseph, Secretary of State for Education and Science, at the North of England Conference, Sheffield, Friday 6 January, 1984. *SEC annual report 1983–1984*. London: Secondary Examinations Council.
- Massey, A.J. (1994). Standards are slippery! *British Journal of Curriculum and Assessment*, 5(1), 37–8.
- Matthews, J.C. (1985). *Examinations: A commentary*. London: George Allen & Unwin.
- Ministry of Education. (1960). *Secondary school examinations other than the GCE. Report of a committee appointed by the SSEC in July 1958 (Beloe Report)*. London: Her Majesty's Stationery Office.
- Montgomery, R.J. (1965). *Examinations: An account of their evolution as administrative devices in England*. London: Longmans, Green & Co Ltd.
- Newbould, C.A. (1994). *Year on year standards*. Unpublished report, Standing Research Advisory Committee of the GCE Boards.
- Northern Examinations and Assessment Board. (1996). *Standards in public examinations, 1975 to 1995: A review of the literature*. Conducted for the School Curriculum and Assessment Authority and the Office for Standards in Education by the Northern Examinations and Assessment Board on behalf of the GCSE and GCE boards. London: School Curriculum and Assessment Authority.
- Nuttall, D.L. (1973). *Mode comparability: Comparability of standards as between modes of examining*. Sheffield: The West Yorkshire and Lindsey Regional Examining Board.
- Nuttall, D.L. (1993). Presentation at Centre for Policy Studies Conference, 21 September 1993. In R. Murphy & P. Broadfoot (Eds.), (1995), *Effective assessment and*

the improvement of education: A tribute to Desmond Nuttall (pp. 236-241). London: Falmer Press.

Office for Standards in Education. (1996a). *A review of vocational GCSE courses*. London: Office for Standards in Education.

Office for Standards in Education. (1996b). *GCE Advanced Supplementary and Advanced level examinations 1996*. London: Office for Standards in Education.

Petch, J.A. (1953). *Fifty years of examining*. London: Harrap.

Roach, J. (1971). *Public examinations in England 1850–1900*. Cambridge: Cambridge University Press.

School Curriculum and Assessment Authority/Office for Standards in Education. (1996). *Standards in public examinations 1975 to 1995: A report on English, mathematics and chemistry examinations over time*. London: School Curriculum and Assessment Authority.

Schools Council. (1971). *A common system of examining at 16+*. Schools Council Examinations Bulletin 23. London: Evans/Methuen.

Secondary School Examinations Council. (1932). *The School Certificate examination. Report of the panel of investigators, appointed by the Secondary School Examinations Council to enquire into the eight approved certificate examinations held in the summer of 1931*. London: His Majesty's Stationery Office.

Secondary School Examinations Council. (1939). *The report of the 1937 H.S.C. investigation. Report on the Northern Universities Joint Matriculation Board*. London: His Majesty's Stationery Office.

Skurnik, L.S. (1974). *Monitoring grade standards in English*. Schools Council Working Paper 49. London: Her Majesty's Stationery Office.

Tattersall, K. (1983). *Differentiated examinations: A strategy for assessment at 16+*? Schools Council Examinations Bulletin 42. London: Methuen Educational.

Whittaker, R.J., & Forrest, G.M. (1983). *Problems of the GCE Advanced level grading scheme*. Manchester: Joint Matriculation Board.

Willmott, A.S. (1977). *CSE and GCE grading standards: The 1973 comparability study*. Schools Council Research Study. London: Macmillan Education.

Willmott, A.S., & Nuttall, D.L. (1975). *The reliability of examinations at 16+*. Schools Council Publications. Basingstoke: Macmillan Education.

Appendix 1 Examining bodies offering public examinations for schools in England¹

University and GCE Boards 1800s to 1990s

UODLE	University of Oxford Delegacy of Local Examinations	1857–1995	
UCLES	University of Cambridge Local Examinations Syndicate	1858–1995	
OCSEB	Oxford and Cambridge Schools Examination Board	1873–1995	
UBSEC	University of Bristol School Examinations Council	1911–1957	
SUJB	Southern Universities Joint Board (for School Examinations)	1954–1990	successor to UBSEC; merged with UCLES 1990
OCEAC	Oxford and Cambridge Examinations and Assessment Council	1995–1998	UODLE+UCLES+OCSEB
UDMSEB	University of Durham Matriculation and School Examination Board	1858–1964	closed 1964
ULEB	University of London Extension Board	1902–1930	} Different names used by London board ²
ULMSEC	University of London Matriculation and School Examinations Council	1930–1951	
UE&SEC and SED	University of London University Entrance and School Examinations Council and School Examinations Department	1951–1984	
ULSEB	University of London School Examinations Board	1984–1991	
ULEAC	University of London Examinations and Assessment Council	1991–1996	ULSEB+LEAG
UB	University of Birmingham	1900(?)–1916	Joined JMB
JMB	(Northern Universities) Joint Matriculation Board	1903–1992	
NEAB	Northern Examinations and Assessment Board	1992–2000	JMB+NEA
AEB	Associated Examining Board	1953–1994	
AEB/SEG	Associated Examining Board/Southern Examining Group	1994–2000	AEB+SEG
CSE Boards 1960s to 1990s			
EMREB	East Midland Regional Examinations Board	1963–1993	
TWMEB	The West Midlands Examination Board	(?)–1998	
EAEB	East Anglian Examination Board	1962–1990	
MREB	Metropolitan Regional Examinations Board	1962–1979(?)	} merged into LREB 1979 (?)
MREB	Middlesex Regional Examining Board	1962–1979(?)	
LREB	London Regional Examining Board	1979(?)–1990	MREB+MREB
ALSEB	Associated Lancashire Schools Examining Board	1964–1992	
NREB	North Regional Examinations Board	1964–1992	
NWREB	North West Regional Examinations Board	1964–1992	
TWYLREB	The West Yorkshire and Lindsey Regional Examining Board	1964–1982	} merged into YHREB 1982
YREB	Yorkshire Regional Examinations Board	1964–1982	
YHREB	Yorkshire and Humberside Regional Examinations Board	1982–1992	TWYLREB+YREB
SEREB	South East Regional Examinations Board	1965–1985	
SREB	Southern Regional Examinations Board	1965–1985	merged into OSEB 1985
SWEB	South Western Examinations Board	1965–1987	
OSEB	Oxford Schools Examinations Board	1986–1995	UODLE+SREB

GCSE Groups 1980s to 1990s

MEG	Midland Examining Group	1985–1998	UCLES+OCSEB+SUIB+TWMEB+EMREB
LEAG	London and East Anglian Group	1987–1991	ULSEB+EAEB+LREB (aka University of London School Examinations Council)
ULEAC	University of London Examinations and Assessment Council	1991–1996	ULSEB+LEAG
SEG	Southern Examining Group	1987–1994	AEB+SEREBS+SWEB+OSEB
AEB/SEG	Associated Examining Board/Southern Examining Group	1994–2000	AEB+SEG
NEA	Northern Examining Association	1985–1992	JMB+ALSEB+NREB+NWREB+YHREB
NEAB	Northern Examinations and Assessment Board	1992–2000	JMB+NEA
Unitary Awarding Bodies 1990s onwards			
OCR	Oxford Cambridge and RSA (Examination Board)	1998	OCEAC+MEG+Royal Society of Arts Examinations Board *
Edexcel	Edexcel	1996	ULEAC+Busines and Technology Education Council *
AQA	Assessment and Qualifications Alliance	2000	AEB/SEB+NEAB with City and Guilds of London Institute *

* Unitary awarding bodies had to offer vocational provision as well as school examinations

Blue – OCR predecessor bodies

Red – Edexcel predecessor bodies

Green – AQA predecessor bodies

(?) – denotes uncertainty. Setting up, closing down and merging examining boards was not a clear cut process. Different sources give different dates which may denote first board meeting, gaining official recognition, first examinations, change of name, agreement to merge or close, last examinations, etc. At different times a body might use one name for trading and another for legal purposes, and might trade in its own right and as part of a group. Changes of name and changes of status are not necessarily related. Confederations may represent mergers or agreements to work together with members maintaining legal independence. Boards appear twice if they offered both GCSE and GCE. By 2000 the complex history of school examining boards in England had resulted in three unitary awarding boards which also made vocational provision.

Endnotes

1 This appendix was prepared by Helen Patrick.

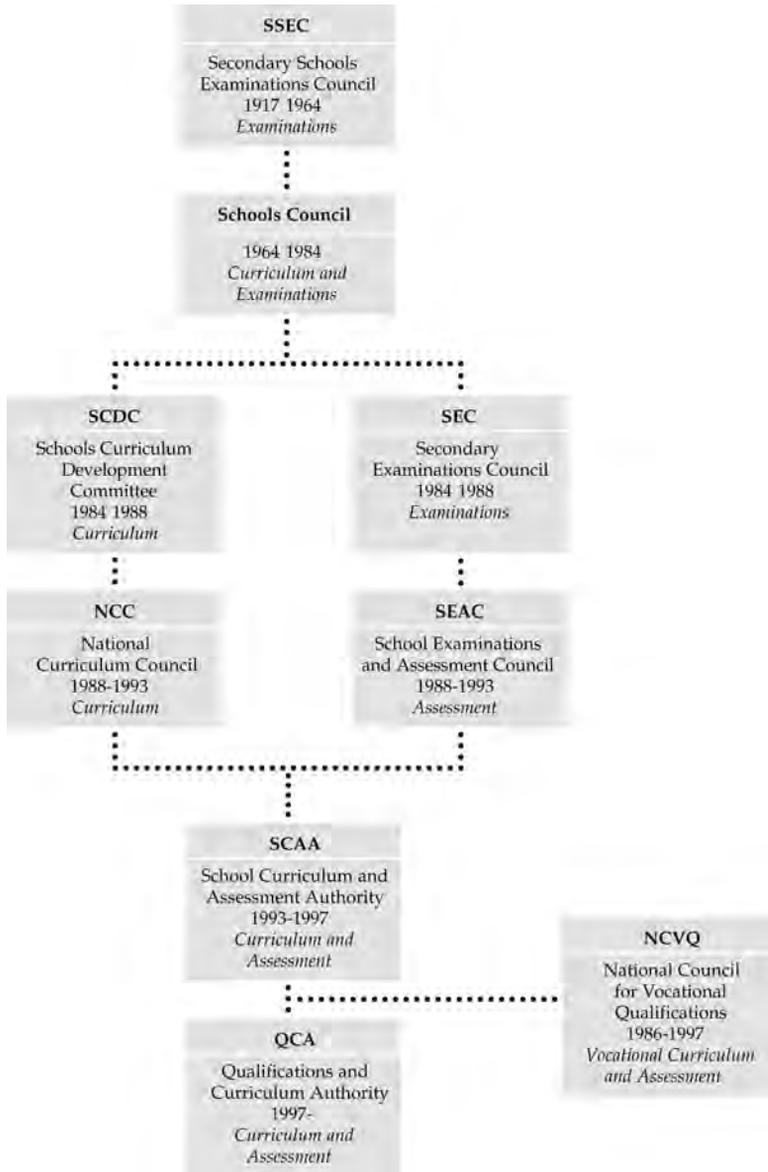
2 The University of London Extension Board (1902) marked the start of a closer relationship with schools than had previously been the case for the London Matriculation examinations which had been held since 1836.

Appendix 2 Geographical distribution of the CSE boards



Reproduced from AQA's *Setting the Standard*, 2003 (AQA material is reproduced by permission of the Assessment and Qualifications Alliance.)

Appendix 3 Advisory and regulatory bodies in England



Reproduced from AQA's *Setting the Standard*, 2003 (AQA material is reproduced by permission of the Assessment and Qualifications Alliance.)

COMMENTARY ON CHAPTER 2

Mike Kingdon

The chapter maps the structural changes in the systems of UK public qualifications since the creation of the first public examination – the University of London Matriculation Examination – in 1838. It is also the story of how, over a century, the succession of regulatory bodies has sought to promote and control the delivery of, first, school examinations and then all qualifications.

The chapter identifies how the traditional British concern about the fairness of examinations came to be expressed through the discussion of different forms of comparability. Concerns about the comparability of the 16+ and 18+ subject examinations offered by different awarding bodies have dominated for over a century, but it has not always been so. From the early 1840s the University of London monitored the pass rates for its Matriculation examination, so the initial focus was on comparability over time¹.

The seeds of today's comparability issues were sown in 1857 when London issued new Matriculation syllabuses, with some choices of subjects, and the University of Oxford introduced its Local (school) Examinations. A year later the University of Cambridge Local Examinations began. The first systematic investigation of any aspect of comparability followed the first examination in July 1859² for the new London Matriculation syllabuses. For the first time ever, the pass rate for the new examination was lower than previous years. The issue was referred to the University's Committee on Examinations for investigation and systems were put in place to monitor the characteristics of the candidature in future examinations. As the author of the chapter has illustrated, issues about the comparability of the London, Oxford and Cambridge examinations were not long in following.

In the 19th century pass/fail decisions were based on the aggregated raw marks for all subjects that a candidate had taken. However, from 1857 candidates were also expected to achieve minimum performances in key subjects. One consequence was that candidates who had passed overall could be 'referred' – invited to retake – one or more of their key subjects. The later School Certificate model, which required success in prescribed combinations of subjects for an overall award, drew attention to inter-subject comparabilities – especially those involving English Language, where standards were generally perceived to be poor (see Bruce's first law of examining in Bruce, 1969, pp. 3–4). Today, students' overall success in the GCSE and A/AS level subject examinations is once again recognised by the aggregation of results. Grade-based points are used as the basis of school performance tables and university selection systems. Therefore, to paraphrase Robert Wood (1976), 'your French has to

equal my chemistry,' and inter-subject comparability is, for many, more important than the inter-awarding body variety.

The author is correct when she identifies the search for comparability – specifically inter-awarding body comparability – as a product of the current UK system and its history. Indeed, all national systems of qualifications face issues of comparability, although the mix of types differs. In the UK case, the reverse is also true. The search for comparability has generated many of our contemporary understandings of the current system, its implementation and structure. The downside of over-emphasising comparability, at the expense of the more general concept of fairness, has been that issues such as validity and special needs have not always received due attention.

In stressing events and structures the chapter can be criticised for obscuring the human achievements that underpin them. It is also the story of how ever-wider access to recognised qualifications has been granted to generations of learners. Further, no history of the structure and regulation of UK qualifications can be complete without reference to the generations of examiners, markers, administrators and later researchers who, through their day-to-day work, created and applied the principles of modern qualifications delivery. Their attempts to make fair and equitable (comparable) decisions as new situations arose generated the practical case law and principles on which our current regulatory codes are based.

My concern is not with the content of the chapter but with what might be inferred from it. Readers may assume that ever-tighter regulation will have produced increasing levels of comparability but this argument has yet to be demonstrated. Similarly, to resolve the issue of comparability by amalgamating all of the awarding bodies into one, would be to destroy other positive features of the current system such as teacher choice and the efficiencies that follow from today's fierce competition between the unitary awarding bodies.

Readers may also assume from the chapter that the three universities were in conflict over standards and methods of examining. Instead the University of London Senate Minutes record the cooperation between the three universities through their membership of the Joint Board for the Promotion of University Teaching³. Premises were shared, examination timetables aligned and responses to Ministers coordinated. This cooperation continued throughout the 19th and 20th centuries via meetings of the Secretaries of the School Examination Boards, and constituted a forum for self-regulation.

Finally, what the chapter has not considered is how Cambridge University and London University (later Edexcel) came to predominate among the English school examination awarding bodies in the provision of examination services to countries, institutions and individual students in the Commonwealth and beyond. The school examinations and other qualification systems of almost all Commonwealth countries began when they 'budded' from the English system at some point in the 19th century. In doing so they imported the contemporary comparability issues, ideas about fairness and qualifications management systems that underpinned their selected

model. How these issues evolved, as countries developed their qualification systems to meet local needs, is probably a book in itself.

Cambridge and Edexcel continue to provide examination services to the Commonwealth and beyond, especially to smaller countries that do not have sufficient resources to fulfil all assessment functions themselves. These examinations also provide alternatives and supplements to national systems of qualifications for individual students throughout the world.

Endnotes

- 1 See statistical appendices at the back of the bound copies of the University of London Senate Minutes from 1842 onwards.
- 2 University of London, Senate Minutes for 1860, *Report of Examiners on the July 1959 Matriculation Examination* – referred to Committee on Examinations for investigation, statistical tables pages 56–58.
- 3 University of London, Senate Minutes for 1876, *Receipt of letter from the London Society for the Promotion of University Teaching inviting the University to appoint representative to a Joint Board of the Universities of Cambridge, Oxford and London*, Minute 121, page 50, representatives appointed, Minute 122. The Minutes of the Board were received and discussed at meetings of the Senate throughout the 19th century.

References

Bruce, G. (1969). *Secondary school examinations: Facts and commentary*. London: Pergamon Press.

Wood, R. (1976, July 30). Your chemistry equals my French. *The Times Educational Supplement*.

RESPONSE TO COMMENTARY ON CHAPTER 2

Kathleen Tattersall

I am grateful to Mike Kingdon for highlighting events in the story of comparability to which the chapter did not make reference. In particular, his reminder that, over the years, thousands of individuals have been involved in the administration, setting and marking of examinations is welcome. It is unfortunate if the text did not make clear that improvements in the design and comparability of assessments owed everything to individual Secretaries to examining boards, chief examiners and researchers, and I am glad to have the chance to rectify this lack of clarity. However, a single chapter cannot do justice to the history of a topic which lies at the heart of assessment and so, in effect, is the history of the examination system itself; the story demands a book in its own right.

Kingdon is also correct in pointing to the influence of the English examination system on students in the Commonwealth and beyond. However, this dimension falls outside the scope of a chapter whose focus is on the history of comparability in the English system. For the same reason the systems of other countries of the United Kingdom received little mention.

Like Kingdon I believe that factors other than comparability are of importance: fairness, validity, equity and, I would add, the involvement of teachers. As someone who started her career in a CSE board I believe strongly that teachers should play a key role not only as examiners and moderators but as assessors of their own students – a huge challenge to comparability but one worth the effort if assessment is to be fit for purpose and part of the teaching/learning dynamic. A single examining board in the interests of comparability, as argued by some, would do nothing to promote those other key values of the examining system.

The chapter covers the period of ever-increasing regulation in the interests of comparability. However, I would be alarmed if the chapter were to be read as support for a further tightening of the regulation noose. Indeed, I would hope that the juxtaposition of tighter regulation and the 2002 crisis of public confidence in A levels would raise questions about the role and purpose of regulation. From the standpoint of the author – a player in the events of 2002 – it seemed that the regulator was, at least in the first instance, unable to safeguard the system from its critics. What was at stake was public confidence in the first awards of the new A level system. An enquiry by the regulator into the grading of one of the three unitary bodies failed to satisfy the critics. A more wide-ranging enquiry was set up, not under the auspices of the regulator but chaired by an independent senior figure, Sir

Mike Tomlinson, recently retired Head of the Office for Standards in Education (Ofsted). That enquiry, with which both QCA and the awarding bodies cooperated, resulted in a more open and transparent awarding process

Sadly, a debate about regulation, like other points which Kingdon makes, falls outside the remit of this chapter – but I am glad to have the excuse to express a personal view on the issue.