

Can vocational A levels be meaningfully compared with other qualifications?

Jackie Greatorex

A paper presented at the British Educational Research Association Conference, University of Leeds, UK, 13-15 September 2001.

Disclaimer

The opinions expressed in this paper are those of the author and are not to be taken as the opinions of the University of Cambridge Local Examinations Syndicate (UCLES) or any of its subsidiaries.

Note

This research is based on data analysed by the Research and Evaluation Division of the University of Cambridge Local Examinations Syndicate for Oxford Cambridge and RSA Examinations.

Contact details

Jackie Greatorex, RED, UCLES, 1 Hills Road, Cambridge, CB1 2EU. greatorex.j@ucles.org.uk.

Can vocational A levels be meaningfully compared with other qualifications?

1 Abstract

Curriculum 2000 represents a time of change for the post 16 sector with, for example, the introduction of Vocational A levels and Advanced Subsidiary Awards. Changes are also reflected in the assessment models that are used for different qualifications, such as, GNVQs becoming more like A levels. These changes to qualifications and assessment models coincide with an expected increase in the number of students involved in education and training post 16 which might lead to a change in the nature of the post 16 cohort. The significant changes in both the assessment systems and the cohort make it very difficult to compare the standards associated with different qualifications. This paper focuses upon vocational A levels. The differences between the assessment models used in vocational A levels and other Curriculum 2000 advanced qualifications will be explored. The limitations of making comparisons between the new vocational A levels and other qualifications will be addressed.

2 Introduction

The aim of this paper is to explore comparability issues which might arise from the changes in the post 16 Advanced Curriculum (Curriculum 2000). In particular this paper focuses upon the Advanced Vocational Certificate of Education (the vocational A level).

GCE A levels were introduced in 1951 for a small elite group of students preparing for single honours study at universities (Young and Leney, 1997). By 2000 nearly 50% of students under 21 years of age had achieved an A level qualification. The government's target is that by 2003 this would grow to 60% (DfEE, 1999). Hodgson and Spours (2000) describe these as dramatic changes and add that: *Advanced level provision now includes a significant proportion of vocational study alongside more traditional academic subjects due to the impact of Advanced GNVQs* (Hodgson and Spours, 2000, 2). A level students in the English, Welsh and Northern Irish systems have a very narrow and specialised educational experience compared with students at the same level in other European countries. Today's global market demands that learners demonstrate a broad range of skills in work and everyday life but, like many other curricula, the Advanced curriculum has been characterised by an academic - vocational divide. Vocational qualifications have been devalued and there has been little mixing of academic and vocational study (Hodgson and Spours, 2000).

The Dearing committee's (1996) review of 16-19 education outlined proposals for reform. Hodgson and Spours (2000) argued that the Dearing Report was the main influence on the present government's approach to this education sector. For example, it recommended the new 6 and 3 unit GNVQs (now called VCEs), which were adopted in *Qualifying for Success*, (QfS) a DfEE consultation paper on the future of post-16 qualifications. The present government's plans for the 16 to 19 qualifications came to fruition from September 2000 when Curriculum 2000 was launched. In summary the changes to the curriculum were that new 3 unit GCE Advanced Subsidiary (AS) awards were introduced as the first part of an A level. Students can then take a 3 unit A2 to transform their AS to an A level. The AS awards aim to be: *...assessed and graded to match the levels of attainment expected from students halfway through an advanced course of study....focus on the skills, knowledge and understanding developed during the first half of an advanced course of study...The conceptually less demanding material will tend to be assessed in AS units. The more demanding A2 units will be assessed and graded at higher standard.* (Stobart, 2000, 17). The AS standard is meant to be below A level but above GCSE. The full A level should be of equivalent standard to the old A levels.

GNVQs have also changed. The 12 unit Advanced level GNVQ became the Advanced Vocational Certificate of Education (Double Award). A new vocational qualification the Advanced Vocational Certificate of Education (AVCE) or Vocational A level has been introduced. The 6 unit and the 3 unit Advanced Subsidiary Vocational Certificate of Education (ASVCE) are equivalent in size to A level and AS respectively. ASVCE is available in Business, Engineering, Health & Social Care and Information & Communications Technology. These Advanced vocational qualifications are deemed to be of equivalent standard to A levels and as a group of qualifications they are known as VCEs. Recently the Qualifications and Curriculum Authority (QCA) launched a National Qualifications Framework. This includes three 'families' of qualifications: -

- General qualifications (A levels);
- Vocationally-related qualifications (Vocational Certificate of Education);
- Occupational qualifications (Level 3 NVQ).

NVQs have been incorporated into the National Qualifications Framework which also includes, for example, A levels, GNVQs and GCSEs (see www.qca.org.uk).

VCEs and GCEs now have the same grades available, which Stobart (2000) has argued will make comparability at the unit level easier. This suggests that although VCEs and GCEs are not in the same family of qualifications there will be calls to compare them. Also they are used together for vocational and educational selection decisions thus making it necessary to know how the qualifications compare with one another. There might also be calls to compare VCE and concurrent GCE qualifications at unit level because: *the most strongly held (concern), which is already part of an existing ministerial remit to the regulatory authorities, has been the lack of alignment in the standard of units in GCE and VCE* (QCA, 2001b, 6).

In the following sections, standards and comparability are defined and discussed along with the curriculum and assessment changes that might affect comparability studies. These definitions are considered in detail in order to see whether they would affect the comparison of concurrent qualifications or comparisons over time. Methods of comparing VCE with other qualifications and their limitations are also discussed.

3 Comparability

Candidates who are entered for examinations are graded against standards of performance. Baird et al. (2000) discuss what actually constitute these standards, suggesting that they might be socially constructed. Cresswell (2000a) argues that standards are socially constructed and that they are therefore subjective, a form of social practice. Wiliam (1996) argues that we should trust Senior Examiners to make expert judgements about standards.

If similar qualifications are to be compared, such as two Business A levels from different Awarding Bodies with the same examination system, then they are said to be comparable if candidates who demonstrate the same level of attainment obtain the same grade (Bell, 2000a). There are other looser definitions of comparability where the objective might be to identify the broad equivalence of qualifications and / or to describe the similarities and differences between them. Another approach is to consider the fitness for purpose of two qualifications e.g. Coles and Matthews (1995). The question here might be how good are the qualifications as a preparation for employment and further study?

Bell (2000a) explains that there are two ways of tackling comparability studies:-

- *Mode A* where the objective is to investigate whether two qualifications are equivalent;
- *Mode B* where the objective is to describe the similarities and differences between the qualifications. The qualifications are not necessarily considered to be equivalent.

There are 5 generic approaches to the mode A comparability problem:-

- 1) Using measures of prior outcomes;
- 2) Using measures of concurrent outcomes;
- 3) Using measures of subsequent outcomes;
- 4) Comparing performance of candidates who have attempted both qualifications *at the same time*;
- 5) Expert judgement of the qualifications.

Some of the factors which affect comparability studies are:-

- The time lapse between qualifications - applying the same standard to assessments becomes more meaningless the further the assessments are apart in time (Newton, 1997b). It is most reasonable to compare a particular examination with its immediate predecessor (Bell et al., 1998; Bramley et al., 1998);
- Differences in the assessment models or changes in the structure of qualifications. For example, Taverner and Wright (1997) compared modular and linear A levels by comparing the mean GCSE and A level grades for modular and non-modular groups;
- The availability of appropriate measures of achievement (e.g. prior / concurrent / subsequent achievement). Thus mean GCSE has been used as a baseline from which to compare the relative performance of candidates in the old A level and the new AS qualifications (Elliott et al. 2001). But mean GCSE is not the best predictor of success for individual A level subjects (Bell, 2000b);
- Whether the qualifications are deemed to be of the same standard. For example, can AS and AVCE be compared given they are a different standard?
- Size and demand of qualifications. For example, is an ASVCE comparable to an AVCE Double Award given that the later demands a sustained effort over 2 years and is 4 times the size of the former?
- The content of qualifications, what is being taught and tested. That is, are Mathematics A level and VCE Art and Design able to be compared?
- The purpose of the qualifications. Some authors believe that if qualifications are for different purposes, as are GNVQ and A level, then it is meaningless and pointless to compare them (Wolf, 1995). But there are also calls to make such comparisons to ensure that neither A levels or GNVQs are an easy route to further study;
- Bell and Greatorex (2000) considered how the size of qualifications might affect comparability;
- Hodgson and Spours (2000) specify how study programs might affect attainment;
- Taverner and Wright (1997) and McClune (2001) have considered how the number of resits taken might affect attainment.

4 Similarities and differences between VCE and other Advanced qualifications

In this section the objective is to identify similarities and differences between VCE and other Advanced qualifications. There are a number of issues that could be considered:-

- *Curriculum changes* - purposes, standards, unitisation factors (i.e. study programs, size of qualifications, choices between units), Key Skills, pedagogy and recommended prior learning;
- *Assessment changes* - internal and external assessment, grading issues (i.e. the grades available, Awarding Meetings, compensation and criterion referencing, uniform mark schemes), the assessment specification (i.e. Key Skills, Quality of Language, Synoptic Assessment) and resits.

It is beyond the scope of this paper to cover all these issues in detail therefore only, some examples will be discussed here. The issues covered have been selected as they are of particular concern to Awarding Bodies and are unlikely to be considered in other literature. For a full discussion of the differences between VCEs, A levels and GNVQs see Stobart (2000).

4.1 Assessment changes

4.1.1 Internal and external assessment

GNVQs were designed as part of the NVQ framework and so external assessment (by employees of an Awarding Body) was seen to be against the competency approach which was the basis of the qualifications (Hodgson and Spours, 1997). Later, external tests were introduced to provide confirmation of knowledge and understanding for each mandatory vocational unit (Hayward, 1995). In the later years of GNVQs: *Most mandatory units are assessed by means of externally set tests as well as internal assessment of coursework. Optional units and additional units are not externally tested and so all assessment is carried out internally by the centre* (OCR, 2000i, 10). Internal assessment tends to be assignment, project and portfolio work which is assessed by the candidates' tutor in the centre. It is referred to here as centre assessment and internal assessment. Generally there were two forms of assessment in each unit (Stobart, 2000) one of which might be an external assessment.

By contrast in VCE the type of assessment is not necessarily attached to whether the unit is optional or compulsory. For example, OCR AVCE Travel and Tourism constitutes 6 units, 5 mandatory and one optional, of which two of the mandatory units are externally assessed and the rest are centre assessed (OCR, 2000c). There also tends to be one type of assessment per unit (Stobart, 2000). Reading specifications for A levels and AS suggests that the majority have one form of assessment per unit, but there are some where there is, for example, a choice between coursework and another form of assessment, usually a written examination. Prior to Curriculum 2000 linear A levels relied heavily on terminal examinations. Modular A levels had examinations at the end of each module, meaning that learning and assessment were very condensed in comparison to VCEs and other A levels.

Prior to 1994 the GNVQ tests were all multiple choice (Hayward, 1995). The external test normally lasted for 1 hour and contained 30-40 items (Walton, 1999). Stobart (2000) reports that this situation in GNVQs did not change radically over time. VCE is assessed using a variety of different types of assessment. Some units are assessed by collecting a portfolio of evidence; Art & Design VCE, for example, employ written work, artwork and written papers, while Travel & Tourism employ unseen case studies, a case study and written paper and short scenarios, and Science uses written papers with short compulsory questions and some which require extended answers (OCR, 2000a, 2000c, 2000d). Similarly, the external assessments for A levels pre- and post-Curriculum 2000 used a variety of approaches to assessment; for example, English Literature employed essay based examination papers which needed a banded mark scheme, while Mathematics used a points mark scheme. Not only were the different subjects testing different subject knowledge they were also testing different skills.

The pass marks for the GNVQ external assessments were high in the order of 70%. The tests were Pass / Fail but they did not contribute to decisions about the final grade of Merit and Distinction (Walton, 1999) and operated as a hurdle that candidates had to pass to gain the qualification. For VCE and A levels pre- and post-Curriculum 2000 the pass mark is unlikely to be as high as for GNVQs. Candidates do not need to pass every external assessment to gain a Pass and the marks that are gained on each internal and external assessment all contribute to the qualification grade. These are the important differences between VCE and GNVQ.

GNVQ assessment was primarily based around portfolio assessment (centre assessment) (Hayward, 1995). In general, 66% of VCE assessment is portfolio assessment and 33% is external assessment (QCA, 2000b; Stobart, 2000), but there are some subjects which do not conform to this rule of thumb, such as OCR Art and Design AVCE where either 66% or 83% of the qualification is centre assessed, depending on the options taken by a candidate. There is generally more external assessment in VCEs than GNVQs (Hodgson and Spours, 2000). The balance between internal and external assessment was that the A level was predominantly externally assessed using terminal examinations. Since 1991 A level modular syllabuses were allowed a maximum of 20%

coursework and a minimum of 30% assessment through terminal examination. They were also subject to end of module tests (Hodgson and Spours, 1997). Traditionally A levels relied upon external assessment while now in Curriculum 2000 *GCE AS and A levels normally contain a proportion of coursework up to 30 per cent, (though some practical or creative subjects have more)*. (QCA, 2000b, 1). These figures illustrate that VCE is almost the reverse of new A levels in the balance between internal and external assessment. This means that there is a big difference between the kind of assessment and therefore the different skills that candidates must exhibit to gain the same grade on an A level and on a VCE.

In GNVQs, centre assessment followed the NVQ approach to assessment: internal assessors assessed portfolio work. If there is more than one assessor in a centre then there is an Internal Verifier who is responsible for maintaining the national assessment standards at the centre. In VCEs there is also a great deal of portfolio work which is assessed by teachers or tutors in centres. For A levels pre- and post-Curriculum 2000 coursework is / was assessed by teachers / tutors in centres. If there is more than one assessor at a centre it is required that the assessors standardise their judgement for the NVQ, VCE and / or GCE.

The NVQ approach to maintaining national assessment standards through consistency of judgement is the verification system. This approach was also used for GNVQs. The Awarding Bodies employ External Verifiers to visit centres to ensure that there is consistency of judgement: *Assessment is carried out internally by assessors and monitored internally by Internal Verifiers. External Verifiers (EV), appointed by OCR, monitor the quality assurance process and ensure that the national standards are being maintained.* (OCR, 2000i, 10). This system has been described as a paper exercise of checking to ensure that the right paperwork is in order rather than checking the judgements of the assessors (Eraut et al., 1996). The change to VCE has also brought a new system of checking centre assessment - the moderation system - that is used for A level coursework. This can involve postal moderation rather than the centre being visited by an External Verifier. For A level coursework centres are asked to ensure that the rank ordering of their coursework is correct (OCR, 2000f, 2000g). Although there does not appear to be a similar instruction for VCE. Another difference is that moderation is much more important for the fairness and reliability of VCE than for A levels as there is more centre assessment in the VCE.

In all of the assessment systems the internal assessor was involved in teaching and assessing the candidate's work. However the amount of influence that the teacher or internal assessor has over the candidate's grade through their involvement in national assessment is higher for VCE and GNVQ than for A levels.

The number of assessments per unit is linked to comparing GNVQ and VCE over time. If VCE grades and GNVQ were compared, the different approaches to gaining a pass might affect the grade distribution and comparability studies. Expert judges might be able to judge the quality of work if the type of assessment is different. The amount of internal and external assessment is an issue for comparing GNVQ and VCE over time and VCE and GCE as concurrent qualifications. The different types of assessment that have been used are relevant to both comparisons over time (between GNVQ and VCE or between VCE and old GCE) and between concurrent qualifications (VCE and GCE). Therefore the difference between the types of assessment means that we are not comparing like with like.

4.2 Grading issues

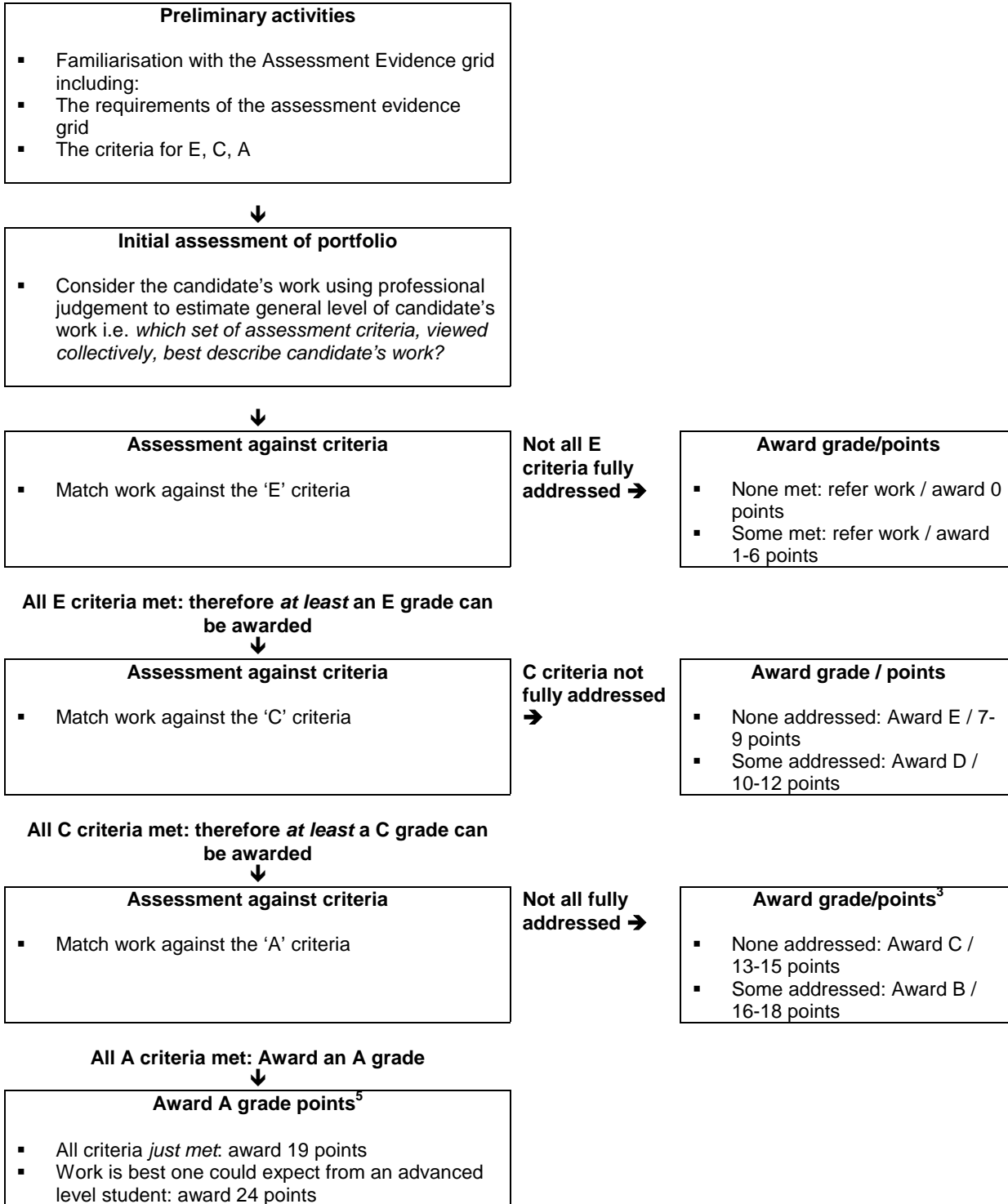
4.2.1 The grades available

GNVQ candidates were awarded grades of Pass, Merit or Distinction (Hayward, 1995). *The broad equivalence between the new grades for Vocational A level / Advanced Subsidiary and the old grades for Advanced GNVQ is as follows: A-Distinction, C-Merit, E-Pass* (Stobart, 2000, 27). Both VCE and A level in Curriculum 2000 will have grades A to E available. *Three-unit and six-unit VCE qualifications are graded on a scale of A-E. Twelve unit VCE qualifications are graded on a double award scale i.e. AA, AB, BB, BC etc.... Candidates not achieving EE or E will receive U (unclassified)* (QCA, 2000a, 50). There is no grade N. In the old A levels grades A to E and U were available, but candidates who just missed achieving grade E were sometimes given grade N (near miss). It has been claimed that *For vocational A level both the unit grades and the overall grades will be on the same A-E scale as Advanced GCE. This will make it easier to compare students' performance across different qualifications.* (Stobart, 2000, 27).

For GNVQ students to gain a Pass grade, students had to pass all the mandatory vocational and optional vocational units and Key Skills units and pass any externally assessed tests. This constituted a hurdle which is a notion derived from criterion referencing and the competency model in the sense that to get past the hurdle a series of criteria must be met. For a Merit or Distinction the student's work for the mandatory vocational and optional vocational units was graded against grading criteria. Key Skills and additional units were not graded. The four grading themes were planning, information seeking and handling, evaluation and quality of outcomes (OCR, 2000i). To achieve an overall Merit (Distinction) grade a candidates' portfolio had to contain a third or more of evidence which met all the Merit (Distinction) criteria for each theme. There is no such hurdle to gain a Pass or E and above in the VCE or A level qualifications at the **award level**. Additionally the VCE and A level approach to *marking* for externally assessed units follows the same way of using the mark scheme for all grades. But for GNVQ the work was *marked* in a different way for different grades. It was marked twice, once to make sure that all criteria were met and therefore a Pass has been achieved; then a second time, in a different manner – taking the best third of the work and identifying whether it matched the criteria for Merit or Distinction using a principle of best fit to apply the criteria. For VCE work at the unit level the process in diagram 1 is used.

Diagram 1 ASSESSING PORTFOLIOS AGAINST THE NEW GRADING ARRANGEMENTS FOR VCE (formerly Advanced GNVQ)

SUGGESTED PROCEDURE FOR CENTRES*



Another difference between VCE and GNVQ is that the GNVQ internal assessment was 'element based' (elements are statements of what a candidate must be able to do e.g. *promote communication within groups*) through evidence indicators (performance criteria and range) (Stobart, 2000). A number of performance criteria were contained within each element. They specified the standard of performance which students had to meet. They had to meet all the performance criteria. The range was also included within each element, indicating the dimensions which must be covered for successful completion of the element. The range was an indication of the breadth of the element. VCE assessment is unit based through criteria within an assessment grid (which specifies the requirements for grades E, C and A).

The changes in GNVQ, VCE and GCE grading arrangements will affect grading standards, perhaps by changing what is required. In terms of losing the GNVQ hurdle this could be seen to make VCE more accessible than GNVQ. The changes in VCE, GNVQ and GCE grading arrangements are linked to comparing qualifications over time e.g. GNVQ with VCE and VCE with old GCE.

Imposing different grading arrangements and grade ranges on qualifications might be reflected in attainment and in statistical comparisons of the qualification. It is unclear whether different grading procedures and grades available would influence the task of comparing the quality of work.

4.2.2 Award Meetings

The external tests for GNVQs did not contribute to the final grades. For VCE externally assessed units grades E and A are awarded by an Awarding Committee using professional judgement and statistical evidence. Other grade boundaries are determined arithmetically (QCA, 2000a). For A levels prior to Curriculum 2000, grade boundaries on each paper were determined in a similar way in the order E, A then B. The remaining boundaries were calculated arithmetically (QCA, 2000d). For A levels in Curriculum 2000 only the E and A unit boundaries are determined judgementally. Other grades are determined arithmetically (QCA, 2000a). Changing GCE awarding practice as described above might affect the grade distribution and therefore statistical comparisons between GCE over time and other post 16 qualifications over time.

Uniform Mark Schemes (UMS) are used because there are a range of choices that candidates can take in unitised schemes and once the boundaries are aggregated from all the different combinations there might be a variety of award level grade boundaries. To avoid this the raw marks are converted to a common scale or Uniform Mark Scheme. For details of how UMS are used in VCEs and other qualifications see Stobart (2000) and QCA (2001c). For GNVQs no Uniform Mark Schemes (UMS) were used. For VCE the grade boundaries for centre assessed units are determined using a nationally agreed UMS (QCA, 2000a). In the case of modular A levels prior to Curriculum 2000 a Uniform Mark Scheme was used to determine grade boundaries. For linear A levels a UMS was not always required. UMS is used for AS and A2. Where it is used UMS will affect the grade distributions of the qualification. Precisely how this will affect statistical comparisons is difficult to tell.

For GNVQs the marks that the candidates gain on the tests do not contribute to the final grade that the candidate receives. The way that the grade is determined is given above. But for VCEs the final qualification level grades are determined by using UMS. Likewise indicator 1 (an aggregation of the lowest boundary marks from each component or unit) was used for modular A levels but both indicators 1 and 2 were required for linear A levels to find the qualification level boundaries. Indicator 2 is a weighted average of the cumulative percentages of candidates achieving each component / unit boundary. Indicator 2 compensates for regression to the mean when the syllabus level boundaries are calculated from the unit boundaries. In other words when the unit boundaries have been chosen and they are aggregated using indicator 1 the syllabus level distribution bunches around the mean. Consequently there are fewer people gaining the extreme grades of A and E at the syllabus level than there are at the unit level when indicator 1 is used. A levels in Curriculum 2000 use the same procedure as is

used for VCE, that is UMS and indicator 1 which does not compensate for regression to the mean. For VCEs and A levels in Curriculum 2000 the compensation for regression to the mean must take place in the judgements that examiners make in setting unit level boundaries. This means that unit boundaries might be harsher at grade A and more lenient at grade E for GCE and VCE than they might have been if UMS were not used. The use of the different indicators affects the position of the boundaries and likewise the grade distributions at the specification level. In turn this might influence statistical comparability of the post 16 qualifications. Whether the difference in the judgements in Awarding Meetings would affect judgements about quality of work in comparability studies is speculative.

Again the differences between the Awarding procedure are relevant to comparing VCE with GNVQs and VCE with GCE over time.

4.2.3 Compensation and Criterion referencing

Criterion referencing means using criteria against which candidates' achievements are matched to identify whether they have passed or failed. The competency model is related to criterion referencing in that it employs a series of criteria which candidates must Pass to be deemed competent. Many assessment systems combine both norm and criterion referencing, but NVQs and GNVQs tended to draw most heavily from the competence and criterion referencing assessment approaches. To pass a GNVQ unit candidates had to meet all the criteria and pass all the components of the unit, such as external tests, as well as offer portfolio evidence of competence. But VCE centre assessment does not work on either a criterion referencing or a norm referencing model. Some units are assessed by tests, marked using a principle of compensation and graded using the Awarding procedure (Awarding Meetings) for A levels. But the situation is different for the internally assessed units; the suggested procedure for grading centre assessment is given above in diagram 1.

The initial assessment of the portfolio follows the principle of best fit that is applied to Key Stage assessment. Then the assessment procedure uses the notion of a hurdle i.e. if all the E grade criteria are met then at least a grade E can be awarded. Hurdles are derived from a criterion referencing approach to assessment (see 4.2.1 above). The next hurdle is the grade C criteria which work in a similar way in that if the grade C criteria are met then at least a grade C can be awarded. If the criteria are not met then a principle of compensation applies between the hurdles; a candidate who has achieved the E criteria but not the C criteria, and has more strengths than weaknesses, can receive a grade D. The same principle applies to the grade A criteria; at the unit level VCE assessment includes compensation and hurdles and the approach to assessment (criterion referencing / compensation / hurdles) varies with the type of unit.

The extent to which criterion referencing was relevant to linear A levels at the paper level and modular A levels at the module level depended upon the subject. In English Literature and other subjects where banded mark schemes were used there was some evidence of criterion referencing. When there was a points mark scheme there would be less. Whatever the mark scheme a principle of compensation would be used (Cresswell and Houston, 1991). For A levels in Curriculum 2000 the same applies.

It is argued below that VCE and GNVQ are derived from the competence model which is meant to ensure that people are competent to do a job. Although they are vocationally related, however, they do not determine whether someone is competent to do a job. VCE and GNVQ have a selection role as they can be a progression route to Higher Education (HE). This is why there were Pass, Merit and Distinction grades in the GNVQ and there are grades A to E available for the VCE. But the existence of a grading system (other than 'competent' and 'not yet competent') is arguably against the philosophy of the competence model and so vocationally related qualifications are said to be experiencing academic drift (Hodgson and Spours, 2000). On the other hand as

semi-vocational qualifications are not a license to practice they do not need to be awarded on a pass / fail basis and can be graded.

To gain the award GNVQ candidates had to pass all units. This is because the qualification was part of the NVQ framework which utilised a competency model and to be deemed competent a candidate must be competent in all areas, there was no compensation for failing one unit. To gain a pass at the award level for a GNVQ there was no compensation allowed, all the criteria must be achieved to gain a Pass. To gain a Merit or Distinction the candidates was initially graded on each grade theme and the lowest grade from each theme were combined and the lowest grade of the combined theme grades provided the overall grade (OCR, 1995). Throughout the above process for gaining a Merit or Distinction the candidates are given the lowest grade possible for the best third of their portfolio work. This does not follow a compensation principle. The process of gaining a grade Pass, Merit or Distinction followed a criterion reference system. In contrast for VCE and new A levels a compensation principle is applied in aggregating the scores from different units. A candidate does not have to pass all the units to pass the VCE / GCE (Stobart, 2000, JCGQ, 2000a). This was also true for the old A levels. As VCE are semi-vocational rather than vocational it is not essential that the candidates are competent in all aspects of the qualification.

GNVQs did not use compensation which is a feature of VCE and GCE assessment procedures at the award level. This affects comparability between GNVQ and VCE over time. The lack of compensation in centre assessment at the E grade at the unit level affects concurrent comparisons between VCE and GCE units. In making judgements about quality of work in comparability studies experts might be able to account for the difference in approaches to grading and grades available. This cannot be controlled for in statistical comparability.

5 The Limitations of Making Comparisons between VCE and other Advanced Qualifications

The far reaching changes in both the curriculum and the assessment system in Curriculum 2000 will mean that it is difficult to identify which factor(s) might affect standards and their comparison. This also means that there could be changes in the cohort, and changes in the qualifications that make it difficult to realistically make comparisons over time. As the vocationally related and general qualifications in Curriculum 2000 are different it is also difficult to make meaningful concurrent comparisons. As explained earlier, there are three main ways of comparing qualifications by statistical comparison, expert judgement and a combination of the two. In this section some methods of making comparisons which utilise statistics and expert judgement are explored and their limitations for comparing VCEs with other qualifications are given. It should be noted that this is not an exhaustive list of methods.

5.1 Statistical Approaches

When qualifications are deemed to be comparable the comparison is made in terms of groups of candidates i.e. *statistical comparability*. Those who use a statistical definition of comparable standards must either accept that qualifications are open to continual revision (or have different features) or argue that the original standards set are preferred (or the standards of one qualification are preferred) (Cresswell, 2000a). Therefore when identifying statistical comparability between VCE and other qualifications (GNVQ and old and new A levels) it must either be accepted that:-

- in the case of making comparisons over time the qualifications GNVQs, VCEs and A levels are open to revision. In the case of comparing concurrent qualifications VCEs and A levels have different characteristics;

- or in the case of standards over time it must be stated that the standards of GNVQ are preferred to VCE or that old A level standards are preferred to VCE. This would imply that the new qualifications recently promoted by the government were of less value than previous qualifications and this is not likely to be a popular political view. When VCEs are compared with new A levels it must be stated that the standards of one qualification are preferred.

Cresswell (2000a) adds that although two grade distributions might be identical for whole groups they might not be identical for subgroups. Therefore educational researchers must defend the choice of the arbitrarily constituted group of candidates that they use to define their standards. Additionally there are systematic differences between the self-selecting groups of candidates who take different qualifications (Cresswell, 2000a). Given the differences between vocationally related qualifications and general qualifications (Stobart, 2000) they are likely to appeal to different groups of students in Curriculum 2000. Therefore, in making statistical comparisons between the qualifications such differences need to be controlled for before standards can be compared. Cresswell (2000a) explains that either an indirect control of attainment can be employed or an independent measure of attainment can be used (Cresswell, 2000a). This relates back to using prior / concurrent / subsequent measures of achievement as suggested by Bell (2000a). Cresswell (2000a) adds that choosing different ways of controlling for these variables or choosing different independent measures of attainment, leads to different operational definitions of comparability as the choices made by researchers involve value judgements (Cresswell, 2000a). This conclusion was also reached by Bell (2000a).

One method of comparing concurrent qualifications is subject pairs. Here the difference between the mean grade achieved by candidates on two qualifications with an overlapping candidature are compared. The assumption behind this approach is that the difference should be 0. This approach is limited as there is no reason why candidates should do equally well in different qualifications. If there appear to be differences over time it is difficult to tell which specification's standard has shifted. Additionally it is unclear whether the standard has changed or whether the aptitude of the cohort has changed. See Forrest and Vickerman (1982) and Newton (1997a) for a full discussion of the limitations of this approach. If this method were applied to VCEs and A levels all the assumptions behind the analysis would **not** be sustained. For example, there is no reason why candidates should do equally well in an A level and a VCE. This would affect the outcomes of subject pairs and make them difficult to interpret.

To compare modular and non-modular A level standards Taverner and Wright (1997) compared mean GCSE and mean A level grades for the modular and non-modular groups. This approach is limited as Bell (2000b) found that the mean GCSE was not the best predictor of success for individual A-level subjects. McClune (2001) found that pupils opting for modular assessment midway through the course might be at a disadvantage compared to those who opt for assessment at the end of the course. Also less mature boys who are examined at the end of a module might be advantaged. Given this evidence AS might advantage less mature boys. Cresswell's (2000a) argument that examinations are not statistically comparable when the grade distributions are different for a definable subgroup can be applied here. It means that the modular and linear syllabuses might not be statistically comparable, despite this they are deemed to be the same standard.

The problems with comparing linear and modular syllabuses can affect the comparison of VCE and linear A level syllabuses or VCE and modular syllabuses. The choices that candidates and centres make about when tests and resits are taken in terms of maturity, might affect the candidates' results and perhaps the performance of subgroups.

One of the strategies that has been suggested for comparing achievement in different qualifications is prior measures of comparability. For example, Elliott et al. (2001) use mean GCSE scores to compare achievement in two different A levels by comparing the A level unit grades of students of the same calibre as measured by mean GCSE. For such methods to be used the two A level units (or other qualifications) would need to be sufficiently

similar to make the comparison sensible. Also the most appropriate predictor of achievement on each qualification would need to be established. A limitation of this method is that it only compares part of the qualifications rather than the whole qualification. However this limitation is also a strength in the sense that this method compares the most similar parts of each qualification which is a precise and feasible method of comparison as the modules are assessing similar knowledge and skills. Other limitations are discussed in Elliott et al. (2001). If there are similar units in AS and VCEs or VCEs and modular A levels this approach could be used to make comparisons between similar units in each of the two qualifications. This would be useful for both concurrent comparisons and comparisons over time.

Jones (1997) used a multilevel model to compare the grading standards of four different concurrent A levels in Mathematics. He explains that the dangers of controlling for factors which can affect standards is that only the easily quantifiable variables are used in the adjustment and that any differences between distribution are attributed to a difference in standards. Indeed conclusions can only be drawn about grading standards when all relevant independent variables are controlled for or if assumptions are made about the consistency of the behaviour of the variables. Some variables can be easily controlled for, e.g., prior and concurrent achievement, but factors like teacher and student motivation and schools effectiveness are less quantifiable from the immediately available data (Jones, 1997). If this method were to be used with VCEs then the changes in the qualifications over time and the differences between the vocationally related qualifications and general qualifications e.g. teaching, learning and assessment of say Key Skills, or their effects cannot be easily measured and controlled. Jones (1997) explains that the effects of motivation, for example, cannot be assumed to be the same for different qualifications, given that some courses are expressly intended to improve it. His argument is substantiated by the results of his multilevel model analysis. One Mathematics specification was shown to gain higher results than the others. Some might argue that the grading standards of this qualification were too lenient. But Jones (1997) argues that the specification was designed to improve performance by enhancing motivation. Therefore it achieved its aim rather than being too lenient. There are other differences between the qualifications which Jones (1997) uses to explain why there might be a difference in the students' attainment.

Pinot de Moira (2000) used a multilevel model to compare the grading standards of GCSE English from different Awarding Bodies. In addition to including the variables used by Jones (1997) and Bell and Dexter (2000) (see below) she utilised a questionnaire which asked about factors like whether candidates were encouraged at home to study English, whether they liked it as a subject, whether the candidates were eligible for free school meals etc. These are not factors which relate to changes in qualifications but they are factors which might not otherwise be accounted for in comparisons between standards. She found that there were many factors which appeared to have a significant role in predicting that a candidate would exceed each key grade boundary besides which English specification they did or what Awarding Body they chose.

Bell and Dexter (2000) also used a multilevel model to compare examination standards. They give a discussion of the advantages and disadvantages of using:-

- particular measures of attainment as variables in the model;
- a continuous, ordinal or binary variable to represent grades;

as variables in the model. They explain that using a multilevel model has the advantage of correctly estimating the standard errors of the parameter estimates and providing an estimate of the school level variance (which is not high in the UK). Jones (1997) and later Bell and Dexter (2000) both note the importance of the use of expert judgement in considering comparability, as judges can account for changes in the specification over time or differences between qualifications which statistics cannot. This is particularly important when comparing VCE with GNVQs and old A levels as described above where the qualifications have changed significantly over time. It is also important given that there are far reaching differences between VCE and other concurrent qualifications, as detailed above. These differences cannot be accounted for by the statistics.

5.2 Expert Judgement

As explained above, one of the main disadvantages of using statistical methods is the inability to account for changes in qualifications over time or differences between different types of qualifications that are not easily measured. For example, how can the effect of the differences between teaching styles in VCE and A level be measured based upon available data? Consequently comparability studies which employ expert judgement are most suitable for comparing VCEs with other qualifications. The expert judges might be able to compensate for differences between qualifications in ways that statistics cannot. The following are two judgmental methods of comparability, the limitations of which are outlined.

Cresswell (2000a) states that expert judgements are really value judgements, they are subjective rather than objective. Paterson (2000) argues that judgements are therefore, inter-subjective rather than subjective as they are shared within a community. Judgements can be made about the relative attainment of candidates in different qualifications. Expert judgements are not amenable to empirical verification or epistemological justification but they can be the results of a rational process supported by reasons. But the assessment systems based upon supposedly more objective judgements, i.e. criterion referenced systems, have not worked successfully in the sense that success means replicating holistic value judgements of quality (Cresswell, 2000a). However, William (1996) argues that we should trust Senior Examiners to make expert judgements about standards. Given that standards are socially constructed and that they are a form of social practice we should be comfortable with using expert judgement as a way of comparing standards. Additionally as standards are socially constructed they cannot be made objective (Cresswell, 2000a) and therefore using non-objective, qualitative judgements to compare standards is a valid approach.

As standards are socially constructed, what is to be assessed and valued at different levels of attainment changes with time. Standards and their meaning are localised in time and space. Therefore when assessments separated in time are different it makes direct statistical comparisons invalid. Only indirect statistical comparisons and qualitative judgements can be made (Cresswell, 2000a). By the same token qualifications that assess different qualities but are taken at the same time cannot be compared using direct statistics - only indirect statistics and expert judgement can be used.

Jones (1997) states that a cross moderation approach to comparing standards in different qualifications is preferable to statistical approaches. He explains that one of the main reasons for this is that much of what should be adjusted for in a statistical comparison e.g. differences in design in the syllabuses cannot be easily measured. In cross moderation exercises experts systematically compare candidates' work from different syllabuses and make a judgement about which is of better quality. They also systematically compare question papers, specifications and judge the similarities and differences in demand between the specifications (Gray, 2000). The advantage of this approach over statistics is that the experts might be able to account for differences between qualifications in the judgements that they make (Jones, 1997). Additionally the experts inspect students' work. For more details see Gray (2000). The main disadvantage of cross-moderation exercises are that they are expensive (Jones, 1997).

Can qualifications deemed to be of higher and lower standards, e.g., AS taken from January 2001 and ASVCE, be compared? In the literature there is a method of comparing qualifications 'vertically' and 'horizontally'. It is called fitness for purpose. Coles and Matthews (1995) have considered using this as a method of comparing qualifications in terms of their fitness for the purpose of progression into employment or higher levels of education. The methodology was tested by comparing GCE A levels in science subjects and Advanced Science GNVQ. In this method, different qualifications are evaluated in terms of an external comparator, as their standards are too different to compare directly. Briefly, the method consists of subject specialists from HE and employers describing the level of knowledge and skills that they require in a qualification at a particular level. Qualifications are scrutinised to identify whether they meet the criteria that have been developed. The method is

flexible as it covers all qualifications – vocational and academic, including those of different sizes at all levels. In this methodology comparability is based on the perception of some of the ‘users’: subject specialists, tutors and employers. However candidates who are possibly the most important group of qualification ‘users’ are not consulted and candidates’ achievement is not used to compare the qualifications.

Coles and Matthews (1995) acknowledged that the subject specific experts were not necessarily representative of the scientific community. It was found in Science that employer and Higher Education tutors had more in common than they had different in terms of their requirements of Advanced qualifications. GCE and GNVQ met their user requirements to the same degree. GNVQ matched more general skill components than GCE A levels. It was found that the users wanted students to do basic things well rather than study a wide range of topics in depth. Again the expert judgements made in this kind of comparability study might be able to account for the differences between qualifications in a way statistics can not. Indeed the result that the different users of the qualifications were able to agree upon provided a set of criteria against which to judge the two qualifications. This suggests that expert judgement can overcome some differences between qualifications, thus the pedagogic and assessment differences between GCE and GNVQ were overcome. This would have been impossible to do with statistics. However there does need to be similarity between the content of the qualifications for experts to make judgements.

A limitation of fitness for purpose is that students’ work is not inspected. This is particularly important as standards are not in the written word, such as, assessment criteria and Key Stage level descriptors. Standards are communicated by interpreting the written word such as assessment criteria and Key Stage levels against examples of work (Wolf, 1995; Cresswell, 2000a). Therefore any comparability study which does not involve the inspection of candidates’ work is inherently flawed. As a methodology ‘fitness for purpose’ is only instrumental, in that it assumes that students are studying ‘for a purpose’, not for interest. For example, students might study at night school for interest or entertainment rather than because the qualification will be useful. In studies using this methodology the question of comparability changes from ‘are the qualifications equivalent and what are their similarities and differences’ to ‘are these qualifications equivalent for a particular purpose?’

5.3 Statistical approaches and judgmental approaches

Ideally both statistical and judgmental approaches would be combined. This would mean that the strengths of both approaches could be used to try to account for the changes that have been made in qualifications that might affect attainment. An example of a combination of statistical and judgmental approaches being combined is that the multilevel model approach used by Pinot de Moira (2000) was complemented by a cross-moderation exercise undertaken by Pritchard et al. (2000). The results of the two studies complemented one another.

6 Conclusions

It has been argued that in Curriculum 2000 there are large differences between GCEs and VCEs and that the changes to (a) GNVQs to create VCEs; and (b) old GCEs to create AS and A2; are substantial. These dramatic changes and differences make it difficult to say with any certainty what factors affect attainment based on the readily available data and therefore to what extent standards have been maintained over time (and are similar for VCE and GCE). When comparing VCEs with other qualifications many factors need to be considered. Sometimes critics only consider one aspect of the changes in their papers and omit the rest. It has been argued that direct statistical comparisons between grade distributions cannot be made and that when indirect comparisons are made the statistics cannot account for the differences between concurrent qualifications and / or how qualifications have changed over time. However expert judgement can be used to account for some of the differences between qualifications and the revisions to qualifications. Although for experts to make judgements

the qualifications must have some similarities e.g. content. As in the case of Coles and Matthews (1995) who successfully used the fitness for purpose methodology to compare GNVQ and A level Science. The conclusion surmises that when qualifications are compared it is useful to use both statistical and expert judgement approaches to utilise the strengths of both.

Can VCE be meaningfully compared with other qualifications? The point about methods of comparability made above is quite a general statement. In the case of comparing VCEs with other qualifications although expert judgement can overcome some differences between qualifications it does have its limitations. Also it is essential that the comparisons made are meaningful therefore. VCE Art and Design might be compared with A level Art and Design or VCE Science might be compared with Science A levels. These comparisons are meaningful in that the general content shows some comparison of like with like and the differences in teaching approaches and other differences like assessment approaches might be overcome using expert judgements.

Given the discussion above crossmoderation methods would be appropriate for comparing VCEs and other Advanced qualifications. This would involve inspecting candidates' work, that would need to be collected and archived. There are likely to be problems collecting GNVQ work as candidates might want it to use in their portfolios for job interviews. Also this work is now quite old given that the qualification has terminated. However making such comparisons raises the question of why two qualifications, say A level and VCE, would be used to teach the same subject, say Business, for the same purpose i.e. selection for work and further study. The answer might lie in the individual differences between candidates e.g. in their preferred work styles and aptitudes. The different qualifications might enable different types of students to achieve similar levels of attainment through varying routes. The point about comparing like with like also holds comparisons over time as well as concurrent qualifications. Therefore VCEs might be meaningfully compared with GNVQs and A levels in the same subject.

In terms of comparing VCEs with other concurrent VCEs this practice would be no less meaningful than comparing the standards of different A level subjects with one another. Comparing the grading standards of different subjects in the general qualification family is already standard practice e.g. subject pair comparisons. If the limitations of these methods can be tolerated for general qualifications then there is no reason why they cannot be tolerated for vocationally related qualifications.

It is likely that any future curriculum and assessment changes will affect comparability studies just as the changes in Curriculum 2000 have affected similar studies.

7 Bibliography

- Baird, J. (1999) *Are examination standards all in the head? Experiments with examiners' judgments of standards in A level examinations*, Paper presented at the British Psychological London Conference 1999. Institute of Education: London.
- Baird, J, Cresswell, M., Newton, P. (2000) Would the real gold standard please step forward? *Research Papers in Education Policy and Practice*, 15, 2, 213 - 229
- Bell J. F. (1997) Question Choice in English Literature Examinations, *Oxford Review of Education*, 23, 4, 447-458.
- Bell, J. F.(2000a) *Review of research undertaken comparing qualifications*. In Bell, J. F. and Greatorex, J. (2000) *A Review of Research in Levels, Profiles and Comparability*. A report to the Qualifications and Curriculum Authority.
- Bell, J. F. (2000b) *Methods of aggregating GCSE results to predict A-level performance*. Paper presented at the British Educational Research Association Annual Conference, Cardiff University, September 7-10 2000.
- Bell, J. F, Bramley, T., and Raikes, N. (1998) Investigating A level mathematics standards over time, *British Journal of Curriculum and Assessment*, 8, 2, 7 – 11.
- Bell, J. F. and Dexter, T. (2000) *Using multilevel models to assess the comparability of examinations*, Paper presented at the Fifth International Conference on Social Science Methodology, October 3 - 6.
- Bell, J. F. and Greatorex, J. (2000) *A Review of Research in Levels, Profiles and Comparability*. A report to the Qualifications and Curriculum Authority.
- Berk, R. (1995) Something Old, Something New, Something Borrowed, A Lot to Do, *Applied Measurement in Education*, 8, 1, 99 - 109.
- Bramley, T., Bell, J.F. & Pollitt, A. (1998). Assessing changes in standards over time using Thurstone paired comparisons. *Education Research and Perspectives*, 25 (2), 1 - 23.
- Busch, J. C. and Jaegar, R. M. (1990) Influence of type of judge, normative information, and discussion on standards recommended for the National Teacher Examinations. *Journal of Educational Measurement*, 27, 145 - 163.
- Business and Technology Education Council (1996) *GNVQ Assessment Your Questions Answered, Guidance for September 1996*, Business and Technology Education Council, City and Guilds, NCVQ and RSA Examinations Board: London.
- Christie, T. & Forrest, G. M. (1981). *Defining Examination Standards*. Schools Council Research Studies: London.
- Coles, M., & Matthews, A. (1995). *Fitness for Purpose: A means of comparing qualifications*. A report to Sir Ron Dearing to be considered as part of his review of 16-19 education.
- Cresswell, M. (1987). Describing examination performance: grade criteria in public examinations. *Educational Studies*, 13, 247 - 265.

- Cresswell, M. J. (1995) *Technical and Educational Implications of using Public Examinations for Selection to Higher Education: Issues and Practice* Educational Research Centre and Princeton, International Association for Educational Assessment: Dublin.
- Cresswell, M. J. (1997) *Examining Judgements: Theory and Practice of Awarding Public Examination Grades*, PhD thesis, University of London, Institute of Education: London.
- Cresswell, M. J. (2000a) The Role of Public Examinations in Defining and Monitoring Standards, in H. Goldstein and A. Heath (2000) *Educational Standards*, Oxford University Press: New York.
- Cresswell, M. (2000b) *Defining Setting and maintaining Standards in Curriculum -embedded Examinations: Judgemental and Statistical Approaches*. In H. Goldstein and T Lewis (Eds) *Assessment: Problems, developments and statistical issues*, John Wiley and Sons: Chichester.
- Cresswell M. J. and Houston J, (1991), Assessment of the National Curriculum - some fundamental considerations, *Educational Review*, 43, 1, 63-78
- Cross, L. H., Frary, R. B., Kelly, P. P., Small, R. C., and Impara, J. C. (1985). Establishing minimum standards for essays: Blind versus informed reviews. *Journal of Educational Measurement*, 22, 137-146.
- Dearing, R. (1996) *Review of Qualifications for 16-19 Year Olds*, SCAA: London.
- DES (1991) *Education and Training for the 21st century, Volume 2: The challenge to colleges*, HMSO: London.
- DfEE (1997) *Qualifying for Success: A Consultation Paper on the Future of Post-16 Qualifications*, DfEE: London.
- DfEE (1999) *National Learning Targets for England for 2002*, DfEE: London, in Hodgson, A. and Spours, K. (2000) *Qualifying for Success: Towards a framework of understanding*, the first report from the project Broadening the Advanced Level Curriculum: Institutional Responses to Qualifying for Success, Institute of Education / Nuffield Project: London.
- Elliott, G., Forster, M., Greatorex, J. and Bell, J. F. (2001), *Back to the future: a methodology for comparing old A level and new AS standards*, Paper presented at the British Educational Research Association Conference, University of Leeds, UK, 13-15 September 2001.
- Eraut, M., Steadman, S., Trill, J. & Parkes, J. (1996) *The Assessment of NVQs. Research Report Number 4*. University of Sussex: Brighton.
- Fitz-Gibbon, C. and Vincent, L. (1994) *Candidates' performance in mathematics and science*. A report for the School Curriculum and Assessment Authority. In Taverner, S. (1997) An investigation into A level entry policy of mathematics departments, *British Journal of Curriculum and Assessment*, 7, 1, 38 - 41.
- Fitzpatrick, A. r. (1989) Social Influences in standard setting: The effect of Social Interaction on Group Judgements. *Review of Educational Research*, 59, 315-328.
- Forster, M. J. (2000) Is the information provided for examiners at award meetings appropriate for its purpose? MEd dissertation, University of Bristol.
- Forrest, G. M. and Shoesmith, D. J. (1985a) *A Second Review of GCE Comparability Studies*, JMB: Manchester

- Forrest, G. M. and Shoesmith, D. J. (1985a) *Monitoring standards in the General Certificate of Education*, JMB: Manchester
- Forrest, G. M. and Vickerman, C. (1982) *Standards in GCE: subject pairs comparisons, 1972-80*, Occasional Publication 39. Joint Matriculation Board: Manchester. In Jones, B. E. (1997) Comparing examination standards: is a purely statistical approach adequate? *Assessment in Education, Principles, Policy and Practice*, 4, 2, 249-263.
- Goldstein, H., & Cresswell, M. (1996). The comparability of different subjects in public examinations: a theoretical and practical critique. *Oxford Review of Education*, 22, 435-442.
- Gray, E. (1999) *GCSE Science (Double Award) 1998 Comparability Study Report: syllabus review and cross-moderation report*. Organised by OCR on behalf of the Joint Forum for GCSE and GCE.
- Greatorex, J. (1999) Generic Descriptors: a health check, *Quality in Higher Education*, 5, 2, 155- 166.
- Greatorex, J. (2000) *Review of research into qualification levels*. In Bell, J. F. and Greatorex, J. (2000) *A Review of Research in Levels, Profiles and Comparability*. A report to the Qualifications and Curriculum Authority.
- Greatorex, J., Johnson, C. and Frame, K. (2001) Making the Grade - Developing Grade Descriptors for Accounting using a Discriminator Model of Performance, *Westminster Studies in Education*, 24(2) 167- 181.
- Guille, R., Wiley A. and Norcini, J. J. (2001) The Occasion Effect in Standard Setting, *A paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, WA*.
- Haladyna, T and Hess, R. (2000), An evaluation of conjunctive and compensatory standard setting Strategies for test decisions, *Educational Assessment*, 6, 2, 129-153.
- Hayward, G. (1995) *Getting to Grips with GNVQs*, London, Kogan Page.
- Hodgson, A and Spours, K. (1997), *Dearing and Beyond 14-19 qualification frameworks and systems*, Kogan Page Limited: London.
- Hodgson, A. and Spours, K. (2000) *Qualifying for Success: Towards a framework of understanding*, the first report from the project Broadening the Advanced Level Curriculum: Institutional Responses to Qualifying for Success, Institute of Education / Nuffield Project: London.
- Jaeger, R. M. (1989) Certification of student competence, in R. L. Linn (Ed.) *Educational Measurement*, 3rd edn, pp. 485-514 (New York, ACE / Macmillan).
- Jaeger, R. M. (1991) Selection of judges for standard setting. *Educational Measurement: Issues and Practice*, 10, 3-14.
- Jaeger, R. M. (1995) Setting performance standards through two-stage judgemental policy capturing, *Applied Measurement in Education*, 8, 1, 15 - 40.
- JCGQ (2000a) *Notice to Centres Vocational Certificate of Education and Foundation, Intermediate and Part One GNVQ for Teaching from September 2000. Entry, Aggregation and Certification Rules, Revised Version*. JCGQ
- JCGQ (2000b) *Advanced And Advanced Subsidiary GCE: Your Questions Answered*
http://www.qca.org.uk/nq/framework/a_as_gce.asp

Jones, B. E. (1997) Comparing examination standards: is a purely statistical approach adequate? *Assessment in Education, Principles, Policy and Practice*, 4, 2, 249-263.

London Mathematical Society (1995) *Tackling the Mathematics problem*, London Mathematical Society: London.

McClune, B. (2001), Modular A-levels - who are the winners and the losers? A comparison of lower-sixth and upper-sixth students' performance in linear and modular A level Physics, *Educational Research*, 43, 1, 79-89.

McEwen, A., McGuinness, C. and Knipe, D. (1999) Comparing teaching and learning in A levels and advanced GNVQs, *General Educator Journal of the NATFHE General Education Section*, 12 - 15.

Murphy, R., Burke, P., Cotton, T. Hancock, J., Partington, J., Robinson, C., Tolley, H., Wilmut, J. and Gower, R., (1996), *The Dynamics of GCSE Awarding*. Nottingham: A report of a project commissioned by the School Curriculum and Assessment Authority.

NCVQ (1993) *NCVQ Information Note: General National Vocational Qualifications*, NCVQ : London.

Newton, P. (1997a) Measuring Comparability of standards between subjects: why our statistical methods do not make the grade, *British Educational Research Journal*, 23, 4, 433-449.

Newton, P. (1997b) Examining standards over time, *Research Papers in Education*, 12, 3, 227-248.

Norcini, J. J., Shea, J. A. (1992). The reproducibility of standards over groups and occasions. *Applied Measurement in Education*, 5, 63-72.

Norcini, J. J., Shea, J. A., and Kanya, D. T. (1988). The effect of various factors on standard setting. *Journal of Educational Measurement*, 25, 57-65.

Oates, T. (2000a) *Issues of unitisation* Paper presented at the Advisory Group on Research into Assessment and Qualifications, Qualifications and Curriculum Authority 2-3-00.

Oates, T. (2000b) *Modularisation / unitisation a structural trend that refuses to go away*, QCA: London.

OCR (1995) *Advanced GNVQ Business*, Cambridge: OCR.

OCR (2000a) *OCR Advanced Vocational Certificate of Education in Art and Design (7760) Advanced Vocational Certificate of Education (Double Award) in Art and Design (7780)*, Cambridge: OCR.

OCR (2000b) *GNVQ and VCE Construction and the Built Environment Portfolio Administration Pack Curriculum 2000, For use from September 2000*. OCR: Cambridge.

OCR (2000c) *Advanced Vocational Certificate of Education in Travel and Tourism (7775) Advanced Vocational Certificate of Education (Double Award) Travel and Tourism (7795)* OCR: Cambridge.

OCR (2000d) *Advanced Vocational Certificate of Education in Science (7775) Advanced Vocational Certificate of Education (Double Award) Science (7795)* OCR: Cambridge.

OCR (2000e) *OCR Advanced Subsidiary GCE in English Literature (3828) OCR Advanced GCE in English Literature (7828)* OCR: Cambridge.

OCR (2000f) *OCR Advanced Subsidiary GCE in Chemistry (3882) OCR Advanced GCE in Chemistry (7882) Coursework Administration Pack* OCR: Cambridge.

OCR (2000g) *OCR Advanced Subsidiary GCE in English Literature (3828) OCR Advanced GCE in English Literature (7828) Coursework Administration Pack* OCR: Cambridge.

OCR (2000h) *OCR Advanced Subsidiary GCE in Chemistry (3882) OCR Advanced GCE in Chemistry (7882)* OCR: Cambridge.

OCR (2000i) *GNVQ Centre Guidelines 2000-2001 Edition For use with Pre September 2000 Specifications only GNVQ L148.* OCR: Cambridge.

OFSTED (1994) *GNVQs in schools 1993 / 94 Quality and Standards of General National Vocational Qualifications.* HMSO: London:.

Ollin, R. and Tucker, J. (1994) *The NVQ and GNVQ Assessor Handbook*, Kogan Page: London,.

O'Leary, J. (2001) *Exams to be cut after AS 'shambles'*. The Times, 9 July 2001, page 1.

Osterlind, S. J. (1990) *College Basic Academic Examination, Form LD.* Chicago: Riverside Publishing Co.

Paterson, L. (2000) Responding to Cresswell (2000) in H. Goldstein and A. Heath (2000) *Educational Standards*, Oxford University Press: New York.

Pinot de Moira, P. (2000) *A Comparability Study in GCSE English Statistical analysis of results by board.* Organised by AQA on behalf of the on behalf of the Joint Forum for GCSE and GCE.

Plake, B. S. (1995) The Performance Domain and the Structure of the Decision Space, *Applied Measurement in Education*, 8, 1, 3-14.

Plake, B. S. (1998) Setting Performance Standards for Professional Licensure and Certification, *Applied Measurement in Education*, 11, 1, 65-80.

Plake B. S. and Hambleton, R. K. (1999). A standard setting method Designed for Complex Performance Assessments: Categorical Assignments of Student Work, *Educational Assessment*, 6(3), 197 - 215.

Plake B. S., Impara J. C. and Potenza, M. T. (1994) Content Specificity of Expert Judgments in a Standard-Setting Study, *Journal of Educational Measurement*, 31, 4, 339 - 347.

Plake, B. S., Melican, G. J. and Mills, C. N. (1991) Factors influencing intrajudge consistency during standard setting. *Educational Measurement: Issues and Practices*, 10 (2), 15-16.

Pole, D. (1961) *Conditions of Rational Inquiry: A Study in the Philosophy of Value.* London: Athlone Press.

Pritchard, J., Jani, A. and Monani, S. (2000) *A Comparability Study in GCSE English, Syllabus Review and Cross moderation exercise.* A study based on the Summer 1998 examination, Organised by EDEXCEL on behalf of the on behalf of the Joint Forum for GCSE and GCE.

QCA (1998) *GNVQ Assessment Your Questions Answered 2*, QCA, CCEA and ACCAC: London.

- QCA (UNDATED) *Comparability across subjects and families: some issues and ideas*, QCA: London.
- QCA (2000a) *GCSE, GCE, VCE and GNVQ Code of Practice 2000 / 01*, First Edition, December 2000. QCA: London.
- QCA (2000b) *Level 3 qualifications* <http://www.qca.org.uk/nq/framework/framework3.asp>
- QCA (2000c) *A level German* http://www.qca.org.uk/nq/framework/tribunal_german.asp
- QCA (2000d) *GCSE and GCE Code of Practice 2000*, QCA: London.
- QCA (2000e) *Regulations for Entry, Aggregation and Certification, GCE AS / A level Examinations for first certification in 2001 / 2002* www.qca.org.uk/nq/framework/regulations.asp
- QCA (2001a) *Qualifications Framework*, <http://www.qca.org.uk/nq/framework/>
- QCA (2001b) *Review of Curriculum 2000 – QCA’s Report on Phase One*. www.qca.org.uk/nq/framework/review2000_phase1.asp
- QCA (2001c) *GCSE, GCE, VCE and GNVQ Code of Practice 2000 / 01*, Incorporating Advanced Extension Awards, Second Edition, May 2001. QCA: London.
- Sadler, D. R. (1987) Specifying and promulgating achievement standards. *Oxford Review of Education*, 13, 191-209.
- Savory, C. (2000) Curriculum 2000 Reform, *College Research*, 4, 1, 55 - 57.
- Sizmur, S. (1997) Look Back in Angoff: a cautionary tale, *British Educational Research Journal*, 23, 1, 3 - 13.
- Stobart, G. (2000) *Changes to Post-16 Qualifications A Briefing for Higher Education on changes to the post-16 curriculum in England, Wales and Northern Ireland*. UCAS: London.
- Taylor, C., Fitz-Gibbon, C.T., & Vincent, L. (1997). Difficulties regarding subject difficulties: developing reasonable explanations for observable data. *Oxford Review of Education*, 23, 3, 291-298.
- Taverner, S. (1997) An investigation into A level entry policy of mathematics departments, *British Journal of Curriculum and Assessment*, 7, 1, 38 - 41.
- Taverner, S. and Wright, M. (1997) Why go modular? A review of modular A-levels Mathematics, *Educational Research*, 39, 1, 104, 112.
- Walton (1999) *Manual of Procedures for Joint Awarding Body Results Determination for GNVQ End of Unit Multiple Choice Tests*, Joint Compliance Unit.
- Young, M. and Leney, T. (1997) From A levels to an advanced curriculum of the future Hodgson, A. and Spours, K. (1997) (eds) *Dearing and Beyond 14-19 Qualifications, Frameworks and Systems*, pp 40 - 56, Kogan Page Limited : London.
- William, D. (1996). Standards in examinations: a matter of trust? *The Curriculum Journal*, 7, 3, 293-306.

Wolf, A. (1995) *Competence Based Assessment*. Open University Press: Buckingham.

Wolf, A. (2000) *Vocational Assessment*. In H. Goldstein and T Lewis (Eds) *Assessment: Problems, developments and statistical issues*, John Wiley and Sons: Chichester.

Woodward, W (2001) *Key Skills school tests sidelined*, The Guardian, 11 July, www.guardian.co.uk

F:\Ears\JACKIE\bera01\AVCEalastairconf.doc