CAMBRIDGE ASSESSMENT

PERFORMANCE IN GCSE EXAMINATIONS IN VOCATIONAL SUBJECTS (GCSEvs) 2004-2005

John F. Bell and Carmen L. Vidal Rodeiro

October 2006

A Report from the
Research Division, Cambridge Assessment

This report has been commissioned by the Qualifications and Curriculum Authority

Note on printing: this report is best printed in colour to ensure that the colour graphs are clear.

UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate

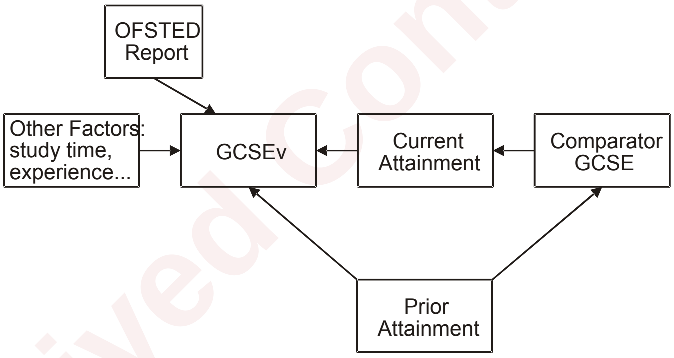
Contents

Executive Summary

1. The research addresses the following key issues:

- Do students tend to be awarded lower grades than appropriate in their GCSEvs examinations?
- Is there a problem with the grading of GCSEvs examinations?
- What are the factors that result in differences in apparent attainment?

2. GCSEv performance was compared with performance on GCSE comparators. For the purpose of this study comparators were defined as GCSE subjects closest to the GCSEvs under investigation in terms of content and skills. The underlying assumption was that the probability of obtaining a particular grade for each value of mean GCSE was the same for both the GCSEv and the comparator GCSEs. The advantage of this model was that it allowed multiple analyses of the interactions between the different factors involved (see below).The analyses included the use of concurrent attainment data (mean GCSE comparators), prior attainment data and additional analyses of variables such as study time and centre experience. Qualitative reports from OFSTED were also considered. From these sources conclusions were drawn.



3. OFSTED monitored the introduction of these qualifications and in their reports (OFSTED 2003, OFSTED 2004) they suggested a number of factors that could affect performance. A provisional cause of variation was overall lack of experience of whole centres in delivering vocational qualifications. There was evidence to suggest that some centres were achieving better results in some vocational subjects than in others. They also suggested that lack of study time and quality of teaching could have an effect on outcomes.

4. We used concurrent attainment rather than measures of prior attainment for comparability analyses in comparing differential performance because it has greater explanatory power. The underlying issue with the use of prior attainment data is that there are motivational and other factors that affect performance and these change differentially over time. Less complex methods can be useful for identifying areas of concern; however, their limitations should be recognised and reported. One of the advantages of using mean GCSE scores as a concurrent measure of attainment is that some of the factors that influence performance are controlled, for example, the same overall school effect is exerted on GCSEs and GCSEvs.

5. Performance in the following GCSEvs was investigated:

Health and Social Care	Applied Art and Design
Applied Business	Applied ICT
Leisure and Tourism	Manufacturing
Applied Science	Engineering

6 Analyses were carried out using 2004 and 2005 data.
There was no overall trend of variations in outcomes between GCSEs and GCSEvs. The following paragraphs outline how outcomes varied across the subjects with different factors leading to varied impact on outcomes.

7. For four of the subjects (Health and Social Care, Leisure and Tourism, Applied Business, and Applied Science) **there was no evidence of a lack of comparability in grading in 2004 or 2005 between GCSEvs and their comparators.**

For a summary of the relative grading of each of these subjects in relation to their comparators refer to Appendix G.

Health and Social Care	fig. G22 p.97
Leisure and Tourism	fig. G26 p.102
Applied Business	fig. G31 p.108
Applied Science	fig. G34 p.112

For each of these GCSEvs the curve is within the range of curves for the comparators. This means that for any mean GCSE value the probability of getting at least that grade is higher than for some of the comparators and lower than for others. This indicates that there are no serious grounds for concern about the grading of these GCSEvs.

8. For Applied Art and Design the analyses suggested that **GCSEv grades were higher overall for given levels of concurrent attainment.** For a summary of the relative grading of Applied Art and Design in relation to its comparators refer to Appendix G fig. G37 p.116. The curves indicate that the students were more likely to obtain grade A* and at least grade A for a given level of mean GCSE performance. Only at grade F in 2005 does the curve indicate that there may be a concern about lack of comparability of the GCSEv.

9. For three subjects (Applied ICT, Manufacturing and Engineering) the **evidence suggested that there may be a lack of comparability.** For a summary of the relative grading of each of these subjects in relation to their comparators refer to Appendix G.

Applied ICT	fig. G5 p.77
Manufacturing	fig. G17 p.91
Engineering	fig. G11 p.84

The curves for these subjects are lower than all of the comparators and this means that there may be grounds for concern about the grading of these GCSEvs.

10. When we studied the probability of obtaining at least grade A or at least grade C, even allowing for school type and gender, there was more variation between centres using GCSEvs than for centres using GCSEs. There were some variables that did explain some of the variation. These included lack of experience of centres in delivering the qualifications and lack of sufficient time allocated for study. OFSTED reports produced in 2003 and 2004 contain conclusions about the quality of provision in some subjects (see p.5). Not only do these reports suggest that there are reasonable grounds for lower performance in GCSEvs, they also suggest that there is more variation between schools for the GCSEvs than for the traditional GCSEs. They also suggest that some of the centres did not allow sufficient study time for the GCSEvs. This factor was investigated as part of the research and the results for grade C in 2004 and 2005 are presented in fig. 6 p.15. There was evidence that this was an influencing factor in ICT and Manufacturing but it does not account for all of the difference (see pp.14-15). The variation between schools (strictly examination centres) was considered. It was found that in the case of Engineering, Manufacturing

and ICT, although most schools performed poorly, there were exceptions where candidates tended to perform very well.

11. Having two years of data meant that it was possible to investigate whether centres with more experience perform better than inexperienced centres. The results (see fig.8 p.17) suggest that experienced centres do, in fact, perform better than inexperienced centres. However, this factor does not fully explain the difference between GCSEs and GCSEvs.

12. These are the factors investigated in the research:

1. Concurrent attainment (mean GCSEs)
2. Amount of study time (total number of GCSEs and GCSEvs taken)
3. Centre experience
4. Motivation
5. Prior attainment (KS2 and KS3 scores)
6. Quality of teaching

There are statistical data on 1, 2 and 3 which suggest that these factors affect performance in some GCSEvs.

There is supplementary information from OFSTED reports on 2, 3, 4 and 6 which suggests factors that could lead to lower GCSEv grades than predicted by concurrent attainment.

There are statistical analyses indicating that the use of prior attainment data (5) can give misleading results possibly as a result of motivational effects (4).

Overall the research indicated some evidence of differences between GCSEv grading against the comparators for some subjects. However, there was evidence to suggest that factors such as study time and experience may have had an effect on the differential grading based on concurrent attainment. For other subjects there was no cause for serious concern.

.

1 Introduction

In 2002 teaching started on a new qualification: GCSEs in vocational subjects (GCSEvs). These qualifications were in Applied ICT, Engineering, Manufacturing, Applied Business, Health and Social Care, Leisure and Tourism, Applied Science and Applied Art and Design.

GCSEvs were designed to:

- replace Part One GNVQs;
- provide an introduction to a broad vocational area;
- enable progression to further education, training or employment;
- be available at Key Stage 4 and post-16.

More details of these qualifications are given in Appendix A.

The first results for these examinations were issued in 2004. Concerns were expressed that candidates taking GCSEvs, except Applied Art and Design, were less likely to obtain A*-C grades than candidates taking traditional GCSEs. There could be two competing explanations for this. First, the GCSEvs were graded harshly in comparison to other subjects or, secondly, the candidates were less likely to produce work that was of the standard required for the grades. For most of the GCSEvs there is some overlap in content with traditional GCSEs.

It is useful to consider how the grade boundaries were set and how the decisions about pass marks were made by awarding committees which had overlapping memberships. Experienced members of the teaching profession, who occupy the senior examining positions including those responsible for writing the examination papers, decided that the candidates had not reached the standard expected for particular grades based on prior attainment. These decisions resulted in fewer high grades than those predicted by the statistics available at the awarding meetings but by their nature statistical predictions do not consider what the candidates actually did. These discussions were monitored by QCA who observed that there were cases in all awarding bodies in award meetings for GCSEvs where the awarders were unable to reconcile the script evidence with the statistically derived expected outcomes. In the case of OCR, there were no written reports that any of these committees had any concerns about the appropriateness of the grade boundaries set in the meeting.

OFSTED monitored the introduction of these qualifications and produced two reports. In the first interim report (OFSTED, 2003), which contained some conclusions that may have an impact on the performance of GCSEvs, they found that:

- Schools had not planned well enough for the introduction of the new courses.
- Schools had provided little or no training for their teachers.
- Schools had guided only lower-attaining students to take the new courses.
- The standards of achievement and the quality of provision, especially teaching, are sound overall, although there is too much variation in standards and quality in particular subjects.
- The satisfactory or better work found, in particular, in Applied Art and Design, Engineering, Manufacturing, Applied Business and Applied Science, is especially encouraging.
- Students are generally positive about the courses and appear to be motivated by them, at least at this stage of their course development.
- Too many teachers are still unsure about the assessment aspects of these qualifications.

In the final OFSTED report on GCSEvs (OFSTED, 2004), the main findings included:

- Students' achievement is satisfactory or better in three quarters of lessons and good or better in a third. It is, however, unsatisfactory in a quarter. This compares unfavourably with the average for all GCSE subjects at Key Stage 4.
- Although there are examples of high achievement in all subjects, there are considerable overall differences among them.
- Students' achievement is often good in Engineering, and sometimes good in Applied Business.
- In Applied Science, achievement is higher than in the traditional double award courses when the prior attainment of students is taken into account.
- In some schools where the target group for the new courses is mainly lower-attaining students, achievement is often unsatisfactory, especially in Leisure and Tourism.
- In a large minority of schools, the volume of work students do, and the breadth and depth of their studies, do not always add up to the weight of a double award GCSE, especially (but not exclusively) in those schools where insufficient teaching time is allocated.
- The quality of teaching is satisfactory or better in nearly nine tenths of lessons, good or better in nearly a half, but unsatisfactory in a tenth. Again, these figures compare unfavourably with those for other GCSE subjects in Key Stage 4.
- Most teachers were unclear about important assessment requirements and lacked confidence in being able to judge students' attainment against the criteria set by the awarding bodies.

Not only do these reports suggest that there are reasonable grounds for the lower performance in GCSEvs but they also suggest that it would be sensible to consider two hypotheses. First, that there is more variation between schools for the GCSEvs than for the traditional GCSEs. Secondly, the study time allocated should be investigated by considering the number of GCSEs and GCSEvs taken by candidates as an approximation. The more subjects studied implies that there is less study time for each individual subject.

The objective of this report was to investigate, with statistical modelling, the evidence for differences in performance between GCSEvs and traditional GCSEs (these will be referred to as 'comparators') and to investigate possible factors influencing differential performance. In particular, the three following research questions were posed:

- Do students tend to be awarded lower grades than appropriate in their GCSEvs?
- Is there a problem with grading of GCSEvs?
- What are the factors that result in differences in apparent attainment?

The questions were addressed by analysing GCSE and GCSEvs results data from 2004 and from 2005.

The next section of this report (section 2) is a description of the issues and techniques used in the investigation of comparability. This is followed, in section 3.1, by a detailed account of the analysis and results for Applied ICT Double Award. It is recommended that this section is read carefully before looking at the results for other GCSEvs. Shorter descriptions of the analyses of the other seven GCSEvs are presented in sections 3.2 to 3.8. The report concludes with a discussion of the results and their implications (section 4).

2 The research methodology and its rationale

This research arose principally out of concerns expressed to QCA by Jesson (2005) about the grading standards of GCSEvs. He was concerned that it was harder for candidates to obtain good grades in GCSEvs compared with the pre-existing equivalents. For example, he compared the distributions of Key Stage 2 scores for various groups of students and presented his results in tables like the one below (Table 1). He proposed that GCSEvs might have been harshly graded e.g. of students with Key Stage 2 scores from 29-30, 75% gained C or better for GCSE Science, whereas only 47% achieved this in GCSEv Applied Science.

Table 1: % of entry gaining grade C or above in Science (GCSE and GCSEv) vs. Key Stage 2 scores

KS 2 Points	Low	Below Av	Average	Above Av	High	Mean
	- 22	23 - 25	26 – 28	29 –30	31++	
GCSE Science	5%	20%	48%	75%	93%	51%
GCSEv Applied Science	9%	21%	38%	47%	54%	27%

Whilst this method is useful as a screening technique for identifying potential problems it is not effective enough to give definitive answers about comparability.

Comparability is a complex issue and there is a considerable body of research that considers it (Goldstein and Cresswell, 1996; Newton, 1997; Dexter and Massey, 2000; Jones, Baird and Arlett, 1997). For this report, two aspects are being considered: comparability between various subjects and comparability between GCSEs and GCSEvs. The latter depends on the nature of the comparators for an individual subject. A comparator in this study means a subject that makes similar demands of a candidate and/or is a potential alternative (e.g. all the GCSE Science specifications can be considered comparators for GCSEv Applied Science). The first stage of this comparability analysis was deciding on the GCSE specifications to compare with each GCSEv (see Appendix B for a list of comparators).

When there are different types of specifications for the same subject and the same type of examination, then the following very strict definition of comparability (Bell and Greatorex, 2000) can be used:

> 'two examinations are comparable if students who demonstrate the same level of attainment obtain the same grade.'

The above definition can be widened to a less strict definition suitable for comparing related subjects rather than variants of the same subject:

> 'two qualifications can be comparable if they either indicate the same degree of knowledge, skills *or* potential, or the same degree of skills *and* potential, or if they indicate the same potential.'

When two qualifications are deemed to be comparable for a particular purpose they are described as equivalent. It is also necessary to consider different situations. Depending on the subject, the nature of the comparability being investigated is different. In some cases the comparability is with traditional subjects in the same subject area (e.g. GCSEv Applied ICT with GCSE ICT), while in other cases this involves comparing similar subjects, for example, different forms of Design and Technology. There may be some overlap in terms of knowledge but this can be variable. It is assumed that the more general, transferable skills are the same or very similar.

The main technical problem with methods of statistical comparability is that the allocations of centres and candidates to the GCSEvs and their comparators are not random processes. Candidates take GCSEvs as a result of a process of channelling and choosing. These selection effects have potential implications for GCSEvs. One of the reasons for their introduction is that they were supposed to improve the motivation of disenchanted candidates. This can be a particular problem for studies involving prior attainment given that attitudes to school can change in the teenage years: two students who enter secondary schools with the same Key Stage 2 scores may develop differently over Key Stage 3. This also has possible implications for the individual subjects.

Because there is no way of distinguishing between a difference caused by differing grading standards or by a specification facilitating greater attainment, statistical methods can never give an absolutely definitive answer to comparability issues. However, statistical analyses can be used to investigate whether some explanations are plausible and as a screening process to identify areas of concern.

Statistical comparability investigates how the relationship between examination performance and a measure of prior/concurrent attainment varies by specification or specification type. One technique is multiple regression which estimates average relationships between response (e.g. educational attainment) and predictor variables (e.g. concurrent ability). Additional variables can be added to investigate how they change this relationship. It is not satisfactory for this data because it is likely that the process of student selection by schools or the education given by schools may influence outcomes. Two students within a particular school could then tend to be more similar than two students from different schools. Generally we are interested not only in the average relationship but in how this relationship varies from school to school. Multilevel modelling is an extension of multiple regression and provides a powerful framework for doing this.

The technique used in the report is multilevel logistic regression modelling. The word 'logistic' refers to the fact that probability is being modelled with the result that it is possible to generate a curve indicating how the probability of obtaining at least a particular grade changes with different values of mean GCSE. Mean GCSE is calculated by assigning points to each GCSE grade: A* = 8, A = 7 … G = 1, U = 0. When a categorical variable (e.g. whether the candidate entered for a GCSEv or GCSE) is added to the model it is then possible to generate a curve for each category. When one curve is to the right of another then this indicates that the probability of obtaining at least that grade is lower for any given value of the explanatory variable (mean GCSE). The curves are explained in more detail in the following paragraphs.

The word 'multilevel' indicates that it is assumed that each centre is considered potentially to have a different curve from the other centres. In the report, curves have been generated for the average centres. Any differences observed between the curves are for average centres. However, the variation between centres is important. When there is a difference between two categories for an average centre, for example, the GCSE curve is higher than the curve for GCSEv, it is still possible for the curve for a given centre taking GCSEv to have a curve that is higher than some centres taking GCSE if there is enough variation. When this is true, the choice of centre is more important than the choice of type of qualification. There are, however, problems of graphing this aspect of multilevel models. Given the number of centres involved it is not practicable to produce interpretable plots with a curve for each centre. Instead it is useful to present the distribution about a meaningful reference value (in this case, the mean GCSE equal to the grade being considered). The probability values generated can then be presented as histograms.

The logistic curve is illustrated in Figure 1 where a binary dependent variable (i.e. only taking the values 0 and 1 – not obtaining and obtaining a particular grade respectively) is related to a continuous independent variable (in this case mean GCSE). In the middle of this plot about half the candidates pass and half fail so the probability is 0.5. Since a probability cannot be greater than 1 or less than 0, a linear relationship is not usually appropriate. The logistic curve is appropriate for the analysis in this report because it is likely that candidates with mean GCSEs above 7 (i.e. a mixture of A* s and As) would be likely to get at least a grade F in the GCSEvs or its comparators. This is a ceiling effect, i.e. candidates with a range of very high values (60 or above in Figure 1) of the independent variable are almost certain to obtain at least the grade in question. Similarly, a floor effect occurs at very low values of the independent variable. All the analyses described in this report involve this type of regression. It can be argued that more complex models can be used but the advantages are limited (see Appendix C).

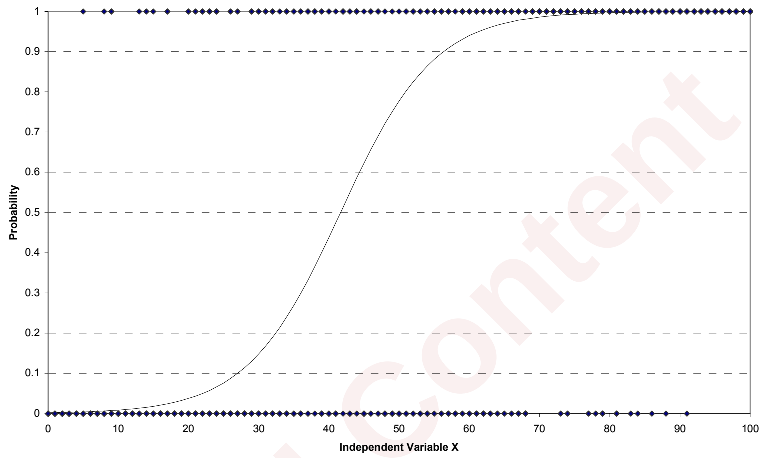


Figure 1: Example of the relationship of a binary dependent variable and a continuous independent variable X

Figure 2 illustrates the multilevel aspect of modelling. Two solid red and blue lines represent the average centre taking GCSE and GCSEv. However, it is possible to fit a line for each centre: these are represented by the dashed red and blue lines. In the presentations of the results, the point at which the red and blue lines cross the black vertical line is considered. The best GCSEv centre has a probability of 0.95 of obtaining at least the grade in question at attainment 0 on the sketch graph and the worst centre has a probability of 0.52.

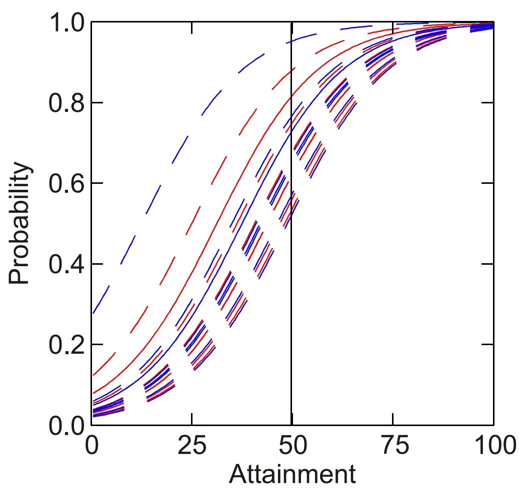


Figure 2: Hypothetical example of a multilevel model

The next issue to consider is how the prior or concurrent attainment influences the results. Appendix D discusses the advantages and disadvantages of using each of these two measures in the comparability analyses. It also includes some analyses using prior attainment. It was found that the probability of obtaining particular grades in mathematics was lower for those entered for GCSEvs than for those entered for the GCSEvs' comparators. Since the candidates in both groups were entered for the same mathematics examination it is clear that the progress of the two groups differed. This means that prior attainment is unsuitable for comparability purposes and also supports the awarding meeting decision not to follow predictions made on prior attainment. The parameter estimates for analyses based on prior attainment are given in Appendix H and show they would have given a false impression of the extent of the lack of comparability of the GCSEvs.

In this report, the main analyses were carried out using concurrent attainment (mean GCSE). Using this type of measure means that some of the factors that influence performance are being controlled. For example, mean GCSE is influenced by the same overall school effect as the GCSEvs and it may be influenced by external factors that can vary for entries for the qualification but occur after a prior measurement was made.

Given this research base, the methodology used in this report involves the statistical modelling of the National Pupil Database using logistic multilevel models. Although there were many variables available, it is important to recognise that some potential important factors cannot be considered. In particular, there is no evidence of the level of motivation for an individual GCSEv and its comparators (general motivation is indirectly controlled for by concurrent attainment since it is assumed to affect all GCSEs) and of the quality of teaching in the subjects. Although there are no data, there is some evidence of differences in the quality of teaching from the OFSTED reports about the introduction of GCSEvs cited in the introductory section. Also, the OFSTED ratings of quality of teaching were shown to be related to examination performance in Bell (2003).

All specifications for all eight GCSEvs will be considered, together with those for the traditional GCSE subjects that are closest to them in content and skills, that is, each GCSEv specification will be compared with a range of traditional GCSE specifications. The choice of comparators will be more obvious for some subjects (for example, Applied Business or Applied Art and Design), than for others (for example, Health and Social Care). However, the aim is to ensure that comparisons are as valid as possible. A further reason for this design is that centres may choose between these subjects when considering what to offer in a subject area.

The following binary variables were created and analysed using a series of multilevel logistic regression models:

- probability of obtaining a grade A*;
- probability of obtaining at least a grade A;
- probability of obtaining at least a grade C;
- probability of obtaining at least a grade F.

The aim of using these models was to respond to the research questions posed in the Introduction, that is, to investigate factors influencing performance on the GCSEvs and to try to explain the differences in the observed grade distributions.

3 Results

3. 1 Applied ICT Double Award

Introduction

Applied ICT Double Award was chosen as the first subject for analysis because it had a large entry and only one GCSE comparator. The analysis of this subject is explained in greater detail than the analysis of other subjects (note that parameter estimates and figures for all analyses at each grade can be found in Appendices E, F and G).

In 2004 there was a total of 36,891 students taking Applied ICT Double Award, which compares with 618,052 students taking at least one GCSE. These students were grouped in 798 schools (out of 3,952 centres offering GCSEs), giving an average entry per school of 46 students, which would suggest that many centres have multiple teaching groups.

In 2005 there was a total of 43,557 students taking Applied ICT Double Award. There was, therefore, an increase of 18% in the entries for this GCSEv. These students were grouped in 968 schools, 170 more schools than in 2004, which suggests that the popularity of this GCSEv is increasing. Note that this increase represents schools starting their students on GCSEvs in September 2003.

Applied ICT Double Award entries are displayed in Table 2.

Table 2: Applied ICT Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	19,007	23,112
OCR	10,012	11,735
AQA	7,872	8,710
Total entry	36,891	43,557

GCSE in ICT (full course only) was used as the comparator for Applied ICT Double Award. A total of 70,566 students took GCSE in ICT in 2004. This means that approximately 34% of ICT entries were for the vocational version. In 2005, the number of students taking the comparator increased to 74,669.

Figure 3 shows the grade comparison between the ICT subjects in 2004. GCSE grades were higher than those for GCSEvs. The percentages of the high grades (A to C) increased in 2005 for the GCSEv.

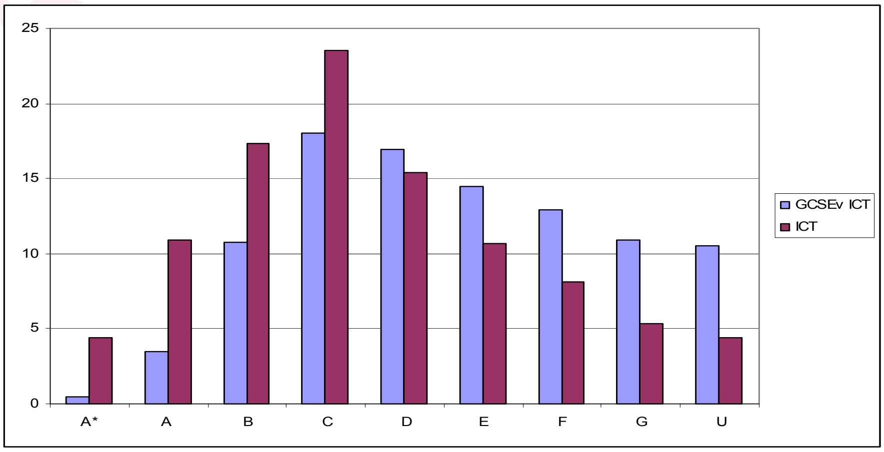


Figure 3: Grade comparison among ICT subjects (% in grade) in 2004

Multilevel model for concurrent attainment

The objective of the following statistical analysis was to explore, using a multilevel model, the reasons for the pattern of performance. In a first step, an exploratory analysis was carried out to investigate whether the model was appropriate. The details of this analysis are given in Appendix C.

OFSTED (2004) makes reference to the fact that there is considerable variation between the centres with regard to the implementation of GCSEvs (the term 'centres' is used in this report because examinations are not only taken in schools but in other kinds of examination centres). To test this hypothesis, we fitted a logistic multilevel model where the school level variance was modelled as a function of the subject: vocational or non-vocational. This means that it was possible to investigate whether there was greater variation for centres using GCSEvs compared to centres using GCSEs. Such differences are not generated by grading standards but by the influence of factors operating at the centre level. These factors include but are not limited to differences in teaching effectiveness. Technical details of this analysis can be found in Appendix C. The model fitted two curves: one for the GCSEv and one for the comparators, and assumed that the variance between centres differed. The parameter estimates are given in Appendices E (for 2004 data) and F (for 2005 data). Note that the number of students taking the GCSEv and getting a grade A* is very small and the model cannot be fitted accurately using MLwiN[1].

Figure 4 shows the probability of obtaining at least grade C for the GCSEv and the GCSE in 2004. Dotted lines have been drawn to illustrate the predictions for a student with a mean GCSE of 4 (i.e. grade D; for example, a student is likely to have a mixture of grades Cs, Ds and Es with the number of Cs equal to the number of Es). For the GCSEv the predicted probability of obtaining at least grade C is just less than 0.2, compared with a probability of 0.3 for students taking the GCSE. Students taking the GCSEv had a lower probability of success for any given mean GCSE. This pattern also appears for at least grade A and at least grade F (graphs for all grades are given in Appendix G). However, this does not mean that the GCSEv was graded harshly. It might be the case that students with equivalent mean GCSEs did not, on average, produce work of as good a quality as the GCSE. Although a statistical analysis of the results cannot answer this question definitively, it is possible to investigate possible alternative explanations.

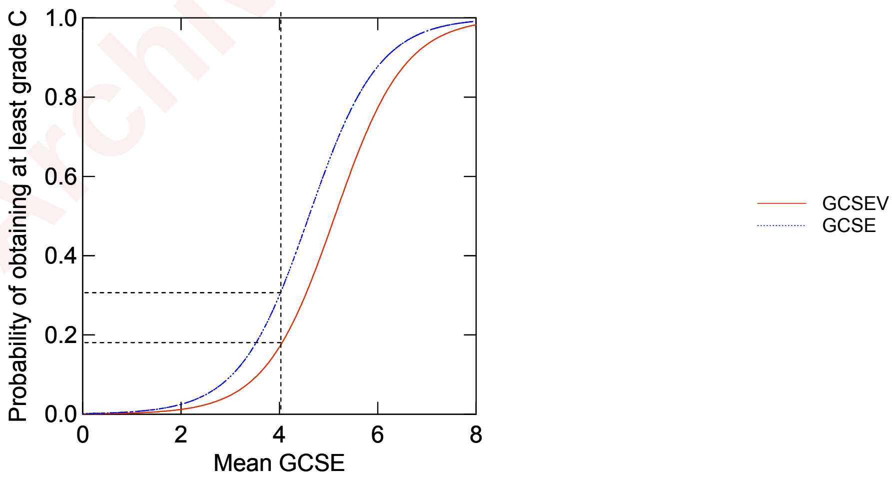


Figure 4: Probabilities of obtaining at least grade C by mean GCSE for Applied ICT Double Award and its GCSE comparator in 2004

The model used in this analysis allows the variation in the probability of obtaining a grade to vary by centre and for the distribution of this variation to differ by centres. If these distributions overlap then in some centres taking GCSEs the students will have the same probability of

[1] Software used to fit the Multilevel models in this report (http://www.cmm.bristol.ac.uk/MLwiN/index.shtml)

success as those in the centres taking GCSEv. In effect, while the average GCSE centre has the curve shown in Figure 4, it is possible that a poorly performing centre could have a curve that coincides with the one for the average GCSEv centre and a high performing GCSEv centre could have a curve equal to the average GCSE centre.

In 2005, there was more variation between centres using GCSEvs than for centres using GCSEs, in particular for grade A where the variation is 3.045 for centres offering GCSEv and 1.777 for centres offering only the GCSE comparator (this variation corresponds to the variance components of the multilevel model described in Appendix C). However, the magnitude of the variation for centres offering the GCSEv in 2005 decreased considerably compared to the variation in 2004.

The effects of the estimated variance components mentioned above are difficult to interpret, so they are illustrated by calculating the expected probabilities of obtaining at least grade A, grade C and grade F for students averaging grades A, C and F, respectively, in their other GCSE examinations in each centre (centres for GCSEvs and centres for GCSEs). Therefore, from Table 3, a student averaging grade A in his/her GCSEs had a probability of obtaining at least grade A of 0.36 in the GCSEv or 0.58 in the GCSE. However, in 2004 there was considerable variation between centres as indicated by the magnitude of the standard deviation (SD). This difference is reflected in the reference interval which indicates the range of probabilities that 95% of the centres lie in. For example, there are centres in which candidates were virtually certain to get a grade A in ICT with an average of grade As on their other GCSEs (i.e. usually a combination of B, A, and A* grades).

Table 3: Predicted probabilities of obtaining at least grade A, grade C and grade F

		2004		2005	
		Centres for GCSEv	**Centres for GCSE**	**Centres for GCSEv**	**Centres for GCSE**
At least a grade A	Mean	0.36	0.58	0.34	0.60
	SD	0.23	0.19	0.32	0.35
	Reference Interval	(0.15,0.95)	(0.31,0.93)	(0.03,0.96)	(0.02,0.99)
At least a grade C	Mean	0.44	0.63	0.39	0.61
	SD	0.23	0.2	0.31	0.32
	Reference Interval	(0.12,0.95)	(0.25,0.89)	(0.01,0.94)	(0.04,0.9)
At least a grade F	Mean	0.45	0.74	0.30	0.53
	SD	0.20	0.18	0.27	0.31
	Reference Interval	(0.09,0.74)	(0.28,0.89)	(0.01,0.83)	(0.03,0.95)

This variation can also be illustrated in a figure. As an example, the distributions of these predicted probabilities in 2004 at grade C are given in Figure 5. The GCSEv students attending centres with the highest predicted probabilities have a greater chance of obtaining a grade C than GCSE students in an average centre. The greater variation for the GCSEv is illustrated by the fact that the distribution is flatter than that for GCSE. The distribution of the predicted probabilities for GCSEv indicates that the bulk of the centres had low predicted probabilities but there is a long thin tail of higher probabilities. That is, in most schools entering GCSEvs, students are predicted low probabilities of obtaining at least grade C, but there are some with high probabilities. For both types of qualifications there are centres where students averaging grade C in their other GCSE subjects are predicted to be almost certain to obtain a grade C in ICT. These results are consistent with the OFSTED report showing that there is considerable variation between the schools with regard to the implementation of GCSEvs. In fact, one of the most disturbing aspects of Table 3 is the fact that for both types of ICT subjects there are centres where students have little chance of emulating their results in other subjects. The equivalent figures for at least grade A and at least grade F in 2004 and 2005 are given in Appendix G. Although there is variation in the shape of the distributions, the predicted probabilities are always wide and overlap considerably. For both GCSE and GCSEv, there is evidence that in some centres grade C average candidates in non-ICT subjects almost always obtain a grade C in ICT whilst in others candidates almost never obtain a grade C in ICT. This is a cause for concern regarding the teaching of these courses.

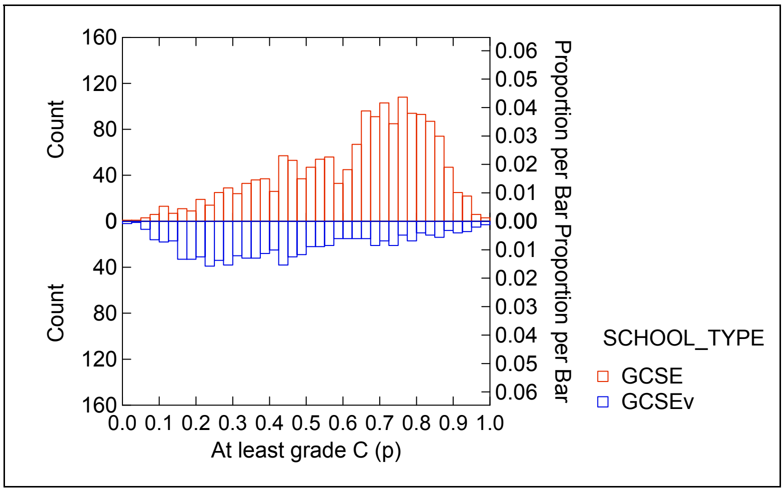


Figure 5: Predicted school effects by type of school[2] (at least grade C, 2004)

Other factors, such as gender or school type (comprehensive, grammar, independent, etc.), might have a role in the performance of students. When school type was introduced into the model, only grammar schools and independent schools had a significant effect on the probability of obtaining at least a certain grade. The number of these schools that offer the GCSEv was small and not representative. The effect of gender was significant but very small. Even allowing for school type and gender, the model shows that there is more variation between centres using GCSEvs than there is for centres using GCSEs. This suggests differences in implementation and teaching.

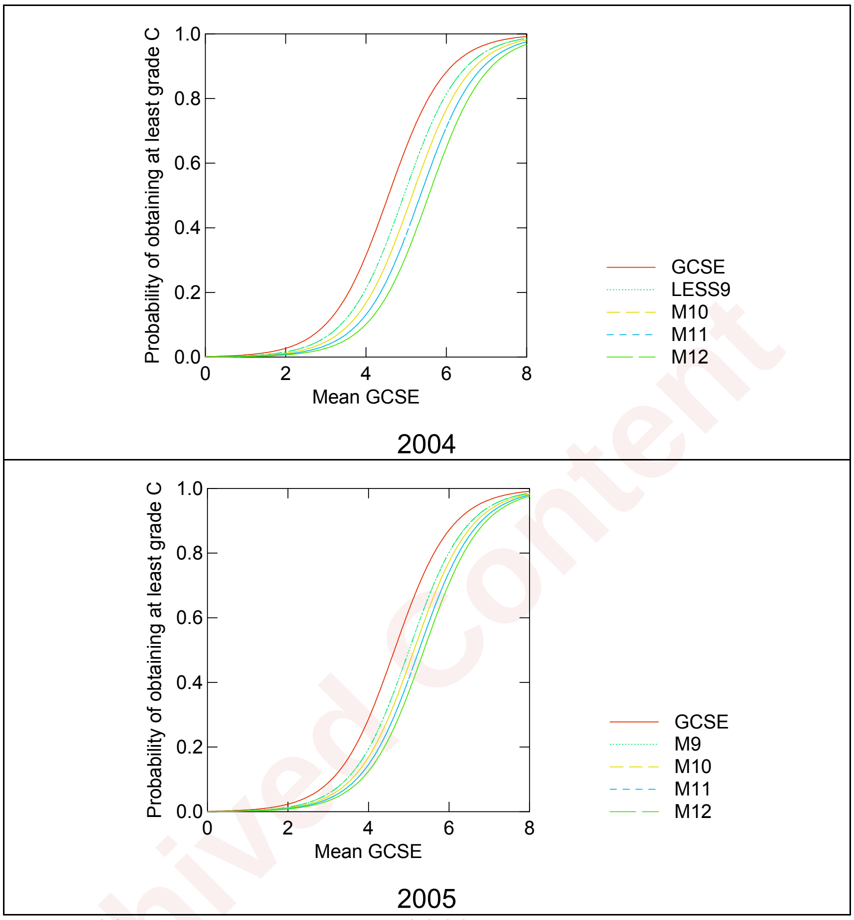
Effect of the number of GCSEs

One of the OFSTED report's findings was that some centres did not allow sufficient time for the GCSEvs (OFSTED, 2004). To test this hypothesis we created a new student-level variable, 'total number of GCSEs', by adding the number of GCSEs and the number of GCSEvs multiplied by two (since they are equivalent to two GCSEs). In the remainder of the report, the term 'total number of GCSEs' will be used to refer to the total number of GCSE and GCSEvs with double awards counting as two. From this variable, we derived a new school-level variable: 'median of total GCSEs per school'. The median was chosen in preference to the mean because it is less sensitive to outliers (unusual candidates who do very small or very large numbers of GCSEs). This new school-level variable was introduced into the model. It is worth noting that the distribution of median GCSE results differs for GCSEs and GCSEvs centres. For GCSE centres, in 2004, 51% of centres had a median of greater than or equal to 9 and 12% had a median greater than or equal to 10. In GCSEv centres the students were studying for more examinations with the equivalent figures being 76% and 40%.

The probabilities of obtaining at least grade C in 2004 and 2005 taking into account the median number of GCSEs are presented in Figure 6. The GCSE line is the predicted relationship for the conventional GCSE examinations (median GCSE was not significant for these qualifications). The other lines are the predicted relationships for the GCSEvs and different levels of median GCSE (less than or equal to 9, 10, 11 and 12 or more - there were not many centres with medians of less than 9 or greater than 12). These figures show a similar pattern. In 2004, the probability of obtaining a particular grade decreased with increasing median GCSE. Although this does support the OFSTED hypothesis of some centres not allowing enough teaching time, it does not explain all the differences between GCSE and GCSEv: the probability curve for centres with a median GCSE less than or equal

[2] School type GCSE means that the school only offers GCSEs in ICT. A school of the type GCSEv offers the Applied ICT Double Award.

to 9 is still below that of the GCSE. In 2005, the variation in the curves for the GCSEv is smaller. This would seem to indicate that centres are better at time management relating to the provision of this course. Note that GCSEvs are two year courses so the 2005 cohort started before the 2004 cohort finished. This means that in 2005 the centres did not have the experience of the full course for the 2005 cohort. Additionally, there would have been little time to re-think curriculum provision.



LESS9/M9 – median number of GCSEs lower or equal to 9
M10 – median number of GCSEs equal to 10
M11 – median number of GCSEs equal to 11
M12 – median number of GCSEs higher or equal to 12

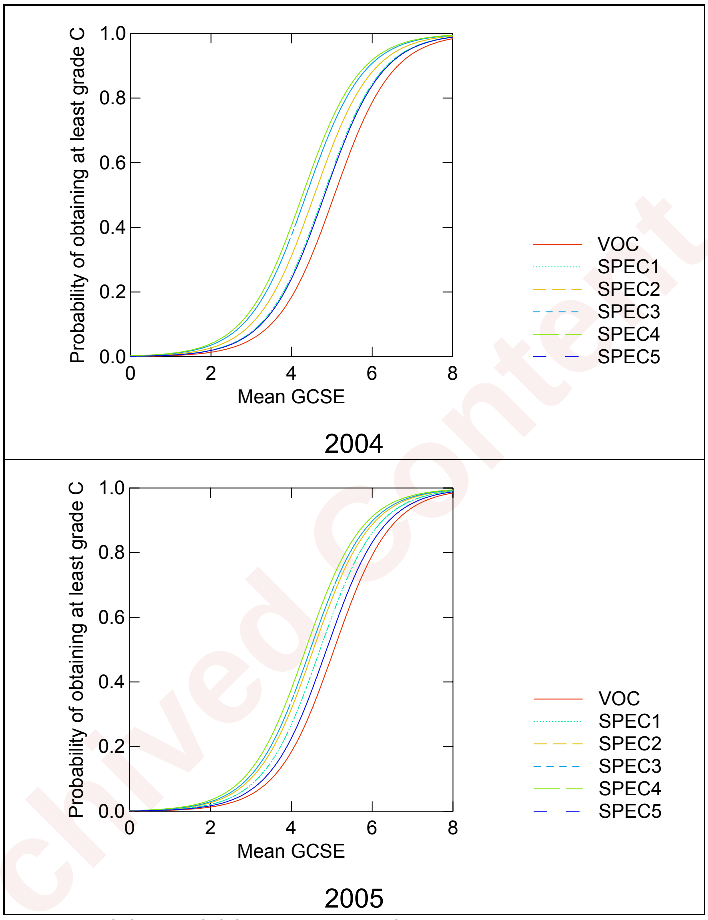
Figure 6: Probability of obtaining at least grade C in Applied ICT Double Award by the median number of GCSEs in the school

Comparison of GCSEv and each comparator

Since statistical comparability studies involve large numbers of students, relatively small differences will turn out to be statistically significant. This means that there has to be a degree of subjective judgement about the level of tolerance. Therefore, it would be reasonable to accept that there is no lack of comparability if the GCSEv is within the range of its comparators. It is reasonable to expect some differences between curves for different subjects because overall level of motivation of students, quality of teaching, level of resources, etc. might all vary from subject to subject. It would be unlikely to have curves coinciding completely. The range of variation between the specifications depends on the set of specifications. When the set is tightly defined, the curves should be closer together than when the set is diverse.

The next step in the analysis was to investigate whether the GCSEv lies in a plausible range of variation within its comparators. For Applied ICT Double Award there is only one comparator so all the different GCSE ICT specifications offered by the three awarding bodies are considered in this section.

In Figure 7 the probabilities of obtaining at least grade C in each of the GCSE ICT specifications and in the GCSEv in both years of the study are presented. The GCSEvs curves are lower than the curves for the other subjects. The curves for at least grade A and grade F are also lower for the GCSEv than for the GCSE. This indicates that the differences in performance between the GCSEvs and the average of the GCSEs are greater than the differences between the individual comparators.



VOC	GCSE in Applied ICT Double Award
SPEC 1	GCSE in ICT A – OCR
SPEC 2	GCSE in ICT A – AQA
SPEC 3	GCSE in ICT B – OCR
SPEC 4	GCSE in ICT B – AQA
SPEC 5	GCSE in ICT – EdExcel

Figure 7: Probability of obtaining at least grade C in Applied ICT Double Award and in each of its GCSE comparators

Effect of experience

Having two years of data means that it is possible to test the hypothesis that centres with more experience in the GCSEv perform better than inexperienced centres. The variation between schools with regard to the implementation of the Applied ICT Double Award should then have diminished over time if it was caused by lack of experience. To test this hypothesis a new variable was created. This new variable takes the value 1 if the school taught the GCSEv in 2004 and 2005, and 0 if 2005 was the first year the GCSEv was taught.
The effect of experience is displayed in Figure 8. There are five curves in the figure. One curve is for the GCSE results in 2005. The other four curves relate to the GCSEvs:

GCSEV_OLD is the curve for 2005 for students in centres that had taught the GCSEv in 2004; GCSEV_NEW is the curve for students in 2005 from centres that entered students for the GCSEv in 2005 for the first time; GCSEV2004 is the curve for the results in 2004 and GCSEV2005 is the curve for the results in 2005.

The same pattern of results is observed for at least grades A, C and F (only results for at least grade C are displayed in Figure 8): experienced centres performed better than inexperienced centres. However, this is only a partial explanation of the differences between GCSE and GCSEv because all lines are still to the right of GCSE 2005.

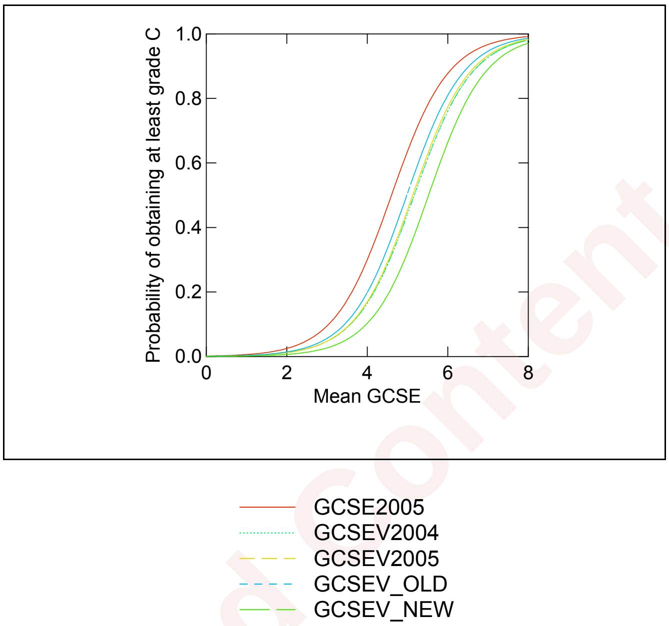


Figure 8: Probability of obtaining at least grade C and in Applied ICT Double Award and in each of its GCSE comparators (including school experience)

Summary of results

The probabilities of obtaining at least grades A, C or F are lower for GCSEv students for a given level of concurrent attainment compared with all GCSEs when no other factors are considered. There were also considerable variations between individual centres with the better GCSEv centres outperforming average GCSE centres.

The results are consistent with the observations made by OFSTED (OFSTED, 2003, 2004) that some centres did not give their students sufficient time for this GCSEv and that many centres were having problems with the introduction of this new and differently structured qualification. The former was illustrated by the results of the median variable analysis and the latter by the variation between centres and the difference between experienced and inexperienced centres.

In conclusion, the statistical analyses have identified differences between Applied ICT Double Award and its comparators but some of these differences relate to the inexperience of the centres with this new qualification and a failure to allow sufficient time for teaching. Given that even the schools classified as being experienced in this report had not completed the full GCSEv course when they started to teach their 2005 cohort, it is possible that there will be a further narrowing of the gap as experience increases.

3.2 Engineering Double Award

Introduction

Only a summary of the results for Engineering Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

In 2004 there was a total of 5,060 students taking Engineering Double Award. These students were grouped in 384 schools, giving an average of 13 students per centre. This would suggest that in some centres the teaching group size must have been very small and that the provision of teaching for this GCSEv might need further investigation.

There was a total of 7,589 students taking Engineering Double Award in 2005, grouped in 546 schools. There was an increase in both the number of students (50%) taking this GCSEv and in the number of centres (42%) offering it. The entries per awarding body are shown in Table 4.

Table 4: Engineering Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	3,203	4,782
OCR	1,857	2,807
Total entry	5,060	7,589

The entry figures for the comparators for Engineering Double Award are displayed in Table 5. There was a drop of 8% in the entry of the GCSE Design and Technology specifications: a total of 422,733 students entered one of these courses in 2005 compared with 458,624 students in 2004. The major decrease was suffered by Design and Technology - Graphic Products; Design and Technology - Product Design showed a slight increase. In 2005, Engineering Double Award was taken by around 2% of the total Design and Technology entry.

Table 5: GCSE comparators

Subject	Entry 2004	Entry 2005
Design & Technology - Electronic Products	18,490	17,119
Design & Technology - Food Technology	114,460	108,247
Design & Technology - Graphic Products	114,279	100,371
Design & Technology - Industrial Technology	1,290	1,270
Design & Technology - Resistant Materials	106,846	99,054
Design & Technology - Systems and Control	27,370	22,461
Design & Technology - Textiles Technology	61,771	57,470
Design & Technology - Product Design	13,583	16,207
Electronics	535	534
Total Entry	458,624	422,733

Figure 9 shows the grade distribution for Engineering Double Award and its comparators in 2004. From it, it is clear that the level of performance was poor for this GCSEv compared with all the other comparators. The percentages of students obtaining grades A*-C in the GCSEv has not changed very much in both years of the study (24.7% students obtained grades A*-C in 2004 and 25.4% students in 2005).

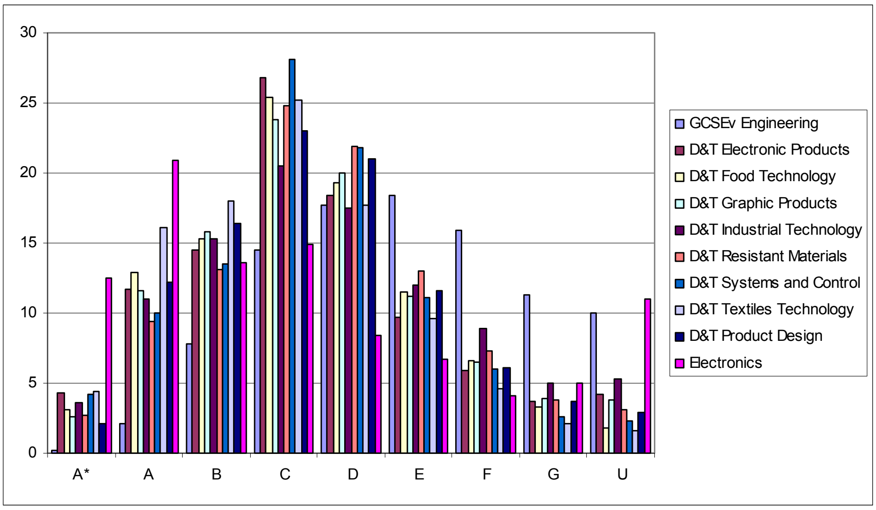


Figure 9: Grade comparison among Engineering Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

For all grades and both years, the probabilities for average students and average centres were lower for the GCSEv than for the comparators at each level of concurrent attainment. The predicted probabilities of obtaining at least grade C in 2004 by mean GCSE are presented in Figure 10.

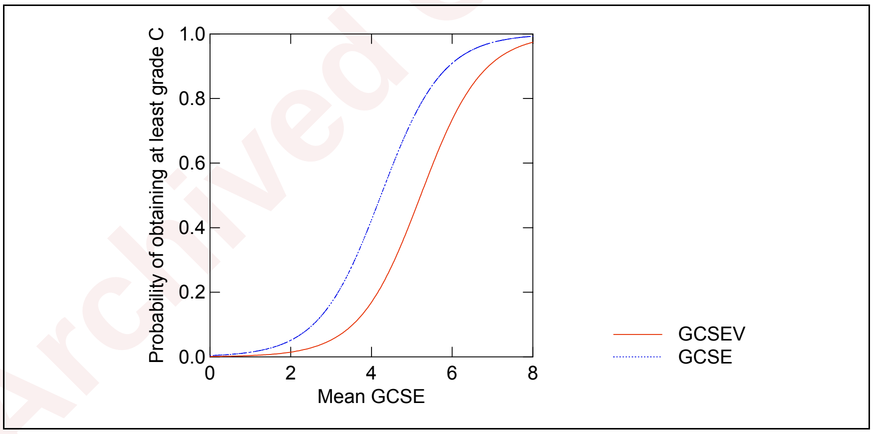


Figure 10: Probability of obtaining at least grade C in Engineering Double Award and in its GCSE comparators in 2004

The next stage in the analysis was to consider the school level variation. For the traditional GCSEs, the ranges of variation in the predicted probabilities of obtaining at least grade A, C and F are wide (see Figure 11 for grade C). For the GCSEv, the centre probabilities tend to be low but there are a few with high probabilities and there are many centres taking traditional subjects which have similar low predicted probabilities and were not very successful in matching achievement in Design and Technology subjects to the achievement in the other GCSEs that they offered.

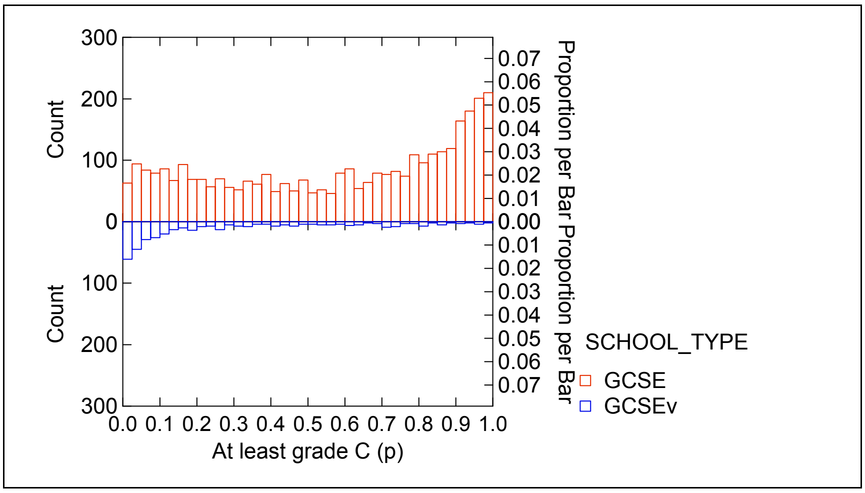


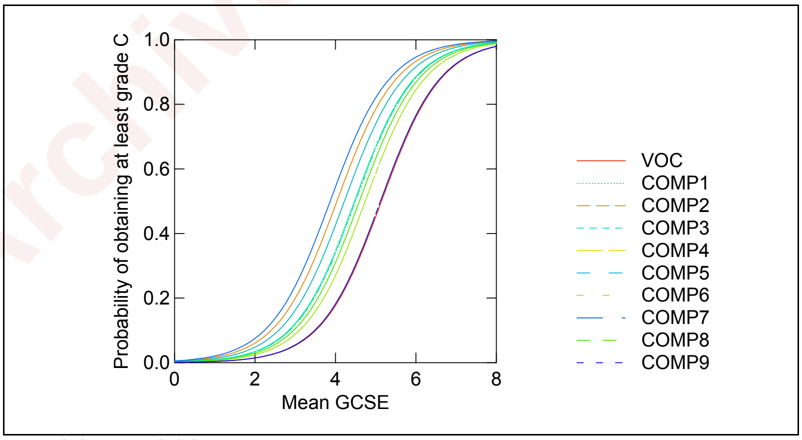
Figure 11: Predicted school effects by type of school (at least grade C, 2004)

Effect of the number of GCSEs

For Engineering Double Award, the effect of median number of GCSEs is not significant. The differences in grades between the GCSEv and the comparators do not seem to be explained by lack of teaching time.

Comparison of GCSEv and each comparator

The variation between the Engineering Double Award comparators is interesting. Figure 12 considers the comparators separately and it indicates that the probability curves at grade C are lower for Engineering Double Award than for all the comparators except for GCSE in Electronics. This was also the case at grade F, but at grade A the GCSEv had the lowest curve. It is worth noting that the differences in skills for the different Design and Technology subjects would be expected to lead to some variation between the curves.



VOC	GCSEv in Engineering Double Award
COMP 1	GCSE in D & T (Electronic Products)
COMP 2	GCSE in D & T (Food Technology)
COMP 3	GCSE in D & T (Graphic Products)
COMP 4	GCSE in D & T (Industrial Technology)
COMP 5	GCSE in D & T (Resistant Materials)
COMP 6	GCSE in D & T (Systems and Control)
COMP 7	GCSE in D & T (Textiles Technology)
COMP 8	GCSE in D & T (Product Design)
COMP 9	GCSE in Electronics

Figure 12: Probability of obtaining at least grade C and in Engineering Double Award and in each of its GCSE comparators

Effect of experience

In 2004 there was considerable variation between schools with regard to the implementation of the Engineering Double Award. These differences should have diminished over time if they were caused by lack of experience. Figure 13 shows that, for grade C students, there is evidence of increased experience being related to improved performance. However, this effect is slight and does not explain all the difference between GCSE and GCSEv.

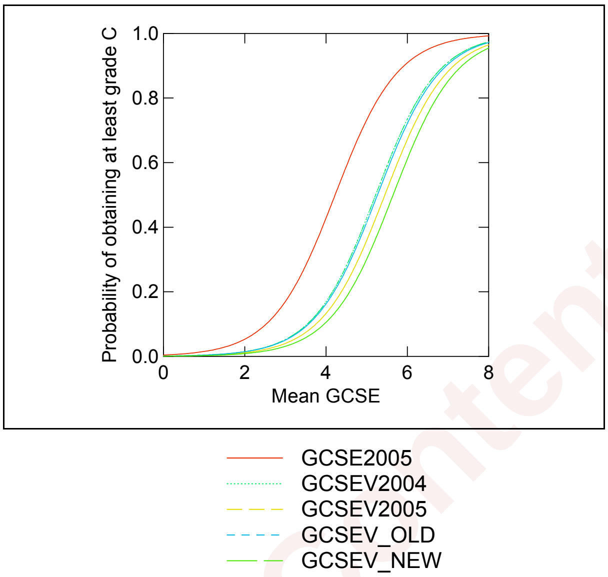


Figure 13: Probability of obtaining at least grade C in Engineering Double Award and in each of its GCSE comparators (including school experience)

Summary of results

This subject requires further monitoring. However, there are limits to what can be achieved by statistical methods and a judgement-based comparability study might be appropriate.

The OFSTED report (OFSTED, 2004) concluded that:

> 'Students' achievement is good or better in three fifths of lessons, which is slightly better than the national picture in Design and Technology (D&T) in 2002/03. However, achievement is unsatisfactory in one lesson in seven, twice as many as in D&T as a whole.'

It also concluded that:

> 'Students' capability in graphics and designing, including the use of computer-aided designing, is sound to good, similar to that in D&T in most schools except that fewer students achieved very high levels.'

and

> 'Higher-attaining students in some courses taught jointly with post-16 providers, however, are often not adequately stretched by the craft teaching.'

The findings in this section suggest that there could be differences in performance for the GCSEv although the OFSTED report suggests that the differences could be at least partly explained by the teaching quality.

3.3 Manufacturing Double Award

Introduction

Only a summary of the results for Manufacturing Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

There was a total of 3,946 students taking Manufacturing Double Award in 2004. These students were grouped in 195 schools. It should be noted that small entry subjects can cause problems at awarding meetings because of the lack of evidence for particular marks for the judgemental aspects and the low number of students for the statistical evidence.

In 2005, there were 4,327 students taking this GCSEv, grouped in 251 schools. There was an increase in both the number of students taking this subject (10%) and in the number of centres (29%) offering it. The entries per awarding body are shown in Table 6.

Table 6: Manufacturing Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	1,787	2,123
OCR	2,159	2,204
Total entry	3,946	4,327

Eight GCSE subjects were considered as comparators for Manufacturing Double Award. The entry figures, displayed in Table 7, show a drop of 8% in the entry: 422,199 students took one of these courses in 2005 compared to the 458,089 in 2004. Manufacturing Double Award was taken by around 1% of the total Design and Technology entry.

Table 7: GCSE comparators

Subject	Entry 2004	Entry 2005
Design & Technology - Electronic Products	18,490	17,119
Design & Technology - Food Technology	114,460	108,247
Design & Technology - Graphic Products	114,279	100,371
Design & Technology - Industrial Technology	1,290	1,270
Design & Technology - Resistant Materials	106,846	99,054
Design & Technology - Systems and Control	27,370	22,461
Design & Technology - Textiles Technology	61,771	57,470
Design & Technology - Product Design	13,583	16,207
Total Entry	458,089	422,199

The grade distributions for Manufacturing Double Award and its comparators indicated that in both 2004 and 2005 the achievements of GCSEv students tended to be poorer than those of the comparators. Very few students obtained grade A* and nearly 10% were unclassified. However, the percentages of higher grades (A to C) increased in 2005 for the GCSEv (see Figure 14 for 2004).

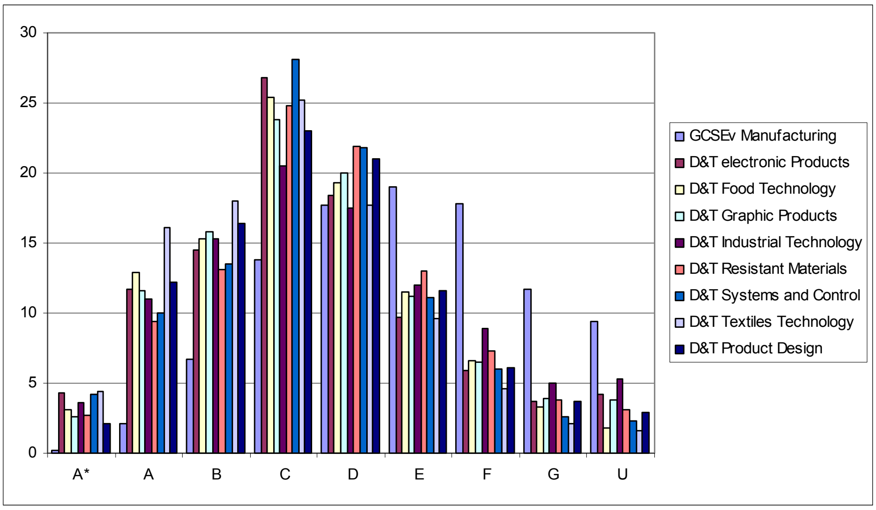


Figure 14: Grade comparison among Manufacturing Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

For all three grades that could be modelled the probability of obtaining at least the grade under consideration was higher for the comparators than for the GCSEv (see Figure 15 for the probability of obtaining at least grade C).

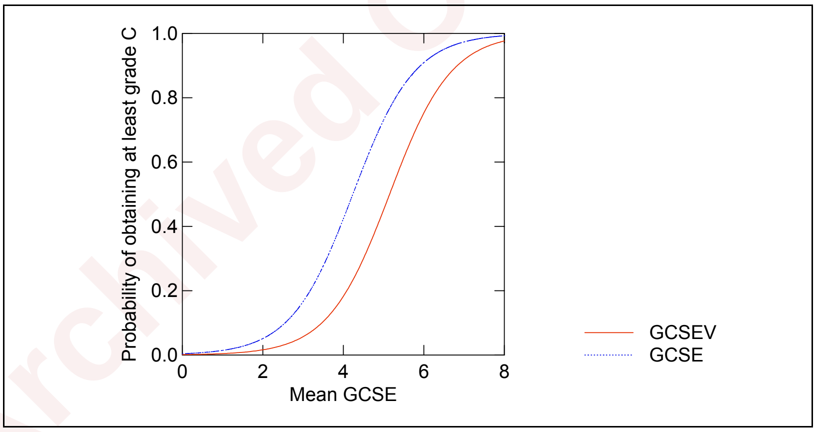


Figure 15: Probability of obtaining at least grade C in Manufacturing Double Award and in its GCSE comparators in 2004

In addition, it was found that, for any given mean GCSE score, the probability of obtaining at least grades A and C was almost the same in both years. It seems that, even with one year more of experience in the implementation of the subject, the students taking Manufacturing Double Award have not improved their performance. For grade F the pattern differed and, in 2005, the probability of obtaining at least grade F in Manufacturing Double Award decreased.

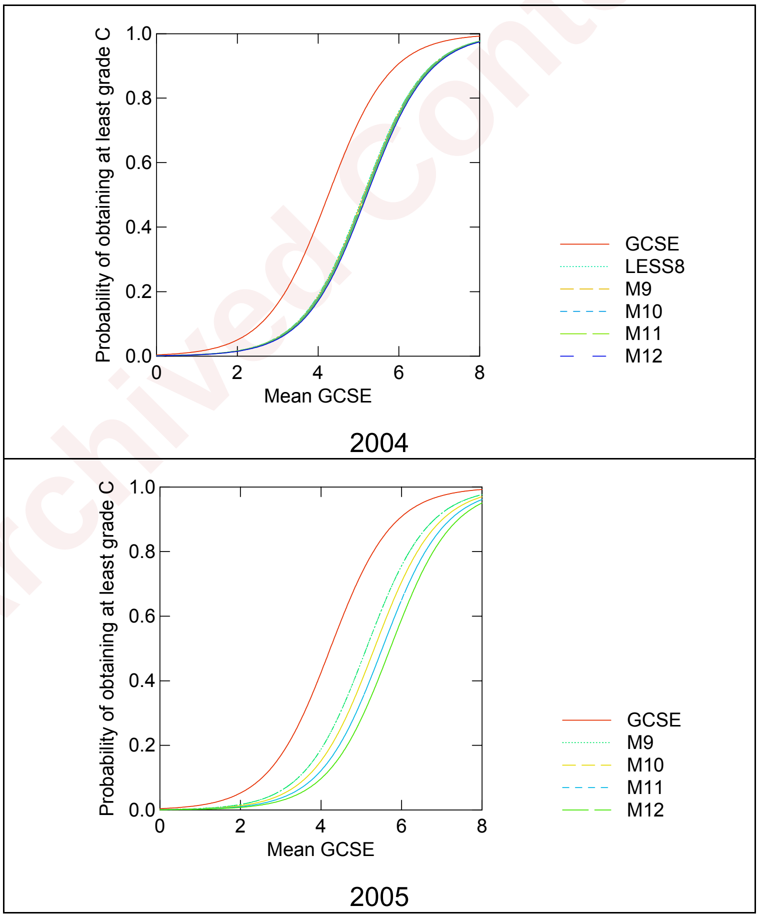
Table 8 displays the expected probabilities of obtaining at least grade A, grade C and grade F for students averaging grades A, C and F, respectively, in their other GCSE examinations in each centre (centres for GCSEvs and centres for GCSEs). From Table 8, in 2005 a student averaging grade A in his/her GCSEs had a probability of obtaining at least grade A of 0.23 in the GCSEv or 0.55 in the GCSEs. For grade C, students averaging a C in their GCSEs have a probability of 0.27 in the GCSEv and 0.58 in the GCSEs.

Table 8: Predicted probabilities of obtaining at least grade A, grade C and grade F

		2004		2005	
		Centres for GCSEv	**Centres for GCSE**	**Centres for GCSEv**	**Centres for GCSE**
At least a grade A	Mean	0.19	0.57	0.23	0.55
	SD	0.26	0.35	0.30	0.35
	Reference Interval	(0.001,0.850)	(0.023,0.989)	(0.001,0.929)	(0.019,0.989)
At least a grade C	Mean	0.25	0.60	0.27	0.58
	SD	0.28	0.31	0.30	0.31
	Reference Interval	(0.006,0.902)	(0.065,0.978)	(0.006,0.859)	(0.063,0.974)
At least a grade F	Mean	0.29	0.64	0.30	0.64
	SD	0.24	0.27	0.26	0.26
	Reference Interval	(0.013,0.764)	(0.114,0.968)	(0.009,0.815)	(0.147,0.961)

Effect of the number of GCSEs

In 2004, the median number of GCSEs was not a significant predictor of the performance in the GCSEvs so there was no evidence of the lack of time dedicated to the subject.



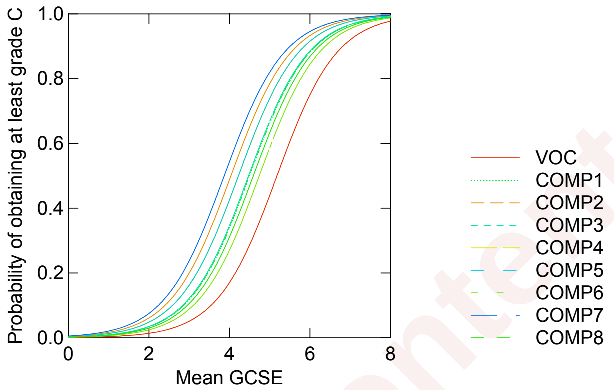
LESS8 – median number of GCSEs lower or equal to 8
M9 – median number of GCSEs equal to 9
M10 – median number of GCSEs equal to 10
M11 – median number of GCSEs equal to 11
M12 – median number of GCSEs higher or equal to 12

Figure 16: Probability of obtaining at least grade C in Manufacturing Double Award by the median number of GCSEs in the school

However, in 2005, for at least grade C (Figure 16) and at least grade F, the probability of obtaining at least those grades decreased with increasing median number of GCSEs in the centre. It seems that for Manufacturing Double Award centres were not allowing enough teaching time in 2005.

Comparison of GCSEv and each comparator

When this GCSEv is compared with each of the comparators individually, it is clear that the probability curves are lower than all the other comparators in both years of the study (see Figure 17 for at least grade C). This means that there is a particular issue with this GCSEv.



VOC	GCSEv in Manufacturing Double Award
COMP 1	GCSE in D & T (Electronic Products)
COMP 2	GCSE in D & T (Food Technology)
COMP 3	GCSE in D & T (Graphic Products)
COMP 4	GCSE in D & T (Industrial Technology)
COMP 5	GCSE in D & T (Resistant Materials)
COMP 6	GCSE in D & T (Systems and Control)
COMP 7	GCSE in D & T (Textiles Technology)
COMP 8	GCSE in D & T (Product Design)

Figure 17: Probability of obtaining at least grade C in Manufacturing Double Award and in each of its GCSE comparators

Effect of experience

Experience with the GCSEv implementation and teaching seemed to make a large difference at grade A but not at the other grades (see Figure 18).

Summary of results

The conclusions of the OFSTED report were that high achievement in Manufacturing Double Award is less common than in the more established Design and Technology courses (OFSTED, 2004). This is consistent with the findings of this report. For example OFSTED pointed out that: the designing of products for manufacture is uneven; students are expected to develop a quantitative approach to designing, within which mathematics and science should be applied as appropriate; this is underdeveloped in most courses; only rarely is the manufacturing activity adequately recorded, such as through use of video or digital photography, to enrich students' portfolios.

This would seem to support the proposition that attainment in this subject area was lower than expected. It would be interesting to monitor the subject to see if greater experience with the specifications leads to improved performance.

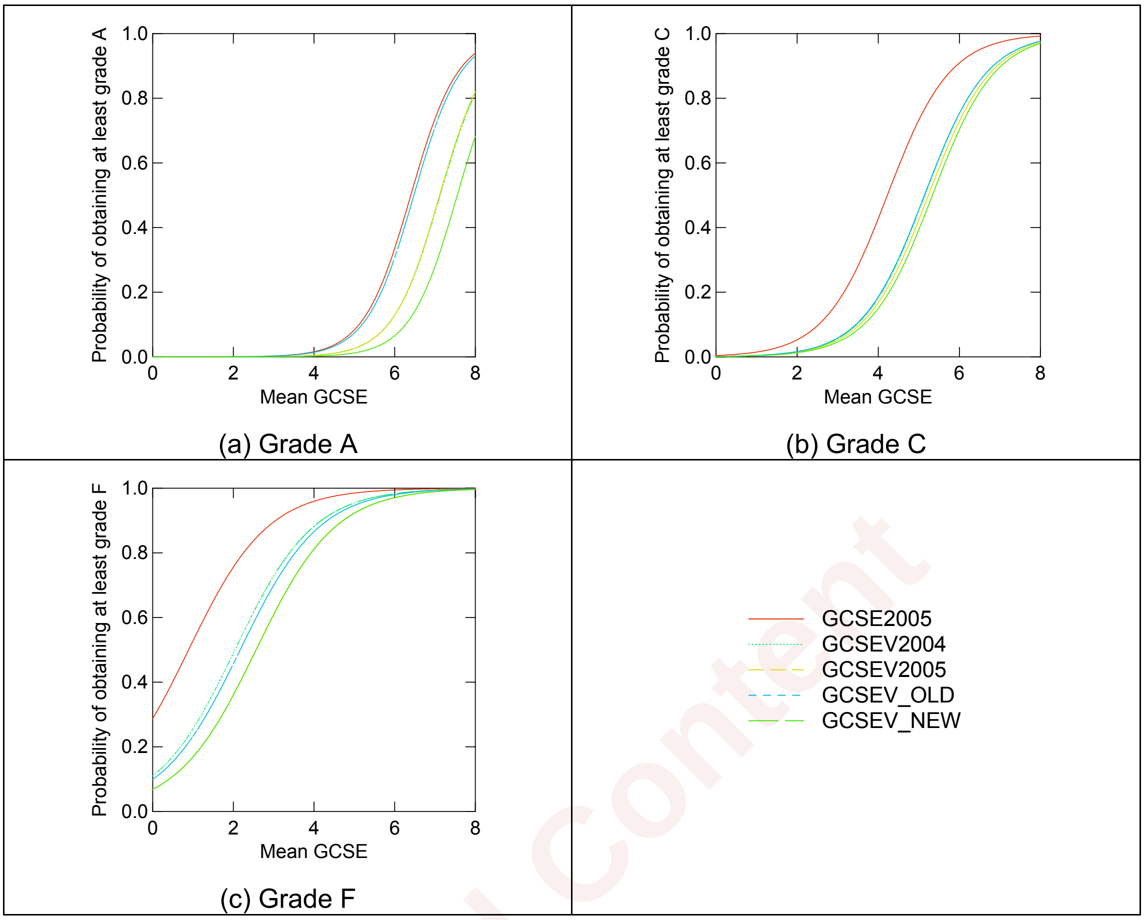


Figure 18: Probability of obtaining at least grade A, C and F in Manufacturing Double Award and in each of its GCSE comparators (including school experience)

3.4 Health and Social Care Double Award

Introduction

Only a summary of the results for Health and Social Care Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

Health and Social Care Double Award is a particularly interesting subject. There is no obvious list of comparators and the comparators chosen are particularly diverse (see Appendix B).

In 2004, there was a total of 16,845 students taking Health and Social Care Double Award. These students were grouped in 844 schools. The entries per awarding body are displayed in Table 9.

In 2005, there was a total of 22,946 students taking this GCSEv; this shows an increase of 36% in the entries. In 2005 there were 265 more schools offering Health and Social Care Double Award than in 2004.

Table 9: Health and Social Care Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	6,722	8,950
OCR	6,144	9,253
AQA	3,979	4,743
Total Entry	16,845	22,946

The entry figures for the comparators are displayed in Table 10. The number of students taking any of the comparators increased, going from 166,161 students in 2004 to 175,534 students in 2005. The comparators with bigger increases in their entries are Biology, Science (Single Award), Sociology and Psychology. Home Economics and Social Science entries decreased. It might be possible that the new GCSEv qualification is capturing students that traditionally were taking these subjects.

Table 10: GCSE comparators

Subject	Entry 2004	Entry 2005
Home Economics - Child Development	36,490	35,766
Home Economics - Food and Nutrition	13,958	13,570
Human Physiology and Health	340	322
Biology	43,665	47,950
Human Biology	673	699
Social Science	1,415	1,289
Sociology	13,446	14,670
Psychology	1,296	2,026
Science (Single Award)	54,583	59,242
Total Entry	166,161	175,534

Figure 19 shows the grade distributions for the GCSEv and for its GCSE comparators in 2004. Although the percentage of students obtaining grade A* is smaller for the GCSEv than for the GCSE comparators, for some of the GCSE subjects, such as GCSE in Human Physiology and Health or GCSE in Science (Single Award), the percentage of As and Bs is smaller than in Health and Social Care Double Award. At the bottom of the grade range, GCSE in Science and GCSE in Human Physiology and Health have a larger percentage of grades F and G than the GCSEv under consideration. The percentages of high grades (A* to C) increased in 2005 for the GCSEv. In particular, the percentage of students obtaining grade A increased by two points. To summarise, in both 2004 and 2005 the performance in the GCSEv is poor compared to the performance in almost all the GCSE comparators.

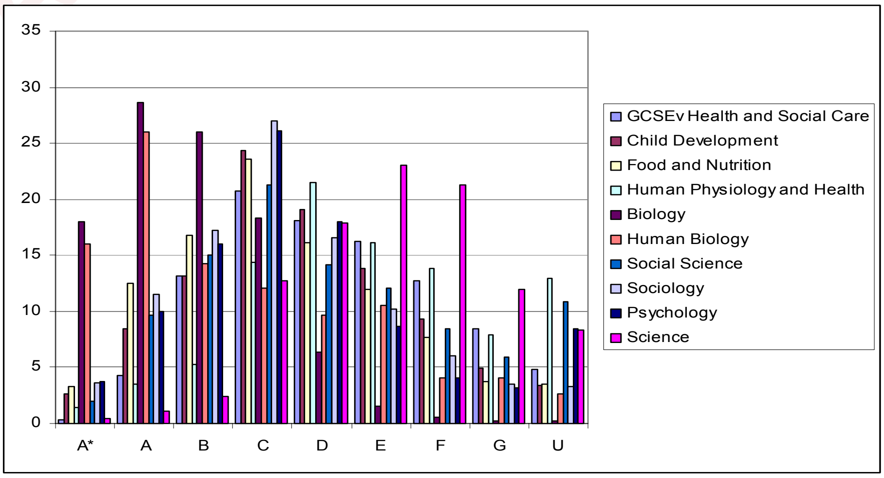


Figure 19: Grade comparison among Health and Social Care Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

Students with the same concurrent attainment taking the Health and Social Care Double Award are more likely to obtain at least grade A and at least grade C than those taking the GCSE comparators (see Figure 20 for grade C).

For any given mean GCSE score, the probability of obtaining at least grades A and C in the GCSEv is almost the same in 2004 and 2005. For grade F the pattern is a little bit different and in 2005 the probability of obtaining at least grade F in Health and Social Care Double Award decreased.

For Health and Social Care Double Award there is evidence of considerable variation between schools after allowing for concurrent attainment (see model estimates in Appendices E and F).

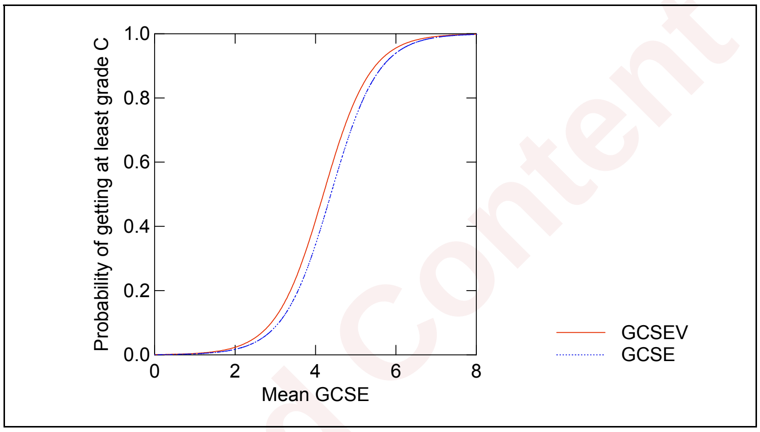


Figure 20: Probability of obtaining at least grade C in Health and Social Care Double Award and in its GCSE comparators in 2004

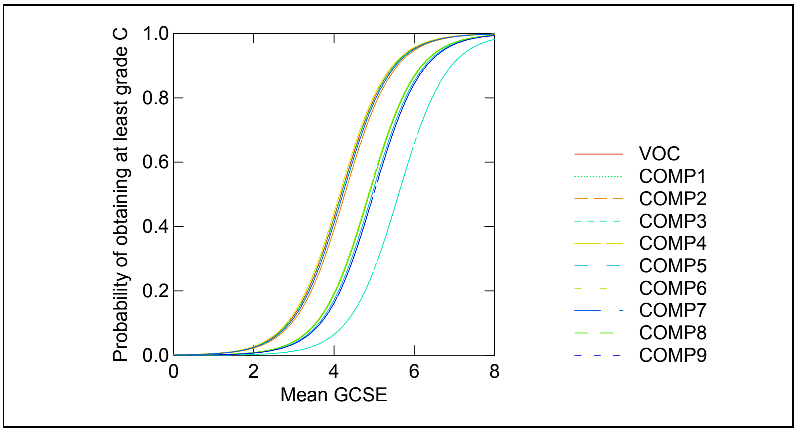
Effect of the number of GCSEs

Except for grade A in 2005, there was no evidence of decreasing performance with increasing median number of GCSEs.

Comparison of GCSEv and each comparator

When Health and Social Care Double Award was compared with its comparators it was found to be graded very similarly to some of the GCSE comparators and students were awarded higher grades than in certain GCSE subjects. For example, in 2004 only GCSE in Home Economics (Child Development) has a higher probability of obtaining at least grade A after adjusting for concurrent attainment.

For at least grades A, C and F, Health and Social Care Double Award is in the 'average' range of performance (see Figure 21 for grade C).



VOC	GCSEv in Health and Social Care Double Award
COMP 1	GCSE in Home Economics (Child Development)
COMP 2	GCSE in Home Economics (Food and Nutrition)
COMP 3	GCSE in Human Physiology and Health
COMP 4	GCSE in Biology
COMP 5	GCSE in Human Biology
COMP 6	GCSE in Social Science
COMP 7	GCSE in Sociology
COMP 8	GCSE in Psychology
COMP 9	GCSE in Science: Single Award

Figure 21: Probability of obtaining at least grade C in Health and Social Care Double Award and in each of its GCSE comparators

Summary of results

The combined conclusion from all these analyses is that although there are differences between Health and Social Care Double Award and its comparators, these differences are within the range of variation between the comparators. There is no evidence that this subject should be a particular area of concern.

3.5 Leisure and Tourism Double Award

Introduction

Only a summary of the results for Leisure and Tourism Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

There was a total of 12,285 students taking Leisure and Tourism Double Award in 2004. These students were grouped in 776 schools.

In 2005, the number of students was 16,073, showing an increase of 31% in the entries. These students were grouped in 1,002 schools, an average of 16 students per school. In 2005 there were 226 more schools offering Leisure and Tourism Double Award than in 2004. The entries per awarding body for both years are displayed in Table 11.

Table 11: Leisure and Tourism Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	5,671	7,836
OCR	3,222	3,902
AQA	3,392	4,335
Total Entry	12,285	16,073

Six GCSE subjects that are close to this GCSEv in contents and skills were considered as comparators. The entry figures for these comparators are displayed in Table 12. The number of students taking any of the comparators decreased from 290,817 students in 2004 to 282,508 students in 2005. The comparators with bigger decreases in their entries are Travel and Tourism, Business Studies and Geography. It might be possible that the new GCSEv qualification is capturing students that traditionally were taking these subjects.

Table 12: GCSE comparators

Subject	Entry 2004	Entry 2005
Travel and Tourism	1,712	1,388
Business and Communications Systems	25,754	28,227
Business Studies	66,269	63,579
Accounting	139	288
Business Studies and Economics	4,124	4,119
Geography	192,819	184,907
Total entry	290,817	282,508

The percentages of students obtaining grades A* to C is smaller for the GCSEv than the percentages in any of the GCSE comparators, showing that students tend to be awarded lower grades in the GCSEv subject than in traditional GCSEs (Figure 22).

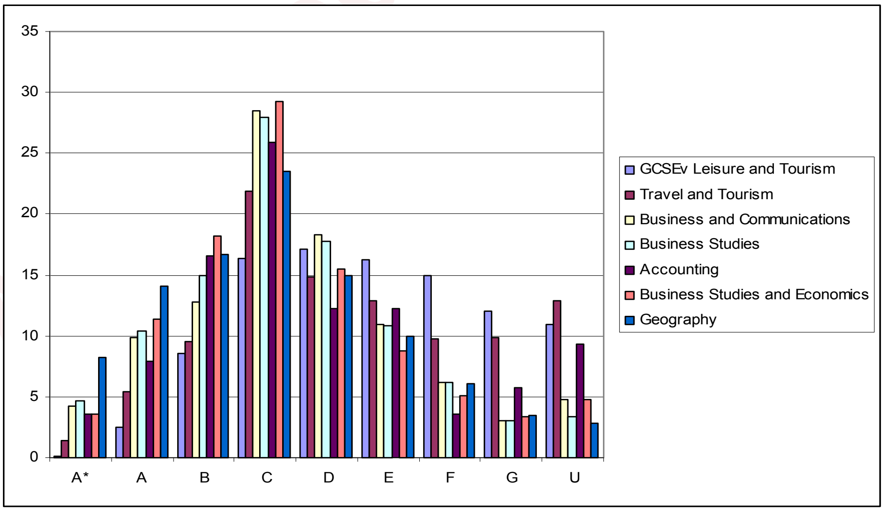


Figure 22: Grade comparison among Leisure and Tourism Double Award and the GCSE comparators in 2004 (% in grade)

According to the OFSTED report (OFSTED, 2004), in the great majority of schools, Leisure and Tourism Double Award is taken by a small number of mainly lower-attaining students. It is offered across a wider range of attainment in only a quarter of schools. This could be one of the reasons for the low percentages of students obtaining high grades. Also, standards of achievement and teacher expectations range from average to low, with few students expected to attain more than a C grade at best. The percentages of grade A* increased in 2005 for this GCSEv (from 0.16% in 2004 to 0.24% in 2005). However, the percentage of students obtaining grade A decreased from 2.47 to 2.35. The percentages of grades B to D increased in 2005.

Multilevel model for concurrent attainment

When all comparators are considered together, the probabilities of obtaining a particular grade are very similar for the GCSEv and the GCSE comparators (Figure 23 shows the results for grade C).

For any given mean GCSE score below 7, the probability of obtaining at least grades A and C is lower in 2005 than in 2004. However, for the most able students the probability is the same in both years.

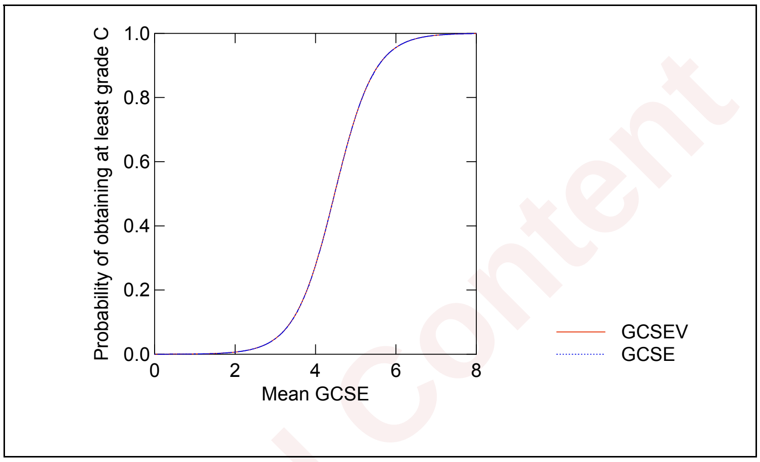


Figure 23: Probability of obtaining at least grade C in Leisure and Tourism Double Award and in its GCSE comparators in 2004

Effect of the number of GCSEs

For Leisure and Tourism Double Award, the median number of GCSEs in the schools had no effect on the probability of obtaining a certain grade.

Comparison of GCSEv and each comparator

When the individual comparators were considered, the curve for the GCSEv was within the range of the comparators.

Summary of results

The results of these analyses indicate that there is no statistical evidence of a lack of comparability for this subject.

3.6 Applied Business Double Award

Introduction

Only a summary of the results for Applied Business Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

This GCSEv is quite popular with a total of 15,028 students in 2004 (grouped in 604 schools) and 18,800 students in 2005 (grouped in 779 schools) taking it. There was an increase in both the number of students taking this GCSEv (25%) and in the number of centres offering it (29%). The entries, fairly evenly distributed per awarding body, are shown in Table 13.

Table 13: Applied Business Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	5,202	6,040
OCR	5,651	6,847
AQA	4,175	5,913
Total Entry	15,028	18,800

The most obvious comparator for this GCSEv is the GCSE in Business Studies. In addition to this GCSE, courses involving Accounting and Economics were included in the list of comparators. This gave a total of five traditional GCSEs as comparators for this subject. The entry figures for the five comparators are displayed in Table 14. The total number in 2005 is very similar to the figure in 2004 although the distribution among the subjects is different (for example, there was a decrease in the entry of Business Studies and of Economics but an increase in Business and Communications Systems). In 2005 Applied Business Double Award was taken by around 16% of the total Business and Economics entry.

Table 14: GCSE comparators

Subject	Entry 2004	Entry 2005
Business and Communications Systems	25,754	28,227
Business Studies	66,269	63,579
Accounting	139	288
Business Studies and Economics	4,124	4,119
Economics	3,188	2,793
Total entry	99,474	99,006

The distribution of grades for Applied Business Double Award and its comparators (Figure 24) indicates that the results were worse for this subject compared with the comparators.

There was a slight change in the grade distribution in 2005. The percentages of students obtaining grades A*-C in the GCSEv increased around 4% (44% students obtained grades A*-C in 2004 and 48% students in 2005). In particular, the percentage of grade A* was 1.2% in 2005 compared to 0.8% in 2004.

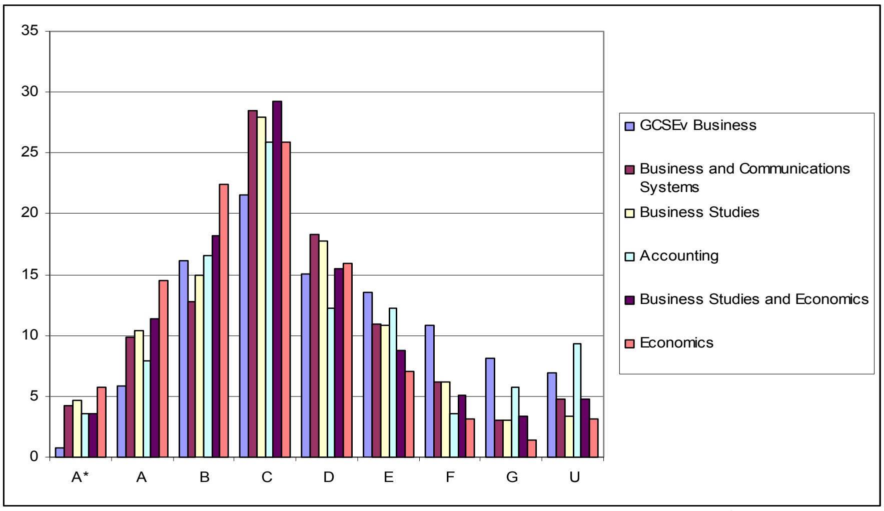


Figure 24: Grade comparison among Applied Business Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

The multilevel analysis indicates the importance of considering concurrent attainment. For both years there were no significant differences in the probabilities of obtaining at least grade A and at least grade C and only a small difference at grade F. This is illustrated in Figure 25 where the two curves for grade C in 2004 are almost identical (same pattern can be observed for the other grades in the figures in Appendix G). There was an exception: in 2005, students taking Applied Business Double Award had a lower probability of obtaining grade A*.

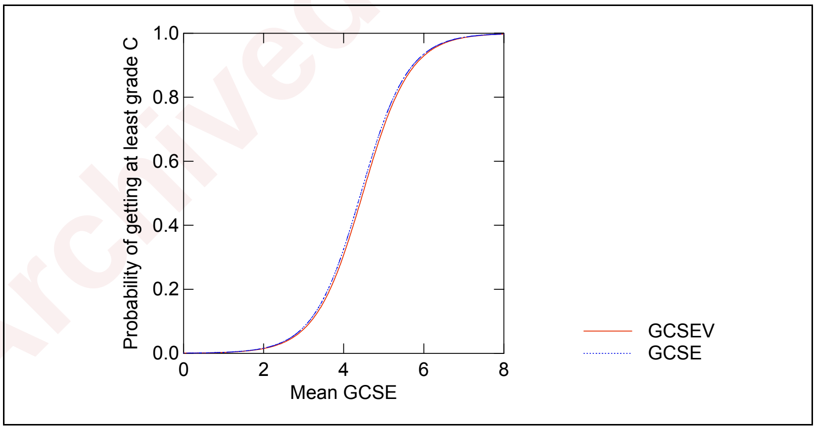


Figure 25: Probability of obtaining at least grade C in Applied Business Double Award and in its GCSE comparators in 2004

Effect of the number of GCSEs

In both years of the study, the probability of obtaining any grade in the GCSEv is almost the same for all values of the median number of GCSEs in the school. It seems that, for Applied Business Double Award, centres were allowing enough teaching time.

Comparison of GCSEv and each comparator

The data were analysed with the comparators identified separately. The curves for Applied Business Double Award were within the ranges for all comparators and very similar to the ones for GCSE in Business Studies.

Summary of results

There is no evidence of lack of comparability using statistical methods for this subject.

3.7 Applied Science Double Award

Introduction

Only a summary of the results for Applied Science Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

There was a total of 8,555 students, grouped in 236 schools, taking Applied Science Double Award in 2004.

In 2005 there was a total of 17,072 students taking the subject. This shows that the entries have almost doubled for this GCSEv. In 2005 the number of schools offering Applied Science Double Award has also doubled. The entries per awarding body are displayed in Table 15.

Table 15: Applied Science Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	1,305	2,960
OCR	3,839	6,196
AQA	3,411	7,916
Total Entry	8,555	17,072

Seven traditional GCSEs were considered as comparators for this subject. The total number of students taking any of the comparators in 2004 was 666,740. The entry figures are displayed in Table 16. The number of students taking any of the comparators decreased to 655,966 students in 2005. There were slight increases in the single science specifications and there was a big decrease in the entry for GCSE Double Science (482,072 students in 2004 compared with 455,511 in 2005). It might be possible that the new GCSEv is capturing students that were traditionally taking this subject.

The grade distributions for GCSEvs indicate that performance in this qualification is again weaker than in the more established double award. However, it is interesting to compare this with the single award science results. The performance level tends to be higher in the middle of the grade range. It would be interesting to investigate which science was used by the school's previous cohort. Figure 26 displays the grade distribution for the GCSEv and its comparators in 2004.

Table 16: GCSE comparators

Subject	Entry 2004	Entry 2005
Science: Single Award	54,878	59,242
Chemistry	42,886	46,272
Physics	42,248	45,773
Biology	43,665	47,950
Human Biology	673	699
Environmental Science	318	519
Science: Double Award	482,072	455,511
Total entry	666,740	655,966

The percentages of grade A* are almost constant for Applied Science Double Award (0.11 in 2004 to 0.12 in 2005). However, the percentages of grades B to D increased in 2005, with the bigger increase in students obtaining grade C (5% higher).

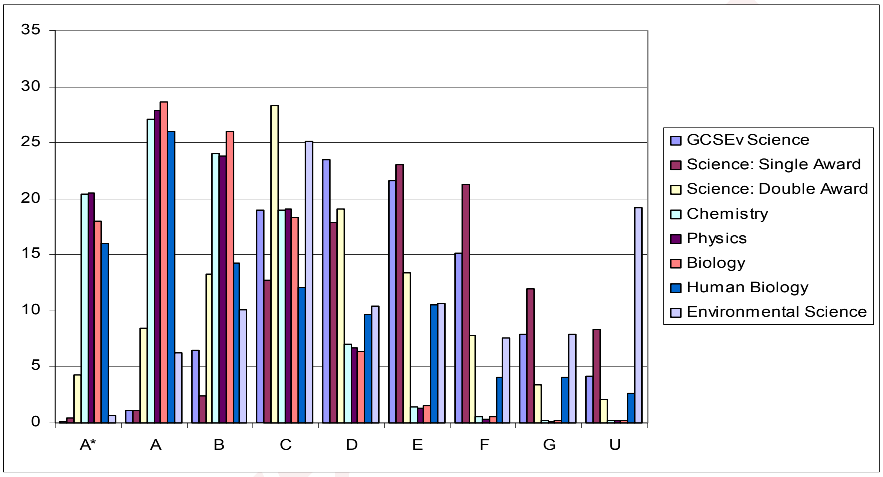


Figure 26: Grade comparison among Applied Science Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

When concurrent attainment is accounted for there is no real difference in the probability of obtaining a GCSEv grade compared with a GCSE grade. However, in general, the grade distributions for GCSEvs indicate that performance in this qualification is weaker than in the more established double award. This is consistent with their attainment in other GCSEs.

Effect of the number of GCSEs

For this subject the effect of median GCSE was non-significant. A plausible explanation for this is that it is used as an alternative to the GCSE Science Double Award. This means that sufficient time was allowed for this GCSEv as we might expect it to fill the same slot in the school timetable.

Comparison of GCSEv and each comparator

When the comparators are modelled separately it is clear that the GCSEv curve is within the range of curves generated for each of the comparators.

Summary of results

In summary, there is no ground for serious concern about comparability for GCSEv Applied Science.

3.8 Applied Art and Design Double Award

Introduction

The final GCSEv to be considered is Applied Art and Design. From the statistical comparability point of view, this subject is more difficult to compare validly i.e there is less relationship with ability measured by mean GCSE.

Only a summary of the results for Applied Art and Design Double Award is presented in this section. Although the analyses for Applied ICT Double Award were replicated entirely, only significant findings will be discussed here. Complete results can be found in Appendix G and more extensive explanation of the interpretation of the figures and tables can be found in section 3.1.

The entry for this GCSEv in 2004 was small. There were only 3,863 students, grouped in 262 schools, taking it. This gives an average entry of approximately 15 students per school.

There was a total of 5,536 students taking Applied Art and Design Double Award in 2005, grouped in 362 schools. There was an increase in both the number of students taking this GCSEv (43%) and in the number of centres offering it (38%).

Table 17: Applied Art and Design Double Award

Awarding Body	Entry 2004	Entry 2005
EdExcel	1,559	2,219
OCR	1,073	1,271
AQA	1,231	2,046
Total Entry	3,863	5,536

The six GCSEs in Art and Design offered by the three awarding bodies were considered as comparators for this subject. The entry figures for the comparators are displayed in Table 18.

Table 18: GCSE comparators

Subject	Entry 2004	Entry 2005
Art and Design	124,831	122,856
Art and Design (Fine Art)	48,530	50,849
Art and Design (Graphic Design)	4,811	4,626
Art and Design (Photography)	2,529	3,116
Art and Design (Textiles)	6,000	6,540
Art and Design (3D Studies)	3,983	3,668
Total Entry	190,684	191,655

The total number of entries for the GCSE Art specifications in 2005 is very similar to the figure in 2004 although the distribution among the subjects is different (for example, there was a decrease in the entry of Art and Design, Graphic Design but an increase in Photography and Fine Art). Applied Art and Design Double Award was taken by around 3% of the total Art and Design entry.

Figure 27 shows the grade distribution for the GCSEv and for its GCSE comparators in 2004. For this subject the percentage of A* is larger than for any of the other GCSEvs. However, this percentage is still smaller than the percentages of grades A* for any of the GCSEs in Art and Design.

The percentages of students obtaining grades A*-C in the GCSEv increased by around 6 percentage points (48% students obtained grades A*-C in 2004 and 54% students in 2005).

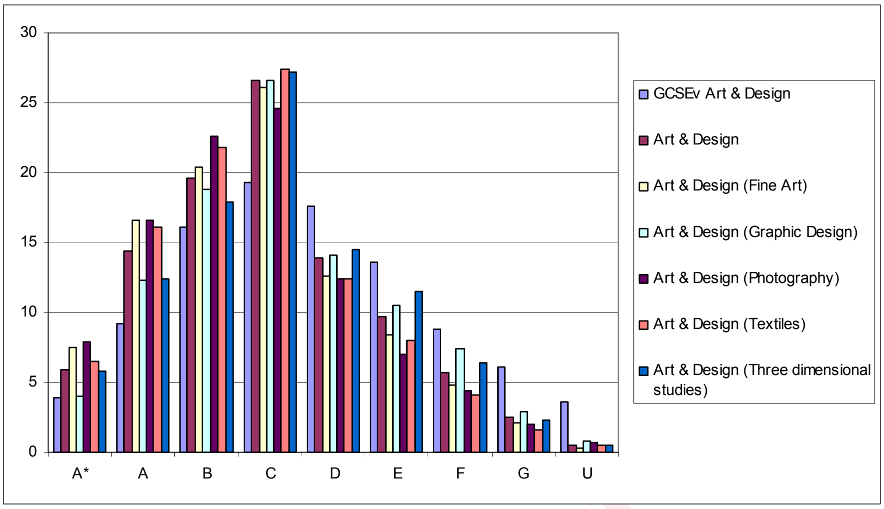


Figure 27: Grade comparison among Applied Art and Design Double Award and the GCSE comparators in 2004 (% in grade)

Multilevel model for concurrent attainment

The probabilities of obtaining grades A*, and at least grade A, in the GCSEv are higher than for the GCSE after adjusting for concurrent attainment (see Figure 28 for at least grade A).

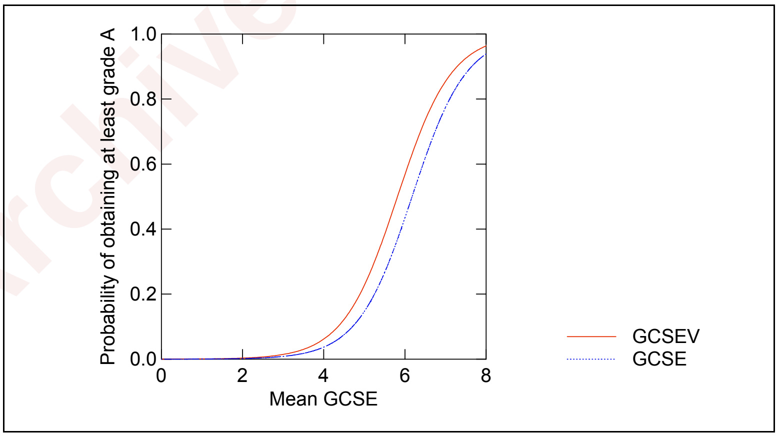


Figure 28: Probability of obtaining at least grade A in Applied Art and Design Double Award and in its GCSE comparators in 2004

In 2005 the probabilities of obtaining grades A* and at least grade A declined with respect to 2004.

In both years of the study, there is more variation between centres using GCSEvs than for centres using GCSEs. However, in 2005, and in particular for grade A, the magnitude of the variation for centres offering the GCSEv decreased considerably compared to the variation in 2004.

Effect of the number of GCSEs

There were no significant effects of the median number of GCSE per school.

Summary of results

The varied nature of the results and the special nature of art make it hard to draw conclusions about this subject.

It is possible that selection effects influence these results. A student who is good at other GCSEs could choose or be channelled into single award Art while students who are poor performers at GCSE and are good at Art might find themselves taking a double award in Art.

4 Conclusions and discussion

This report investigated, with statistical modelling, the differences in performance between GCSEvs and traditional GCSEs and the possible causes of such differences. Of the eight GCSEvs, four (Health and Social Care, Applied Business, Leisure and Tourism and Applied Science) did not produce any results that cast serious doubts on their comparability with GCSEs. For one of the GCSEvs, Applied Art and Design, the students were more likely to obtain grades A and A* for a given level of mean GCSE performance. However, there is a plausible selection mechanism that might account for these results. Art and Design uses some skills that are not widely used in other GCSEs and it is reasonable to assume that talented artists who are weaker academically might opt for a double award in Art. Finally, there are three GCSEvs where the initial analysis indicated that there might be a problem with comparability. These subjects are Applied ICT, Engineering and Manufacturing. However, OFSTED identified some problems with the introduction and implementation of the GCSEvs which could provide an explanation for these differences.

The OFSTED research suggested that the teaching of the GCSEvs tended to be less good than other GCSEs in many centres but that there were some examples of very good teaching. In addition, OFSTED suggested that some centres did not allow sufficient time for these double courses. The analyses carried out in this report are consistent with the OFSTED findings. The results from this report suggest that some centres did not allow sufficient time for these double award courses, in particular in Applied ICT and Manufacturing. This could be a reason for the lower attainment in those GCSEvs.

Using the preferred technique of statistical comparability based on concurrent attainment, the issue of comparability for GCSEvs is very different from that found by less complex analyses. In particular, there are sound reasons for preferring analyses using concurrent attainment (GCSE results) instead of prior attainment (Key Stage 2 or 3 scores). The analyses used in this study showed that mean GCSE, a concurrent measure of attainment, is a better predictor than the earlier Key Stage data. Concurrent measures should be used when they are available providing they have meaningful measurement characteristics. Using prior attainment in this study seems to generate differences between specifications that do not occur when concurrent attainment is used. One counter-argument against the use of prior attainment data that can be proposed is that if there is a difference between specifications when prior attainment is used then this has important implications for value added calculations. In fact choosing GCSEv instead of GCSE does not have implications if there is no difference between GCSEvs and the comparators when concurrent attainment is considered. Whilst the GCSEv students may have underperformed relative to their prior attainment in relation to other students, the fact that there is no difference using concurrent performances means they have underperformed across all the subjects that they have studied. Choosing a GCSE instead of a GCSEv would not improve the value added.

One argument that is sometimes made is that allowances should be made when setting grade boundaries for lack of experience with a new specification. However, this is a complex issue because it depends on the purpose of the examination and the nature of the problem (e.g. dealing with a misunderstanding about the rubric). For vocational qualifications this is a particular issue because of the concept of ‘parity of esteem’. If this is perceived to be lacking, a damaging feedback process can develop. If a subject is perceived to lack credibility then able students may opt to take other, more credible, subjects. This leads to the qualification being associated with weaker students. This can then reinforce the lack of credibility, making it even less likely that able students will choose to take the subject. Another problem with adjusting grade boundaries is that centres that start with the specification in the first year benefit but not those who start in subsequent years.

Less complex analyses identified that there were potential problems with GCSEvs. There is nothing intrinsically wrong with this since such methods are often used to screen the large number of qualifications used by awarding bodies. However, it is important to recognise the limitations of the methodology and that there can be multiple explanations for the results. Less complex analyses should not be used to make definitive statements about the comparability of examinations. Even sophisticated methods have their limitations. Statistical comparability can be used to test likely explanations for differences in the grade pattern. The effectiveness of this depends on the quality and the availability of variables in the analysis. However, ultimately, comparability has to be based on the judgements about the quality of the actual work of the students. This is particularly important in the early years of a qualification when the awarding process is one of standard setting rather than standard maintenance. Statistical evidence has an important role in identifying potential boundaries and indicating changes in entry. However, it is more than possible that greater familiarity with the specifications and greater experience of the teachers can lead to improved student learning. This means that more students might satisfy the criteria associated with the grade. The relationship between attainment in one subject and other subjects should not be assumed to be unchanging. This implies that there is a danger of statistical data being given unwarranted prominence in decisions about awarding.

If it is assumed that the lower performance on Applied ICT, Engineering and Manufacturing is the result of lack of experience with these new specifications, then there are two implications. First, the awarding bodies should be careful not to put too much emphasis on the statistical evidence generated in the first award when making subsequent boundary decisions. Secondly, if lack of experience is the cause of the differences, then they should diminish over time. However, given that the GCSEvs are two year courses, this might be a long process because it might take time to make organisational changes and for feedback to take effect.

For the qualifications in this report, it is important that Applied ICT, Engineering, and Manufacturing are monitored to see if the differences diminish as the experience of the requirements for these subjects increases. If there is still evidence of a problem, it might require further investigation using a judgement-based study involving inspection of the students' work. There may also be a case for providing more training to ensure the requirements of these courses are more fully understood.

5 References

Bell, J.F. (2001) *Visualising multilevel models: the initial analysis of data*. Paper presented at the Third International Conference on Multilevel Analysis, Amsterdam, The Netherlands. Available at http://www.cambridgeassessment.org.uk/research/confproceedingsetc/MA2001JB

Bell, J.F. (2003) Beyond the school gates: the influence of school neighbourhood on the relative progress of pupils. *Oxford Review of Education*, 29, 4, 485-502.

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, Cologne, Germany. Available at http://www.leeds.ac.uk/educol/documents/00001528.htm.

Bell, J.F. and Greatorex, J. (2000) *A review of research into levels, profiles and comparability*. QCA.

Clarke, A.D.B. (1978) Predicting Human Development: Problems, Evidence, Implications. *Bulletin of the British Psychological Society*, 31, 24-258.

Cook, R.D. and Weisberg, S. (1999) *Applied Regression including computing and graphics*. New York: John Wiley and Sons, Inc.

Dexter, T. and Massey, A. (2000) *Conceptual issues arising from a comparability study relating IGCSE grading standards with those of GCSE via a reference test using a multilevel model.* Paper prepared for the 22nd Biennial Conference of the Society for Multivariate Analysis in the Behavioural Sciences, London School of Economics.

Goldstein, H. and Cresswell, M. (1996) The comparability of different subjects in public examinations: a theoretical and practical critique. *Oxford Review of Education*, 22, 435-442.

Jesson, D. (2005) *Vocational GCSEs: Comparisons of outcomes with conventional GCSEs*. Communication to QCA.

Jones, B, Baird, J and Arlett, S (1997) *A Comparability Study in GCSE Art and Design (Unendorsed): A study based on the Summer 1996 examinations*. NEAB/Joint Forum for the GCSE and GCE.

Jones, B (ed.) (1997) *A review and evaluation of the methods used in the 1996 GCSE and GCE comparability studies, and the issues which this raises.* SCR/Joint Forum.

Jones, B. E. (1997) Comparing Examination Standards: is a purely statistical approach adequate? *Assessment in Education*, 4, 2, 249-262.

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

Office for Standards in Education (2003) *Developing new vocational pathways. Interim report on the introduction of new GCSEs*. Report HMI 1630.

Office for Standards in Education (2004) *Developing new vocational pathways. Final report on the introduction of new GCSEs.* Report HMI 2051.

Appendix A: GCSEvs

The regulatory authorities in consultation with NTOs (National Training Organisations), subject associations and other interested parties developed the subject criteria. Each subject consists of three common, compulsory, and normally equally weighted units. Assessment and awarding arrangements are governed by these national subject criteria which apply across all awarding bodies.

The qualification content is the same across awarding bodies. However, the awarding bodies have used the subject criteria to develop their own individual approaches to internal and external assessment, and guidance on delivery and assessment.

GCSEvs are designed to:

- replace Part One GNVQs;
- provide an introduction to a broad vocational area;
- enable progression to further education, training or employment;
- be available at Key Stage 4 and post-16.

The qualifications were offered in the following subjects: Applied Art and Design, Applied Business, Applied ICT, Applied Science, Engineering, Health and Social Care, Leisure and Tourism and Manufacturing. Although the content of the GCSEvs is quite similar to that for Part One GNVQs (albeit updated and reorganised), they are structurally and technically quite different from Part One GNVQs. However, the work a student produces towards a GCSEv could be re-assessed against the criteria for a six-unit GNVQ.

These qualifications are equivalent to two GCSEs (same size as Part One GNVQs) and are graded A*A* to GG, with a U (Unclassified). They cover both level 1 (foundation) and level 2 (intermediate), as does GCSE.

Although QCA does not prescribe teaching/contact hours, these GCSEs are double awards and it was suggested that schools and colleges might want to allow double the amount of time they allow for other GCSEs. They might also want to take into account that longer blocks of time (double or triple periods, for example) will allow for work placement, or other practical activities, such as visits, role plays, presentations by students and/or visitors to be undertaken successfully.

Some subjects relate closely to the national curriculum at Key Stage 4. This is particularly relevant to the decisions on the choice of comparators used in this report. For example, Manufacturing and Engineering meet the national curriculum programme of study for Design and Technology; Applied ICT meets the national curriculum programme of study for ICT, and Applied Science can be used to meet the statutory requirements for Science (although it does not meet the full national curriculum programme of study).

The summer 2004 examination was the first time that certificates in these subjects were offered.

Appendix B: GCSE comparators

For each GCSEv, traditional GCSE subjects that are close to them in content and skills are used as comparators.

Applied Art and Design Double Award

GCSE Art and Design
GCSE Art and Design (Fine Art)
GCSE Art and Design (Graphic Design)
GCSE Art and Design (Photography)
GCSE Art and Design (Textiles)
GCSE Art and Design (Three Dimensional Studies)

Applied Business Double Award

GCSE Business and Communications Systems
GCSE Accounting (AQA specification)
GCSE Business Studies and Economics Nuffield (EdExcel specification)
GCSE Economics
GCSE Business Studies

Engineering Double Award

GCSE Design and Technology (Electronic Products)
GCSE Design and Technology (Food Technology)
GCSE Design and Technology (Graphic Products)
GCSE Design and Technology (Industrial Technology)
GCSE Design and Technology (Resistant Materials)
GCSE Design and Technology (Systems and Control)
GCSE Design and Technology (Textiles Technology)
GCSE Design and Technology (Product Design)
GCSE Electronics

Health and Social Care Double Award

GCSE Home Economics (Child Development)
GCSE Home Economics (Food and Nutrition)
GCSE Human Physiology and Health
GCSE Biology
GCSE Human Biology
GCSE Social Science (AQA specification)
GCSE Sociology
GCSE Psychology
GCSE Science: Single Award A
GCSE Science: Single Award B (Staged Assessment)
GCSE Science: Single Award C (Salters)

Applied ICT Double Award

GCSE ICT

Leisure and Tourism Double Award

GCSE Travel and Tourism
GCSE Business and Communication Systems
GCSE Business Studies
GCSE Accounting (AQA specification)
GCSE Business Studies and Economics Nuffield (EdExcel specification)
GCSE Geography

Manufacturing Double Award

GCSE Design and Technology (Electronic Products)
GCSE Design and Technology (Food Technology)
GCSE Design and Technology (Graphic Products)
GCSE Design and Technology (Industrial Technology)
GCSE Design and Technology (Resistant Materials)
GCSE Design and Technology (Systems and Control)
GCSE Design and Technology (Textiles Technology)

Applied Science Double Award

GCSE Science: Single Award A
GCSE Science: Single Award B (Staged Assessment)
GCSE Science: Single Award C (Salters)
GCSE Science: Double Award A
GCSE Science: Double Award B (Staged Assessment)
GCSE Science: Double Award C (Salters)
GCSE Chemistry
GCSE Physics
GCSE Biology
GCSE Human Biology
GCSE Environmental Science

Appendix C: Technical Details

Appendix C covers some technical details of the analyses carried out in this report.

The first issue considered is why a series of logistic regression models were used rather than a single analysis for an ordinal variable like the examination grades.

Various models have been proposed for regression modelling with ordinal dependent variables. One of the commonest is the proportional odds model. For this type of model, the cumulative probabilities are used, for example, the probability of obtaining at least a grade C. In a proportional odds model, the shape of the relationship between the cumulative probability and the independent variable is assumed to be the same for each probability as in Figure C1 (for the cumulative logits, the relationship is assumed to be a series of parallel lines). In Figure C1, for an individual with a score of 35 on the independent variable, the probability of getting a grade A is effectively 0 and the probability of getting at least a grade F is approximately 0.75.

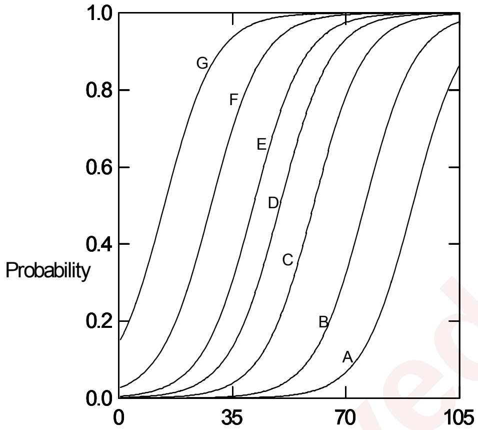


Figure C1: Probability curves resulting from fitting a proportional odds model

The problem with fitting a proportional odds model is the assumption that the curves are the same shape (experience suggests that this is not the case with examination data). This has to be verified and the easiest way of doing so is to carry out a series of separate logistic regressions at each grade boundary (e.g. modelling the probability of getting at least a grade D with the probability of getting a grade E or less). If the slope parameters were the same for each regression, then a proportional odds model could be fitted. Although a proportional odds model may be more efficient because fewer parameters are estimated, this advantage is minimal for the large data sets available for statistical comparability studies. It is for this reason that a series of separate logistic regressions can be carried out instead of a proportional odds model.

The use of separate binary logistic regression models does have some disadvantages. This approach can lead to final models with different sets of covariates for different grades making interpretation difficult (e.g. a sex difference at one grade but not at another, though this finding is inherently interesting). Categories at the ends of the scale may have very low or very high probabilities, and parameter estimates may not be statistically significant due to less power, that is, the ability of a statistical test to detect differences (usually this will not be a problem because the data sets used in comparability studies are large). The proportional odds model, when it fits, is superior because few parameters are fitted. This is better because the standard errors of these parameters are smaller. However, if the data set is large, the standard errors for the separate binary logistic models are likely to be small (meaning that there is sufficient power to detect small differences between syllabuses). In comparison with the proportional odds models the presentation of the results of binary logistic regressions is much simpler and more interpretable for less experienced users. Finally, it might not be possible to fit models for extremely small or large probabilities. However, if they were fitted, the parameters would have large standard errors and the results would not be significant.

Given that separate logistic regressions are used, the next stage was to explore the data to decide on plausible relationships.

The exploratory analysis (Bell, 2001) used interactive graphics and it is not really practicable to analyse large data sets with such techniques. Therefore, an approximate 2% sample of the total ICT entry in 2004 was extracted and analysed with the interactive graphic package ARC (Cook and Weisberg, 1999). There were two main reasons for doing this. First, the objective was to investigate the shape of the relationship between the probability of obtaining at least a grade X and the mean GCSE. Secondly, the plots would indicate whether the relationship was of the same shape for all specifications, that is, checking for interaction terms. Interactive graphics allow users to change plots quickly to explore features of the data. Only some illustrative plots have been included in this report.

The first objective was investigated by comparing the logistic fit with local log-likelihood smoothers (the equivalent for binary dependent variable to a lowess fit for a continuous variable). ARC was used to check all the relationships but for illustrative purposes only the plot for specification 1494 has been presented. The outcome variable '*probability of obtaining at least a grade C*' only takes the value 0 and 1 but and since this would lead to the points overlapping, jittering has been used to add a small amount of random scatter. It is clear from the plot (Figure C2) that the logistic regression fit and the smoothed line almost coincide. This provides evidence that the relationship is a simple linear one when the logistic transform is applied.

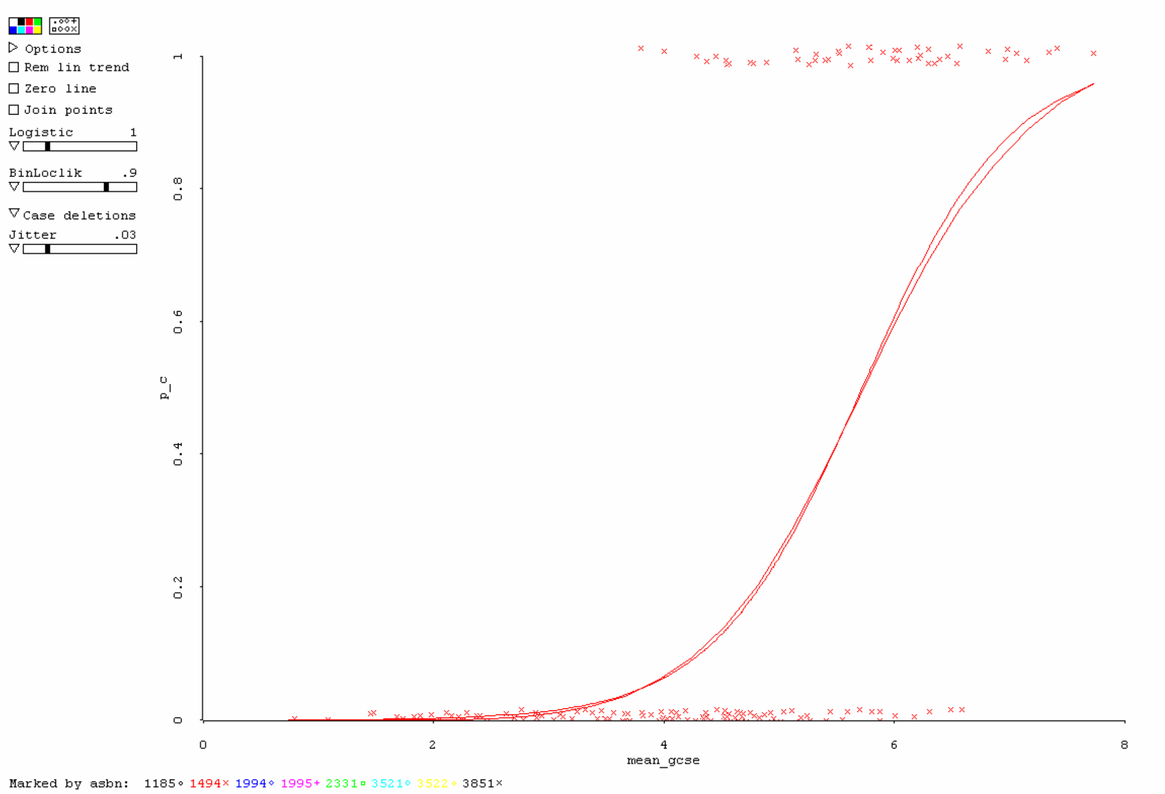


Figure C2: Comparison of logistic regression and smoothed line

In figure C3 a logistic regression line for each specification has been added. It can be seen that there is some evidence that there may be a need to fit interaction terms. Although the final interpretation should be based on the final analysis of the full data, this exploratory analysis does give some indication of what could be happening. Students taking GCSE in ICT with mean GCSE scores of 5 or more have a high probability of obtaining at least a grade C. This is not the case for the three GCSEvs in ICT.

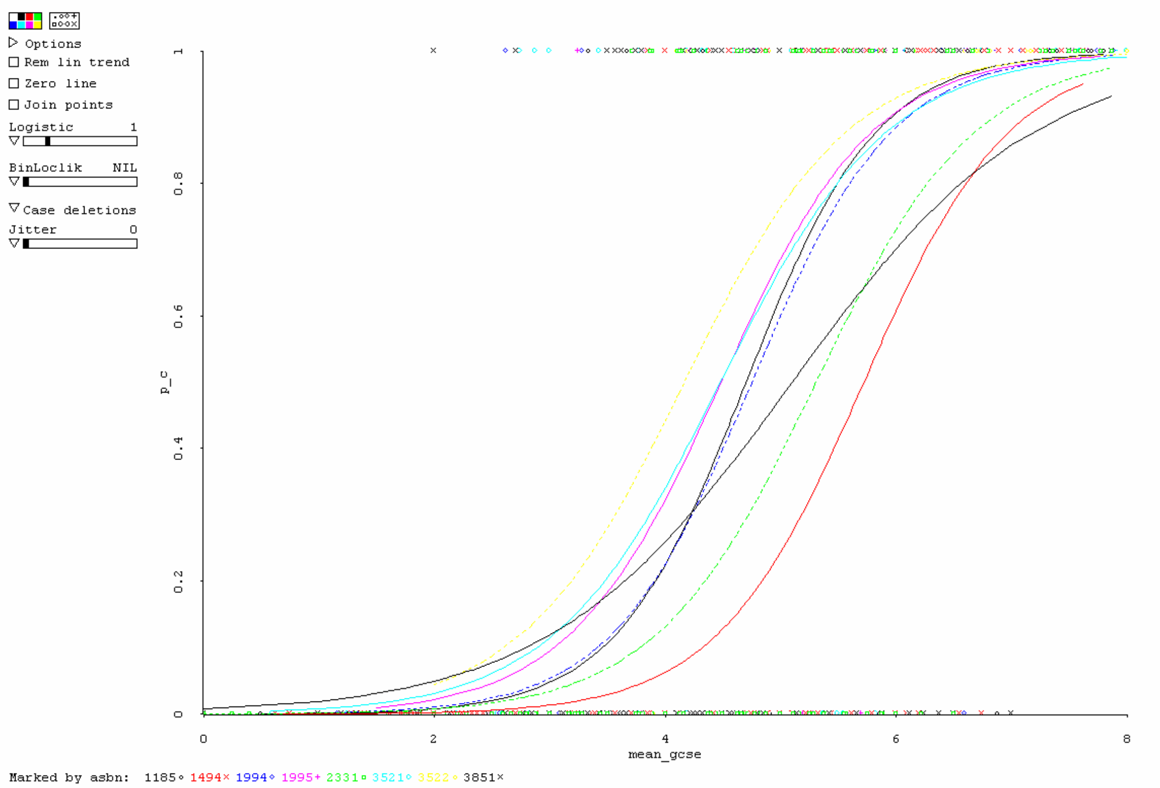


Figure C3: Logistic regression lines for each ICT specification

The exploratory analysis also suggested that there were no sex differences but that the relationships by school type could be complex and involve interactions (however, they could also be the result of interactions between school type and specification). It should be noted that this exploratory analysis is based on a sample of the data and is only intended to guide subsequent analyses. It is possible that the final analysis reaches conclusions that may differ from the exploratory analysis.

After completing the exploratory analysis, multilevel models were fitted using MLwiN. The model used in the main analyses is described in Figure C4.

$$\text{p\_a}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{1j}\text{gcsev}_{ij} + \beta_{2j}\text{gcse}_{ij} + \beta_3\text{mean\_gcse}_{ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\beta_{2j} = \beta_2 + u_{2j}$$

$$\begin{bmatrix} u_{1j} \\ u_{2j} \end{bmatrix} \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u1} & \\ \sigma_{u12} & \sigma^2_{u2} \end{bmatrix}$$

$$\text{var}(\text{p\_a}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

Figure C4: Multilevel model for the main analysis

The probability of obtaining at least grade A for a student i attending centre j, p_a_{ij}, is assumed to be distributed with a binomial distribution. The logit of this probability is modelled by three fixed terms defined by two dummy variables gcsev$_{ij}$ (taking the value 1 if the student sat a GCSEv and 0 otherwise) and gcse$_{ij}$ (taking the value 1 if the student sat a traditional GCSE and 0 otherwise) and a continuous variable representing the centred mean GCSE (i.e. the mean GCSE minus 5). The two dummy variables have random terms at the centre level. This results in three fixed parameters (β_1, β_2 and β_3). A necessary condition for the GCSE and GCSEv to be comparable is for β_1 and β_2 to be equal.

Reversing the logistic transformation of the coefficients above gives the probability of obtaining at least grade A for a student with an average GCSE score and attending an average centre. The more positive the values, the more likely the student is to obtain at least grade A.

There are three random parameters σ_{u1}^2, σ_{u2}^2 and σ_{12}^2 in the model. The first two are the variance components for centres taking GCSEv and GCSE respectively. The third random parameter is a covariance. If this is large and positive, and if the centre level residual for GCSEv is large, then the centre level residual for GCSE is also large for centres offering both. Centre level residuals are sometimes used as measures of value added but it should be recognised that they depend on the model and different models result in different values (Bell and Dexter, 2000).

In Table C1 the estimates for the parameters for the model described in Figure C4 are displayed. The important feature of this table is that the estimate for GCSEv is lower than that for GCSE indicating that the probability of obtaining at least grade A is lower for these students. The random effects indicate that the school level residuals are more varied for GCSEv centre entries. Although the covariance is positive, it is not significant so it should not be interpreted as evidence that centres that are good at teaching GCSEv tend to be good at teaching GCSE. However, it is not easy to interpret the parameters of a multilevel logistic regression and it is more useful to consider a series of plots derived from them (the parameter estimates for all models described in this report are included in Appendices E and F).

Table C1: Probability of obtaining at least grade A (Applied ICT Double Award, 2004)

Fixed terms

Parameter	Estimate	SE
Mean GCSE	1.717	0.016
GCSEv	-3.878	0.102
GCSE	-2.957	0.042

School variance

GCSEv		GCSE		Covariance	
Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
5.540	0.371	1.732	0.080	0.403	0.350

The graphs presented in this report were generated by calculating predictive probabilities. For example, the predicted probability of obtaining a grade A for students taking GCSEv in an average school is given by the following formula:

$$\hat{p}\_a_{ij} = 1 \Big/ \left(e^{-(-3.878 + 1.717 mean\_gcse_{ij})}\right)$$

Appendix D: Prior and concurrent attainment comparison

As a result of the greater interest in school effectiveness and improvement, there has been an increase in the use of value-added measures. This has led to the development of linked databases of educational test/examinations results (e.g. KS3/GCSE database, 16+/18+ database). These databases enable the relative progress of candidates to be assessed.

One of the main problems with prior attainment measurements is the effect of the difference in time. Clarke (1978) notes that two laws have evolved which seem to apply to all behavioural data:

> 'the earlier the measure, the less the long term reliability;
> the longer the period predicted, the less the reliability.'

There are many reasons for these laws. For studies involving children and adolescents, there is the issue of differing rates of development. Using a concurrent measure means that some of the factors that influence performance are being controlled. For example, mean GCSE is influenced by the same overall school effect as the GCSEv but the Key Stage 2 score is not. The GCSE score may be influenced by external factors that can vary for entries for the qualification but occur after the prior measurement was made. Another disadvantage with prior attainment is that this requires data matched across years. This is a more complex process than matching within years. Some candidates may not have prior attainment and it is not possible to match the data for them. This is a particular problem with Key Stage tests because they tend not to be taken by candidates from independent schools. Another problem with prior attainment is that it can lead to more complex models. For example, if Key Stage 2 data is included, a full analysis would have to consider much more complex models known as cross-classified random models. These would consider both the effects of primary and secondary schools for all students. Motivation may change between Key Stage 3 and Key Stage 4 if candidates become disenchanted with schooling in general. This is a particular issue given that GCSEvs may be more likely to be offered to such candidates. In general, prior attainment tends to control for fewer confounding factors compared with concurrent measures. Unless the prior attainment measures have much better measurement properties, preference should be given to concurrent measures.

In this report the main analyses use concurrent attainment (section 3) but they were also carried out using prior attainment (estimates of the latter are given in Appendices H and I). In many cases, the analyses based on prior attainment suggested that there may have been a problem with grading of GCSEvs and the difference between the GCSEvs and their comparators is larger when prior attainment is used. It can be argued that prior attainment does not control for changes in overall motivation. One way of investigating this hypothesis is to consider the progress in a common subject for the group of students taking GCSEv in the subject against the group taking the comparators. To demonstrate this, for the cohort that obtained an Applied Science Double Award or a GCSE in Double Science in 2005, the progress made in Mathematics from Key Stage 2 to Key Stage 4 is considered. Given that students sitting Applied Science and students sitting Double Science have to achieve the same pass marks for any given specifications, then any differences in Mathematics performance are not related to grading standards. For simplicity we do not report the results of models that include individual Mathematics specifications because the uptake of the different specifications was virtually identical for the two groups and so had no effect on the overall difference between them.

In a first step we computed the mean of the GCSE for both groups (Table D1).

Table D1: Mean GCSE for the GCSE and GCSEv cohorts (Science)

Subject	Number of students	Mean GCSE	Standard deviation
GCSE in Double Science	419,591	4.73	1.49
Applied Science Double Award	15,687	3.85	1.27

There is almost a point difference (equivalent to a grade) between both groups of students. The students doing the GCSEv have, on average, lower GCSE results. This difference is statistically significant.

In a second step, we fitted a multilevel logistic model where the dependent variable was the grade in Mathematics at GCSE. The independent variables were sex, type of subject (whether they studied GCSEv Applied Science or GCSE Double Science), Key Stage 2 result in Mathematics and school type (comprehensive being the baseline). Table D2 presents the results of the modelling.

Table D2: Modelling of the grade in Mathematics for the GCSE and GCSEv cohorts (Science)

Fixed

Model	Constant		Gender		Type of subject		Key Stage 2 (Maths)	
	Est.[3]	SE	Est.	SE	Est.	SE	Est.	SE
I = gender + subject + KS2	0.092	0.006	-0.107	0.002	-0.148	0.006	0.666	0.001
II = I + school type	0.010	0.005	-0.106	0.002	-0.146	0.023	0.665	0.001

Model	Grammar School		Independent School		Secondary Modern School		Other	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
I = gender + subject + KS2	-	-	-	-	-	-	-	-
II = I + school type	0.393	0.023	0.584	0.014	-0.018	0.023	-0.241	0.025

Random

Model	School		Student		-2 log likelihood
	Variance	SE	Variance	SE	
I = gender + subject + KS2	0.117	0.003	0.396	0.001	844345.1
II = I + school type	0.073	0.002	0.396	0.001	842609.8

Once prior attainment in Mathematics (Key Stage 2) has been accounted for, the model shows gender and type of subject to be significant predictors of the grade in GCSE Mathematics. Students taking Applied Science Double Award appear to achieve a lower grade in Mathematics compared to students who take GCSE Double Award.

The analysis thus suggests that students who took the GCSEv made less progress than students that took the non-vocational subject. An alternative model (model II) reveals significant effects for the school type, indicating that students in grammar and independent schools made more progress than students in comprehensive schools or other types of centres. However, the interaction of the school type with the type of subject is not significant.

The same analysis was performed for students who obtained an Applied ICT Double Award or a GCSE in ICT (any specification from any awarding body).

The mean of the GCSE for both groups is shown Table D3.

[3] Estimate

Table D3: Mean GCSE for the GCSE and GCSEv cohorts (ICT)

Subject	Number of Students	Mean GCSE	Standard deviation
GCSE in ICT	68,399	4.94	1.41
Applied ICT Double Award	40,926	4.63	1.38

The students taking the GCSEv have, on average, lower GCSE results, although in ICT, the means are much closer than for students taking Science courses. This difference is also statistically significant.

The results of the multilevel analyses are shown in Table D4.

Table D4: Modelling of the grade in Mathematics for the GCSE and GCSEv cohorts (ICT)

Fixed

Model	Constant		Gender		Type of subject		Key Stage 2 (Maths)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
I = gender + subject + KS2	0.139	0.008	-0.077	0.004	-0.132	0.009	0.672	0.002
II = I + school type	0.063	0.008	-0.076	0.004	-0.097	0.009	0.671	0.042

Model	Grammar School		Independent School		Secondary Modern School		Other	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
I = gender + subject + KS2	-	-	-	-	-	-	-	-
II = I + school type	0.399	0.033	0.490	0.022	-0.075	0.031	-0.168	0.042

Random

Model	School		Student		-2 log likelihood
	Variance	SE	Variance	SE	
I = gender + subject + KS2	0.105	0.004	0.375	0.002	208333.3
II = I + school type	0.079	0.003	0.375	0.002	207725.8

As before, gender and type of subject are significant predictors of the grade in GCSE Mathematics. Students taking Applied ICT Double Award appear to achieve a lower grade in Mathematics compared to students who take the comparator GCSE.

The analysis thus suggests that students who take the GCSEv made less progress than students who take the non-vocational subject. An alternative model (model II) reveals significant effects for the school type, indicating that students in grammar and independent schools make more progress than students in comprehensive schools or other types of centres.

The overall conclusion of both analyses is that analysis based on prior attainment can potentially generate misleading results. There is also a case for reviewing the use of prior attainment in the awarding process when there are potential differences in motivation between specifications or options. When prior attainment is used it is clear that apparent lack of comparability between specifications that are not generated by differences in the grading standard can be identified by statistical methods. For this reason, the analyses in this report concentrate on the concurrent attainment models.