Table 7: Proportion of candidates getting a higher or lower grade if marked by a different marker

Observed grade	*Proportion Higher*	*Proportion Lower*
A	0.000	0.265
B	0.272	0.409
C	0.302	0.374
D	0.336	0.339
E	0.330	0.255
U	0.227	0.000

would be a relatively easy way of understanding how much reliance should be put on the results given. A table like Table 7 is a more informative version of a reliability coefficient. Like a reliability coefficient it is not a fixed property of the test, but depends on the distribution of scores, the grade bandwidth and (in this case) the inter-marker correlation. The proportions cannot be interpreted as probabilities for individual candidates, however, because this would depend on how close the individual was to the grade boundary. The proportions apply to the grade scale as a whole.

Finally, some limitations of this study should be mentioned. First, we mainly looked at levels of inconsistent classification in one unit only. In reality this may not be as important to candidates, as we have shown the effect is almost certain to be diluted when aggregating over the three units of AS. This would be even more the case when aggregating over six units of A-level. Arguably, it is at the aggregate level that any inconsistent classification is particularly serious: for example, when grades are used to create point scores for university selection. Secondly, it may be that using a normal distribution to simulate the data is not the ideal method. For instance, having to truncate the distribution at zero and the maximum mark meant losing some of the data, and may have slightly distorted the distribution. It may be that other distributions would better match the distribution of the data in reality, such as the beta binomial (see Livingston and Lewis, 1995; Lee *et al*., 2000). Finally, this research only considered inconsistent classification arising from differences in correlation between markers' scores, not differences between markers in severity or bias. Future research could address some of these issues, and widen the scope to other assessments, such as GCSEs or admissions tests.

References

Bramley, T. (2007). Quantifying marker agreement: terminology, statistics and issues. *Research Matters: A Cambridge Assessment Publication*, **4**, 22–28.

Cresswell, M. (1986). Examination grades: How many should there be? *British Educational Research Journal*, **12**, 1, 37–54.

Lee, W-C., Hanson, B.A. & Brennan, R.L. (2000). Procedures for computing classification consistency and accuracy indices with multiple categories. *ACT Research Report Series*. Available online at http://www.act.org/research/reports/pdf/ACT_RR2000-10.pdf (accessed 23 October 2006)

Livingston, S.A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, **32**, 2, 179–198.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Murphy, R.J.L. (1978). Reliability of marking in eight GCE examinations. *British Journal of Educational Psychology*, **48**, 196–200.

Murphy, R.J.L. (1982). A further report of investigations into the reliability of marking of GCE examinations. *British Journal of Educational Psychology*, **52**, 58–63.

Newton, P.E. (1996). The reliability of marking of General Certificate of Secondary Education scripts: Mathematics and English. *British Educational Research Journal*, **22**, 4, 405–420.

Newton, P.E. (2003). The defensibility of national curriculum assessment in England. *Research Papers in Education*, **18**, 2, 101–127.

Newton, P.E. (2005). The public understanding of measurement inaccuracy. *British Educational Research Journal*, **31**, 4, 419–442.

Please, N.W. (1971). Estimation of the proportion of examination candidates who are wrongly graded. *British Journal of Mathematical and Statistical Psychology*, **24**, 230–238.

Vidal Rodeiro, C.L. (2007). Agreement between outcomes from different double marking models. *Research Matters: A Cambridge Assessment Publication*, **4**, 28–34.

Wiliam, D. (2000). Reliability, validity and all that jazz. *Education*, **29**, 3, 9–13.

EXAMINATIONS RESEARCH

Statistical Reports: Patterns of GCSE and A-level uptake

Joanne Emery and Carmen L. Vidal Rodeiro Research Division

Two new statistical reports have been added to the 'Statistics Reports' series on the Cambridge Assessment website (http://www.cambridgeassessment.org.uk/ca/Our_Services/Research/Statistical_Reports):

Statistics Report Series No. 4: *Uptake of GCSE subjects 2000–2006*

Statistics Report Series No. 5: *Uptake of GCE A-Level subjects in England 2006*

Data for these reports were extracted from the 16+/18+ databases. These databases are compiled for the Department for Children, Schools and Families (DCSF) from data supplied by all the awarding bodies in England. They contain background details and national examination data for all candidates who have their 16th, 17th and 18th birthdays in a particular school year. Candidates are allocated a unique number that remains the same throughout their Key Stage tests, allowing matching of examination data for longitudinal investigations. Records are present only if the candidate has sat an examination in a particular subject, not just attended classes.

This brief article outlines some of the results from both reports.

Uptake of GCSE subjects 2000–2006

There were a total of 561,407 students that attempted at least one GCSE examination in 2000. This number increased 12% to reach 629,523 students in 2006. The average number of GCSEs taken by candidates in the database was 8.36 in 2000 and 7.95 in 2006. This slight decline might be due to the increase in flexibility in GCSE studies, with new applied options for traditionally academic subjects (recorded as different qualifications), changes in the National Curriculum requirements, increased use of entry level qualifications or new 'hybrid' GCSEs that allow students to study on either academic or applied tracks.

An example of the results in the report is described here: the uptake of GCSE science subjects. In Statistics Report Series No. 4, similar analyses for almost all GCSE subjects are available.

The uptake of the separate sciences (biology, chemistry and physics) increased slightly from 2000 to 2006 but, on the other hand, the uptake of the double award in science fell almost 8 percentage points from 2000 to 2006. Girls were less likely to take the separate sciences at GCSE (which will limit their opportunities to progress onto science-based advanced level study). The uptake of biology, chemistry and physics was higher for the higher attaining students. This may explain why the percentages of students entered for the single and double award science courses was lower for the high attaining group compared to the medium and low attaining groups. The uptake of the separate sciences was much higher in independent and grammar schools than in comprehensive and secondary modern schools. With regard to the science double award, the uptake increased in independent schools (around 11 percentage points) but decreased in other types of schools. The uptake of the science single award increased only in comprehensive schools.

Other variables, such as the school gender, school boarding status or the characteristics of the neighbourhood in which the school is situated, were considered in this report and the uptake of the science subjects in 2006, according to candidates' school gender and various of the school neighbourhood variables, is presented in Table 1. Neighbourhood variables were downloaded from the Office of National Statistics Census 2004 data and were matched to the examination data according to the postcode of the school.

Table 1 : Uptake of GCSE science subjects in 2006 (percentages of students taking GCSEs)

		Biology	*Physics*	*Chemistry*	*Science: double award*	*Science: single award*
School gender	Boys	25.2	24.9	24.8	58.8	8.1
	Girls	13.2	12.4	12.7	70.7	9.3
	Mixed	6.5	6.2	6.3	71.0	11.6
Urban/rural indicator	Urban	8.0	7.7	7.8	69.1	11.6
	Town	6.9	6.7	6.7	78.3	8.9
	Village	9.5	8.8	9.0	74.6	9.6
Income deprivation affecting children	Bottom (lowest deprivation)	9.7	9.4	9.5	72.6	9.4
	Middle	7.1	6.8	6.8	70.1	12.2
	Top	5.2	4.8	5.0	64.5	14.1
% working-aged people with no qualifications	Bottom (lowest deprivation)	12.5	12.1	12.2	71.5	8.3
	Middle	7.1	6.8	6.8	72.5	10.8
	Top	4.4	4.1	4.2	66.2	14.7

Results for classifications based, for example, on school boarding status, multiple deprivation, employment rate and the percentage of people with Level 4/5 qualifications, are available in Statistics Report Series No. 4.

Uptake of GCE A-level subjects in England 2006

A total of 223,710 students in England attempted at least one A-level examination in 2006 (an increase of 7,897 students, or 3.7%, from the previous year). This figure equals less than a third of the number taking GCSE examinations in 2006. The modal number of A-level examinations taken was 3 (representing 49% of all candidates), followed by 4 (24% of candidates). If General Studies is excluded then 63% of all candidates attempted only 3 A-level examinations. These figures are similar to those reported previously for 2002 to 2005 in Statistics Report Series No. 3.

Statistics Report Series No. 5 lists the 30 most popular A-level examinations taken in 2006 and tabulates the percentages of candidates taking each of these subjects according to their school type, school gender and various school neighbourhood factors (mostly indicators of deprivation). The number of subjects and 'LEP' subjects taken (subjects listed by the University of Cambridge as providing 'less effective preparation' for their undergraduate courses) are also tabulated by these factors. The top 30 combinations of 3 or more A-level subjects is also presented.

The uptake of A-level science subjects and maths is presented in Table 2, categorised by candidates' school gender and a selection of school neighbourhood variables. Continuous variables (such as the percentage of working-aged people with no qualifications) were divided into three equal-sized groups using percentile values. The groupings here do not represent England as a whole because those from disadvantaged backgrounds are less likely to take A-levels. The full report additionally contains classifications based on school type and boarding status, estimates of neighbourhood income and the percentage of people with Level 4/5 qualifications.

Table 2 : Uptake of A-level science subjects and maths in 2006 (percentages of students taking A-levels)

		Biology	*Chemistry*	*Physics*	*Maths*
School gender	Girls' Schools	26.1	20.4	7.2	22.1
	Girls in Mixed Schools	17.7	10.5	2.8	11.9
	Boys' Schools	22.4	22.5	20.0	36.2
	Boys in Mixed Schools	15.4	14.4	16.0	24.8
Urban/rural indicator	Urban	17.8	13.8	9.3	19.6
	Town	18.8	13.0	10.8	18.2
	Village	19.1	15.4	11.8	22.6
Income deprivation affecting children	Bottom (lowest deprivation)	18.5	14.4	10.4	20.7
	Middle	17.5	13.2	9.6	19.6
	Top	17.8	13.7	8.6	18.5
% working-aged people with no qualifications	Bottom (lowest deprivation)	19.3	15.8	10.5	22.4
	Middle	17.5	13.3	10.0	19.3
	Top	17.0	12.3	8.1	17.0

The uptake of A-level science subjects and maths was higher in girls' schools than for girls in mixed schools. The uptake of English Literature and foreign languages was higher in boys' schools than for boys in mixed schools. However, single-sex schools are much more likely to be independent or grammar schools and these factors themselves were associated with higher uptakes of these subjects (some of the complexities of interpreting the examination results for single sex schools are discussed in a recent paper by a former Cambridge Assessment research officer (Malacova, 2007). Students attending schools and colleges in areas of higher deprivation were more likely to take fewer A-levels and more likely to take a higher number of LEP subjects. This will limit their opportunities to apply to courses at the University of Cambridge (a student will normally need to offer at least two non-LEP subjects). However, the differences are relatively small and did not take into account their previous attainment at GCSE.

Reference

Malacova E. (2007). Effects of single-sex education on progress in GCSE. *Oxford Review of Education*, **32**, 2, 223–259.

QUALITY AND STANDARDS

The OCR Operational Research Team

Elizabeth Gray OCR

To those within OCR (Oxford, Cambridge and RSA Examinations) the Operational Research Team (ORT) provides a constant source of advice, data and statistical support on all technical matters; to those outside OCR their work is largely unknown. This short sketch is an introduction to the main areas of interest of the team and its involvement in the life of OCR.

The outline will start, since at the time of writing the summer awarding series for general qualifications has just been completed, at the end, with the support provided to Awarding Committees and, crucially, Subject Officers and Chairs of Examiners. General assessments are becoming increasingly technical and the use of prior attainment measures to predict outcomes for both GCE and GCSE examinations requires technical manipulation of the highest order. Modelling aggregation (subject level) outcomes in unit based assessments is an essential part of awarding preparation and one which would cause problems were EPS (Examinations Processing System) to be used. In addition, where new subjects are awarded, additional data are provided to help with decision making. The awarding session also brings with it malpractice cases and the ORT supports the malpractice process and helps with the running of malpractice and appeals committees.

This work, though very intense, actually only represents a relatively small part of the ORT's programme. Vocational qualifications are awarded on a more continuous basis than general qualifications and again the ORT provides support for that process. This may, for some assessments, include producing question papers from a library of questions using complex statistical techniques to ensure standards are maintained.

New qualifications provide a source for much of the ORT's work and technical advice is sought regarding the assessment structures and marking regimes. When new specifications are proposed, for example the four unit A-levels, preparatory work is done to gauge the effect of the new assessment structure – in this example the effect of the decrease in the number of units on specification grade distributions. The outcomes from the work will again feed into awarding committees, and new developments, to aid the decision making process. When the issue is likely to affect all awarding bodies, for example the A* at GCE, then the research will be in collaboration with the Joint Council for Qualifications (JCQ). Indeed, many of the investigations undertaken by the ORT are at the behest of the Qualifications and Curriculum Authority (QCA) or the JCQ and contribute to a pool of knowledge shared by all awarding bodies.

QCA often want new qualifications to be trialled or piloted, as is the case for functional skills, and these trials/pilots have to be evaluated both for our own requirement and also for QCA as part of the pilot contract. The ORT has a standing programme of such evaluations which focuses mainly on the innovatory aspects of the trial or pilot and equivalence with existing qualifications. It was on QCA's behalf that a 'Stretch and Challenge' trial was conducted recently on new A-level assessments. This initiative was led by the ORT who will also be analysing the data once the scripts have been marked. The results of the analysis of this trial will be shared with all awarding bodies and QCA at a seminar in November 2007.

National Curriculum testing is now declining, but OCR took over that responsibility from the Assessment Research and Development division (ARD) of Cambridge Assessment in September 2005. This has led to a build-up of expertise in item level analysis which will stand OCR in good stead in the new e-environment. Collaboration across business streams on electronic script management (ESM) research has also enhanced knowledge in that area which can now be put to practical use.

A new member of the team, recruited in March 2006, has allowed more investigation into Malpractice, Appeals and Result Enquiries to take place. By identifying those subjects which attract the greatest number of events and changes arising from those events, research into underlying root causes can feed into specification development and strategies for improving marking reliability.

The quality of marking is always of concern, so much so that an internal OCR committee has been set up to consider the issues and identify investigations to be carried out by the ORT. Led by an ORT member, this committee also has presentations given by ARD members when their research relates to marking issues when the practical application of the research findings is considered.

When time permits, some of the issues raised by straightforward technical investigations lead to more detailed research. For example, as part of the continuous statistical monitoring of awarding decisions, research into awarding judgements showed that awarders cannot easily differentiate scripts which are only 2 or 3 marks apart. This finding lends support to the current awarding process where a zone of marks is defined by judgement of scripts and statistical considerations help to identify the final boundary mark within that zone.

The more OCR knows and understands about its processes the fewer errors are likely to be made and although it is the ORT's role to anticipate assessment issues and provide information to mitigate them, there is no doubt that trouble shooting is also required. In order to reduce this, the ORT is heavily involved in training Subject Officers and Chairs in all