Table 6: Inter-marker agreement, Oxford and Cambridge Assessment patterns

Item	*% exact agreement (uncorrected spellings)*					
	Oxford			*Cambridge Assessment*		
	Live v auto	*Auto v exr1*	*Auto v exr2*	*Live v auto*	*Auto v exr1*	*Ext1 v exr2*
q2biii	91.1%	91.1%	89.4%	93.2%	92.8%	91.1%
q4a_fur	91.3%	93.6%	90.6%	89.2%	92.3%	88.4%
q2cii	71.2%	82.9%	85.9%	72.4%	88.2%	87.9%

Agreement levels between the automatic marker and human markers were also broadly similar – for these items – to those found between human markers. We could find no simple explanation for why the remaining two 1-mark items were marked less well by the system – suitability for automatic marking does not appear to depend simply on item difficulty or the number of alternatives given in the examiners' written marking scheme. However, the 200 sample answers used for pattern-writing appear likely to be sufficient for screening 1-mark items for automatic marking. The system was generally less often correct, and there were bigger differences between auto-human and human-human agreement levels, for 2-mark items.

Patterns were written for three of the items by a temporary worker recruited by Cambridge Assessment. This worker was highly qualified in psychology and computing, but had had no previous exposure to the project or computational linguistics. The correctness and inter-marker agreement levels were similar for both sets of patterns, implying that it is possible to transfer pattern-writing skills from the developers to new staff. This is an important step for the commercialisation of the system.

We conclude that automatic marking is promising for 1-mark items requiring a short, textual response. More work is needed to see how the findings generalise to subjects and qualifications other than GCSE Biology, and to investigate why some items are less suitable for automatic marking using this system than others.

References

Pulman, S. G. and Sukkarieh, J. Z. (2005). Automatic short answer marking. In *Association for Computational Linguistics: Proceedings of second workshop building educational applications using NLP*, Ann Arbor, Michigan, 9–16.

Sukkarieh, J. Z., Pulman, S. P. and Raikes, N. (2005). Automatic marking of short, free text responses. *Research Matters: A Cambridge Assessment Publication*, **1**, 19–22. http://www.cambridgeassessment.org.uk/research/

EXAMINATIONS RESEARCH

The curious case of the disappearing mathematicians

John F. Bell and Joanne Emery Research Division

It is not unusual for claims to be made that some aspect of education is getting worse. Mathematics is often cited as a particular area of concern. There have been a number of reports about this issue including Roberts (2002), Smith (2004) and the UK Mathematics Foundation (2005). The declining number of A-level mathematicians is often cited as a particular concern, for example, in the *Times Educational Supplement* Gardiner (2006) wrote

> *'the number of A-level Mathematics students has slumped from 85,000 in 1989 to 66,000 in 2001, and (thanks to the misconceived Curriculum 2000 reforms) to just 52,000 in 2004.'*

A simple calculation would suggest that there has been a fall in numbers of the order of 33,000 students taking A-level mathematics, that is, a 39% decline. However, the interpretation of educational statistics is not a predictable 'one-piece jigsaw' but is instead a fairly simple multi-step problem. The first step is to identify the source of the statistics and check that they are comparable. It is not surprising or unreasonable that the source is not given in a newspaper story. However, an inspection of the available statistics would suggest that no identical definition of A-level mathematics students could simultaneously give a number as high as 85,000 in 1989 and as low as 52,000 in 2004. To investigate this problem, we decided to use the Summer Inter-board Statistics which have been compiled for A-level since 1990 in their present form (some earlier figures were obtained for 1989 but these may be a slight undercount).

After identifying a comparable source of statistics, the next issue is to consider the definition of A-level mathematics students. It is reasonable to assume that from the point of view of Higher Education and employment this should be based on the number with passing grades (A-E). This is important because in 1989 30% failed A-level mathematics and this was only 4% in 2004. A change in failure rates is unsurprising given that the introduction of modular A-levels led to candidates dropping mathematics rather than completing the course, obtaining a U and appearing in the statistics. Another relevant factor is the number of 17-year-olds in the population. This varied considerably over the period in

Table 1: Statistics for successful A-level mathematicians

Year	*A-level Maths passes*[1]	*Further Maths passes*	*No. of 17 year olds (government estimates in thousands)*[2]	*Maths passes/ No. of 17 year olds × 100*	*Further Maths passes/ No. of 17 year olds × 100*
1989[3]	50,570	5,022	715.2	7.07	0.70
1990	53,954	5,314	644.4	7.91	0.78
1991	51,185	5,144	599.3	7.94	0.80
1992	50,530	4,826	576.0	8.43	0.81
1993	50,129	4,553	551.5	8.70	0.79
1994	50,988	4,271	529.4	9.25	0.77
1995	51,708	4,465	537.2	9.77	0.84
1996	54,674	5,086	582.7	10.18	0.95
1997	58,514	5,216	604.0	9.68	0.86
1998	56,270	5,540	600.8	9.32	0.92
1999	56,192	5,306	586.2	9.35	0.88
2000	54,243	5,164	593.7	9.25	0.88
2001	54,193	5,223	614.7	9.13	0.88
2002	45,398	4,819	643.1	7.39	0.78
2003	47,059	5,224	645.4	7.32	0.81
2004	49,052	5,620	644.4	7.60	0.87

1. All mathematics except further mathematics 2. Population estimates – current releases – datasets T-04: England 3. These figures may be a slight undercount

question. In 1989 the cohort size was 715,200 and in the 2004 it was 645,400 and had fluctuated in between. To control for this two indices were calculated: the ratio of A-level mathematics passes over the number of 17-year-olds times one hundred and the same formula for further mathematics. Note these are not percentages since not all A-level candidates are in the 17-year-old age cohort.

The data derived from the above sources and calculations are presented in Table 1. The second column shows the number of A-level passes for any mathematics except further mathematics. This has not declined massively but has been around 50,000 for the period covered. Furthermore, it increased in the late 1990s and declined with the first set of A-level results in 2002 arising from the introduction of Curriculum 2000. The next column is the number of successful candidates for further mathematics. These have varied by around 5,000 and have shown no sign of a decline. In the fourth column, estimates of the number of 17-year-olds have been presented. The next two columns are the indices that control for cohort size. The index for mathematics passes increased until 1996 then started to decline. However, the latest available figure is not that much different from the figure in 1990. There is no evidence of a spiral of decline or an impending catastrophe. For further mathematics, the trend is less clear but there are now more further mathematicians than there were in 1989.

It is clear that there was a decline in mathematics passes associated with the introduction of Curriculum 2000. This curriculum change meant that typically candidates started studying four or five A-levels and then at the start of the upper sixth chose three of them to study for A-level. This has created problems for the interpretation of statistics about examinations. The awarding of modular A-levels is based on the results for candidates who have taken sufficient modules to obtain AS and A-levels. However, candidates do not necessarily claim the qualification preferring to resit modules or not bother if the overall grade is a U or they are taking the subject as an A-level. Thus, it is possible to have candidates in the matched databases who have AS only, both AS and A-level, and A-level only (although these candidates could have claimed an AS).

To understand the process it is necessary to analyse large matched databases of individual examination results (these are generated for England only to provide data for performance tables). By analysing these databases, we calculated that in 2003 there were approximately 15,000 year 12 candidates in England who obtained a passing grade at AS but did not proceed to A-level (note the numbers in Table 1 contain results for all ages and Wales and Northern Ireland as well as England). The total with a combination of AS and A levels was approximately 53,000. In 2000 there were approximately 4,000 candidates who succeeded in AS only (strictly, this is a different qualification, the Advanced Supplementary rather than the Advanced Subsidiary) and only 47,000 with either combinations of AS and A-levels. The introduction of Curriculum 2002 has been a success in increasing the numbers studying mathematics beyond GCSE. The decline in A-level success as a result of Curriculum 2000 can be explained by candidates not opting to take the A2 modules. One possible explanation is that there were several thousand candidates in the years prior to 2002 who would have dropped mathematics if they had been given the option and that these candidates would not have opted for a highly numerate discipline in higher education.

Another issue that is often raised is how to increase the number of A-level mathematicians. QCA (2006) argues that it is necessary to make A-level mathematics appeal beyond a 'clever core' of mathematicians to less enthusiastic and less able students. However, Gardiner (2006) criticises this and argues that there is a pool of able students:

> *'There are 31,500 students achieving A* grades in GCSE maths, yet the authors [referring to QCA, 2006] have no idea how many of these take maths A-level. One might expect between 10,000 and 15,000 to go on to A-level (the current number is probably much lower), and one can imagine incentives that would increase this to 20,000-plus.'*

Gardiner's estimates convert to between 31% and 48% for his expectations and 63% for his target with incentives. It is possible to calculate the figure using the matched database for England only. For those taking A-levels in 2004, the percentage of those obtaining A* going on to take A-level mathematics is 62%. To put this into perspective, this is higher than equivalent percentages for other subjects, for example, Chemistry (51%), English (48%), Biology (43%), Geography (38%) and

French (37%). In addition, 85% of successful mathematics students who progressed from GCSE mathematics had an A or A* at GCSE mathematics. This is comparable with some other subjects (for example, French – 93%, the sciences - ~80%) but is much higher than for other subjects such as English – 50% and history – 49%. These figures support the argument that if there is to be a large increase it must come from beyond the 'clever core'.

The findings of this research may come as a surprise. There has been no large scale decline in the number succeeding in A-level mathematics. The disappearing mathematicians can be accounted for by changes in the structure of mathematics leading to candidates dropping out rather than failing, and to demographic changes.

References

Gardiner, T. (2006). Make maths count. *Times Educational Supplement*, 10 March 2006, p.21. http://www.tes.co.uk/search/story/?story_id=2206638

QCA Research Faculty (2006). *Evaluation of the participation in A-level mathematics interim report. Autumn 2005*. London: QCA. http://www.qca.org.uk/downloads/qca-06-2326-gce-maths-participation.pdf

Roberts, G. (2002). SET for success. *The supply of people with science, technology, engineering and mathematics skills. The report of Sir Gareth Roberts' Review*. London: HM Treasury. http://www.hm-treasury.gov.uk/documents/enterprise_and_productivity/research_and_enterprise/ent_res_roberts.cfm

Smith, A. (2004). *Making mathematics count. The report of Professor Adrian Smith's inquiry into post-14 mathematics education*. London: DFES. http://www.mathsinquiry.org.uk/

UK Mathematics Foundation (2005). *Where will the next generation of mathematicians come from?* Preliminary Report of a meeting held in Manchester. 18/19 March 2005. http://www.ma.umist.ac.uk/avb/pdf/WhereFromPreliminaryReport.pdf

STANDARDS OVER TIME

What happens when four *Financial Times'* journalists go under the eye of the invigilator?

Miranda Green *Financial Times*, Education Correspondent, additional research by Anna Metcalf

In the weeks leading up to A-level results day Cambridge Assessment Research Division and its UK examination board, OCR, worked with *The Financial Times* to illustrate certain aspects of A-levels. On Saturday August 20, 2005, *The Financial Times* published the following two articles in *FT Weekend* which are reproduced here with their permission.

Four *FT* experts, and four, surely, of the most awkward and disputatious candidates ever likely to grace an examination hall. Earlier this summer, as the exam results season approached and, with it, the inevitable annual debate over standards, *FT Weekend* had a bright and apparently simple idea for getting to the truth behind the increasingly ritualistic argument: why not get a handful of the *FT*'s brightest and best to sit some of this year's papers? These writers, who live, breathe and even, dare we say it, pontificate about the subjects under consideration every day of their working lives, would then be able to give us their impressions of how today's exams compared with those that crowned their own school years.

Several twisted arms later, Lucy Kellaway, work columnist, James Blitz, political editor, Chris Giles, economics editor, and John Lloyd, editor of the *FT Magazine* and commentator on the media, had agreed to face their demons, and possible public ridicule, by submitting to an ordeal that most of the experts, politicians and critics who annually bemoan falling standards have long put behind them.

Accusations of dumbed down questions, grade inflation and lenient marking have dogged the A-level, once the unassailable 'gold standard', for years now. Every August, opposition MPs, employers, universities and independent schools voice their suspicions that ever-higher results (2005 is the 23rd year of improved pass rates) do not represent a true step forward in education or attainment. Their comments are reported just as families are waiting anxiously for an envelope from the exam board. Teachers and government ministers then reproach the doom-mongers for casting a cloud over the latest crop of good results, which they insist have been fairly earned by hard-working students.

But this year the pass rate edged up again to 96.2% – with A grades up to 22.8% – and the exam watchdog, the Qualifications and Curriculum Authority (QCA), has admitted that it will hit 100% in the near future.

If all candidates sitting an A-level – still well below half of the age-group – are deemed worthy of an E or above, that will deal a fatal blow to the credibility of certificates, say critics of the system. But, at bottom, the debate is about what an A-level with a near-universal pass rate is measuring and how marks and grades – particularly the A on which top universities have traditionally relied to 'sort' applicants – are awarded. It was this issue that our 'volunteers' agreed to probe.

"They're not going to ask me about Plato's *Republic*, are they?" said an anxious James as he agreed to put his strengths as the leader of the *FT*'s Westminster team to the test by sitting a paper on UK government.

Chris – economics editor of "the world's business newspaper" and preparing to answer questions on the national and international economy designed for 17 and 18 year olds – spotted the downsides with the ruthless insight for which he is feared and famed: "O God, go on then. Put me down for humiliation."

Lucy and John, both parents of exam-age children, took more of a scientific interest and approached their papers – on business studies and media studies respectively – with rather more equanimity.

After much debate about how to carry out the experiment, we decided to work in co-operation with an exam board, largely because we would