Changes in Examination Grades over Time: Is the same worth less?

British Educational Research Association Annual Conference
Brighton, September 1999

Dr Robert Coe
Curriculum, Evaluation and Management Centre, Durham University
r.j.coe@dur.ac.uk

Abstract

The performances in public examinations of large, nation-wide samples of candidates are considered. The average A level grade achieved has risen by about three quarters of a grade since 1988, while the scores of the same candidates on a test of developed ability used each year has fallen slightly. When the effects of measured ability are held constant, the change in grade level rises to well over one grade. The equivalent finding for GCSE data since 1994 is an increase of nearly half a grade, after adjusting for ability.

Possible explanations and implications are explored.

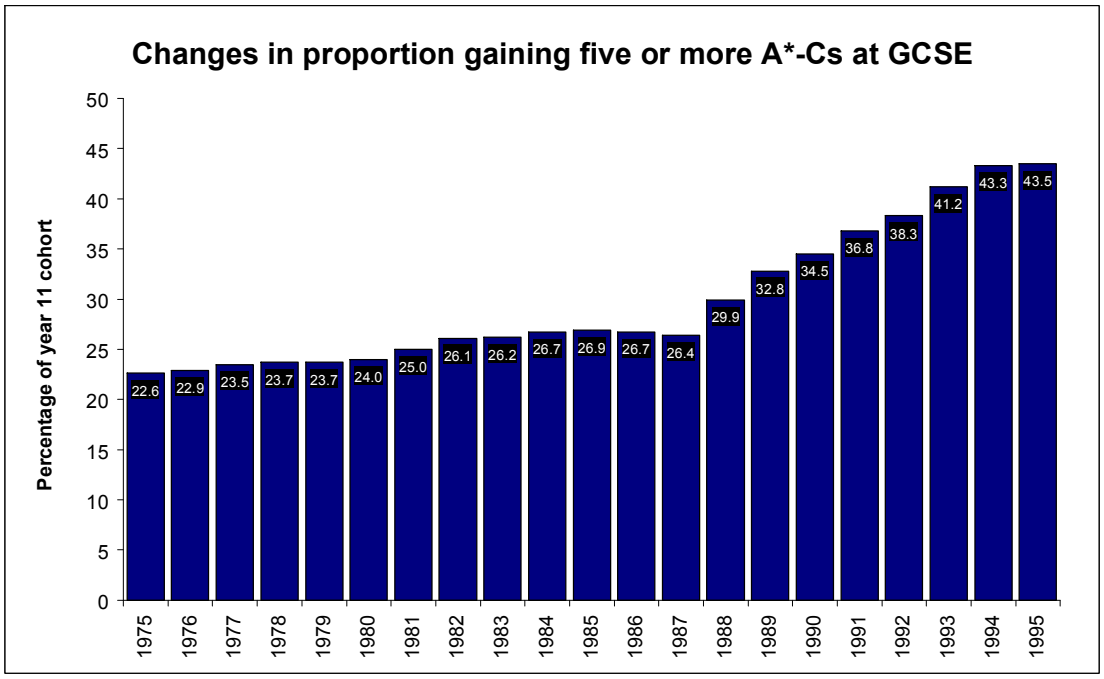
Introduction

The annual publication of A level and GCSE results has, in recent years, been accompanied by claims of falling standards as each year more people achieve higher grades. Despite this concern, and a number of attempts to address the issue, the evidence about whether standards have actually fallen is unclear. This paper presents new evidence which matches the performance of a sample of students in public examinations with their performance on a stable aptitude test.

The rise in achievement

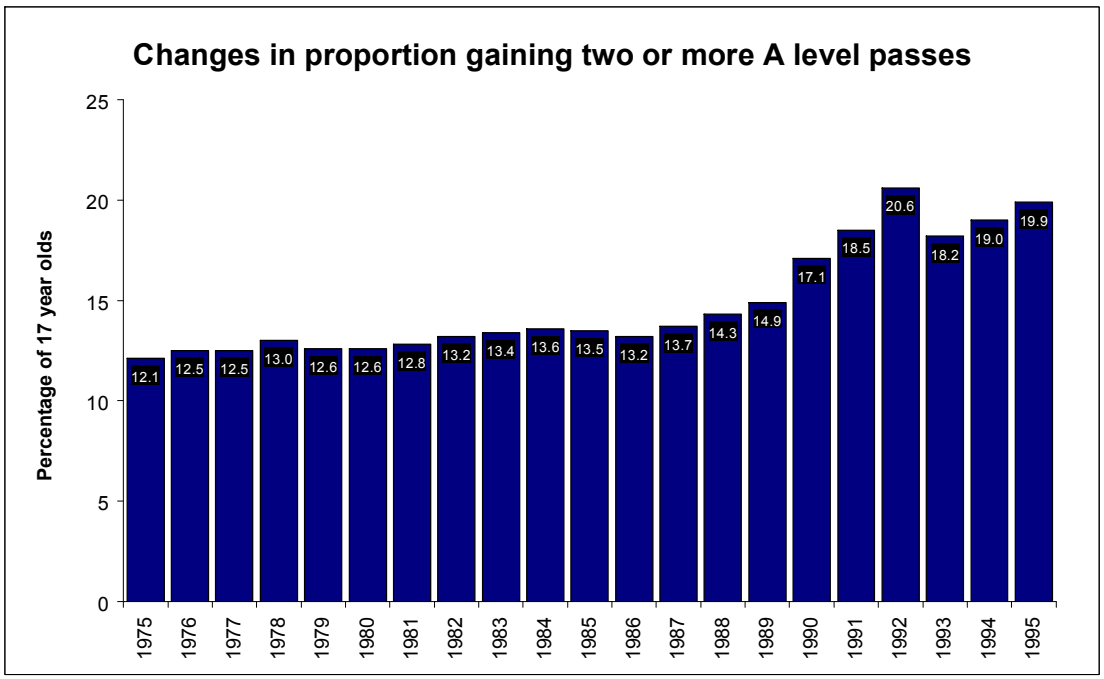
In 1996 the School Curriculum and Assessment Authority (SCAA) published two analyses of achievement in public examinations over time, one looking at GCSE (SCAA, 1996a), the other at A level (SCAA, 1996b). These clearly showed, as illustrated in Figure 1 and Figure 2, a significant rise in the proportions of people gaining qualifications at both levels.

Figure 1



(From SCAA 1996a, p9)

Figure 2



(From SCAA 1996b, p12)

In the same year, SCAA published jointly with the Office for Standards in Education (OFSTED) an inquiry to 'establish whether standards in English, mathematics and science (chemistry) in public examinations at 16+ and 18+ have been maintained over time.' (SCAA, 1996c). Their intention was to compare examination papers, mark schemes and candidates' scripts between 1975 and 1995, but they were somewhat

hampered by substantial syllabus changes over the period in question and the lack of availability of archived scripts and evidence about grade boundaries. Inevitably, therefore, their findings were rather inconclusive and equivocal. A quotation from the analysis of mathematics at 18+ illustrates their difficulty:

> The fact that the marks required to achieve grade boundaries were little changed for two of the syllabuses, in spite of the papers becoming less demanding between 1985 and 1995, suggests some lowering of standards for these syllabuses. However, this is somewhat at variance with the findings from the review of scripts, where a majority of judgements pointed to comparability of performance at the grade A boundary. (SCAA, 1996c, p49)

The latter view, however is itself somewhat at variance with the views of university teachers of mathematical subjects that, 'the mathematical achievement of students entering University with given grades is substantially worse than those entering with the same grades five or ten years ago.' (SCAA, 1996d, p7.1)

Evidence from ALIS and YELLIS

About ALIS and YELLIS

ALIS (the A Level Information System) began in 1983 as a system for helping schools to compare the progress their students have made with that of students in other schools. Currently over 900 schools and colleges participate in the project, which processes about a third of the A levels taken in the UK. Schools receive value added analysis for individual subject entries, based on simple residual gains when A level scores are regressed on average GCSE scores, as well as data on a range of student attitudes and perceptions. An optional part of the scheme is the International Test of Developed Abilities (ITDA), which is offered free of charge to participants in ALIS, should they wish to have an additional base-line measure from which to calculate value added.

There are four main reasons why they might wish to do this. Firstly, if a significant number of their students have not taken GCSE they would not otherwise be able to calculate value added performance. Secondly, a school that had been particularly successful at GCSE might feel that using this as a base-line for judging their A level achievements could put them at an unfair disadvantage. Thirdly, if their A level students have taken GCSEs in a number of different institutions, a school or college might feel the grades were as much a reflection of the institution as of the student, and might therefore want some kind of fair comparison. Finally, a school might accept the validity of GCSE as a measure of prior attainment, but still want the additional information provided by a test more focused on ability. Nevertheless, despite these reasons, many schools do not choose to use the ITDA, and those who do are possibly not representative of the whole ALIS sample.

The YELLIS (Year Eleven Information System) began in 1994 and now analyses the GCSE results of about 1300 schools. YELLIS uses its own test, taken by students in year 10 or year 11, as a base-line from which to calculate value added. The YELLIS test comprises two sections, mathematics and vocabulary.

ALIS data on A level changes

Choosing and defining the A level subjects to be included

Data showing the relationship between ITDA scores and performance in a range of A levels are available from 1988. The six subjects with the largest entries in ALIS were chosen to be included in this analysis: Biology, English, French, Geography, History and Mathematics.

In 1988 the analysis was done separately for each of these six subjects. However, as the ALIS project grew, more subjects were included, and some of those already included were subdivided. For example from 1992, English Literature was analysed separately from English Language, the former accounting for by far the majority of what had been classified as 'English'. To make things more complicated, however, some syllabuses could not easily be classified as either 'Language' or 'Literature', so a small number of entries were still given the generic title 'English'. For the purposes of this analysis, all entries in English up to 1991 and those designated 'English Literature' from 1992 have been included. A similar issue arises for Mathematics, with some syllabuses (e.g. SMP, MEI) being classified separately from 1993, and some option combinations (e.g. 'Pure Maths', 'Pure and Mechanics') also being classified separately. Once again, the category containing the vast majority of entries processed (in this case 'Mathematics') has been taken to represent the whole subject. In Biology and Mathematics, modular syllabuses were classified and analysed separately from 1995, so the figures quoted for these subjects do not include modular syllabuses.

A level results and ITDA scores

A level results and ITDA scores for all students in the sample between 1988 and 1998 are shown in Table 1, together with the number of entries included each year. 'ITDA avg' denotes the average percentage scored on the test by those students. 'A grade avg' is the average A level grade achieved by the sample, coded on the UCAS scale (A=10, B=8, C=6, D=4, E=2, N=0, U=-2). 'Correlation' is the Pearson product moment correlation coefficient between ITDA score and A level grade for that year, and 'Slope' and 'Constant' are the (Ordinary Least Squares) regression coefficients. 'SD' is the residual standard deviation.

Table 1

Biology

Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	2482	53.4	5.24	0.36	0.1000	-0.09	3.20
1997	2616	54.9	5.01	0.41	0.1300	-1.85	3.14
1996	2335	55.5	4.96	0.41	0.1300	-2.27	3.16
1995	1191	55.6	4.27	0.43	0.1400	-3.34	3.24
1994	2340	57.3	4.52	0.46	0.1490	-4.02	3.21
1993	4074	56.4	4.28	0.46	0.1500	-4.18	2.84
1992	1464	57.0	4.26	0.47	0.1490	-4.23	3.20
1991	1271	55.3	3.86	0.30	0.0947	-1.38	3.55
1990	549	60.8	4.31	0.46	0.1308	-3.65	3.37
1989	400	60.0	4.58	0.37	0.1056	-1.76	3.38
1988	221	63.7	4.33	0.44	0.1380	-4.46	3.48

English (Lit)

Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	4830	51.5	5.96	0.40	0.0950	1.06	2.60
1997	4365	51.9	5.76	0.42	0.1100	0.13	2.63
1996	3563	52.2	5.63	0.38	0.1000	0.26	2.78
1995	1275	53.6	5.54	0.20	0.1100	0.26	2.80
1994	3106	53.7	5.45	0.39	0.1080	-0.35	2.80
1993	5835	53.3	5.26	0.37	0.1040	-0.28	2.90
1992	2048	53.4	5.13	0.34	0.0976	-0.09	2.98
1991	550	53.6	4.42	0.18	0.0550	1.47	3.11
1990	655	54.8	4.79	0.31	0.0840	0.18	3.22
1989	539	54.8	4.83	0.24	0.0715	0.91	3.28
1988	336	57.0	4.59	0.10	0.0306	2.85	3.28

French

Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	2021	55.2	5.90	0.35	0.0920	0.81	3.00
1997	1821	55.3	5.58	0.35	0.1000	0.02	3.16
1996	1712	56.0	5.32	0.39	0.1200	-1.30	3.08
1995	645	56.7	5.32	0.29	0.0800	0.51	3.15
1994	1491	56.9	5.14	0.38	0.1220	-1.80	3.24
1993	2543	56.1	4.72	0.36	0.1100	-1.45	2.71
1992	1040	55.3	4.64	0.37	0.1170	-1.83	3.30
1991	814	55.3	4.47	0.22	0.0690	0.66	3.34
1990	334	57.3	4.54	0.42	0.1296	-2.89	3.31
1989	237	56.3	3.43	0.33	0.0948	-1.91	3.58
1988	132	58.2	3.77	0.11	0.0340	1.79	3.62

Geography

Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	2808	52.2	5.62	0.33	0.0990	0.47	2.95
1997	3439	52.7	5.28	0.34	0.1000	0.06	2.92
1996	2129	54.2	4.94	0.38	0.1300	-2.30	3.31
1995	793	54.5	4.73	0.35	0.1100	-1.50	3.42
1994	1411	55.4	4.60	0.34	0.1180	-1.94	3.46
1993	4012	54.5	4.51	0.35	0.1200	-2.03	2.89
1992	1706	54.4	4.25	0.32	0.1050	-1.46	3.34
1991	1201	55.1	4.34	0.25	0.0830	-0.23	3.50
1990	479	58.0	3.85	0.32	0.0858	-1.13	3.49
1989	404	58.5	4.07	0.31	0.0948	-1.48	3.36
1988	265	60.2	3.73	0.27	0.0860	-1.45	3.40

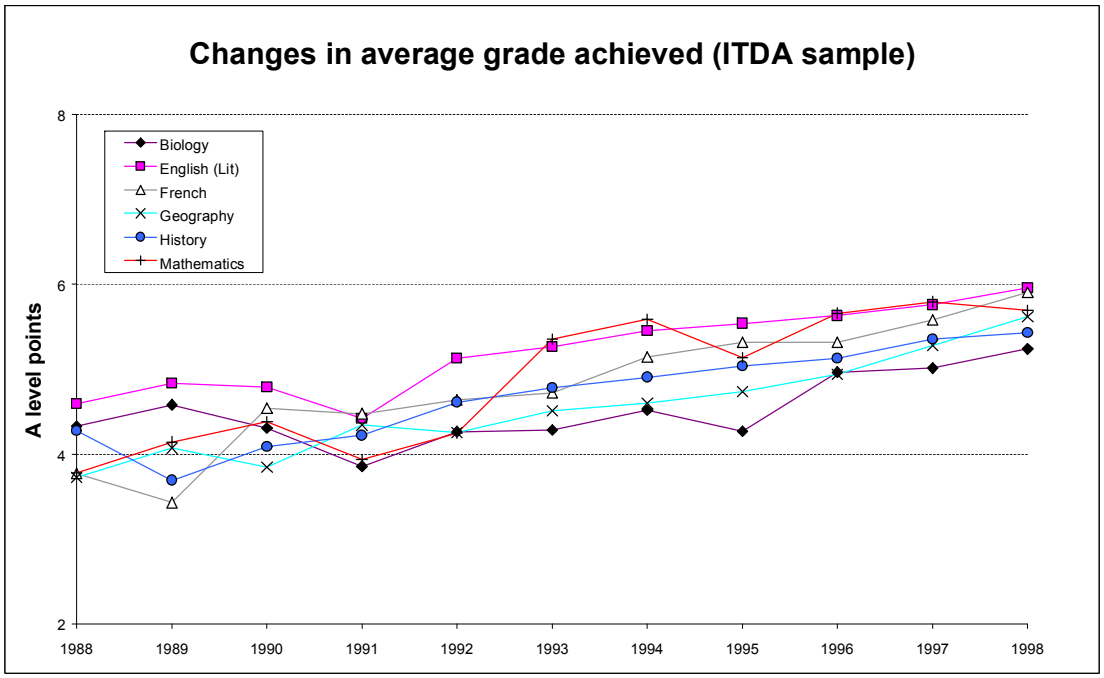
History

Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	3437	53.2	5.43	0.36	0.1040	-0.13	3.09
1997	3116	53.7	5.35	0.39	0.1200	-0.89	3.05
1996	3022	53.7	5.13	0.34	0.1000	-0.35	3.16
1995	1138	54.1	5.04	0.35	0.1000	-0.57	3.09
1994	2387	54.8	4.90	0.33	0.1030	-0.74	3.26
1993	4007	54.6	4.78	0.33	0.1050	-0.95	2.88
1992	1580	55.0	4.61	0.31	0.1000	-0.89	3.32
1991	1061	54.2	4.22	0.18	0.0570	1.13	3.37
1990	531	56.0	4.09	0.29	0.0906	-0.98	3.46
1989	469	55.7	3.69	0.21	0.0660	0.02	3.38
1988	251	56.8	4.28	0.10	0.0320	2.46	3.74

Mathematics

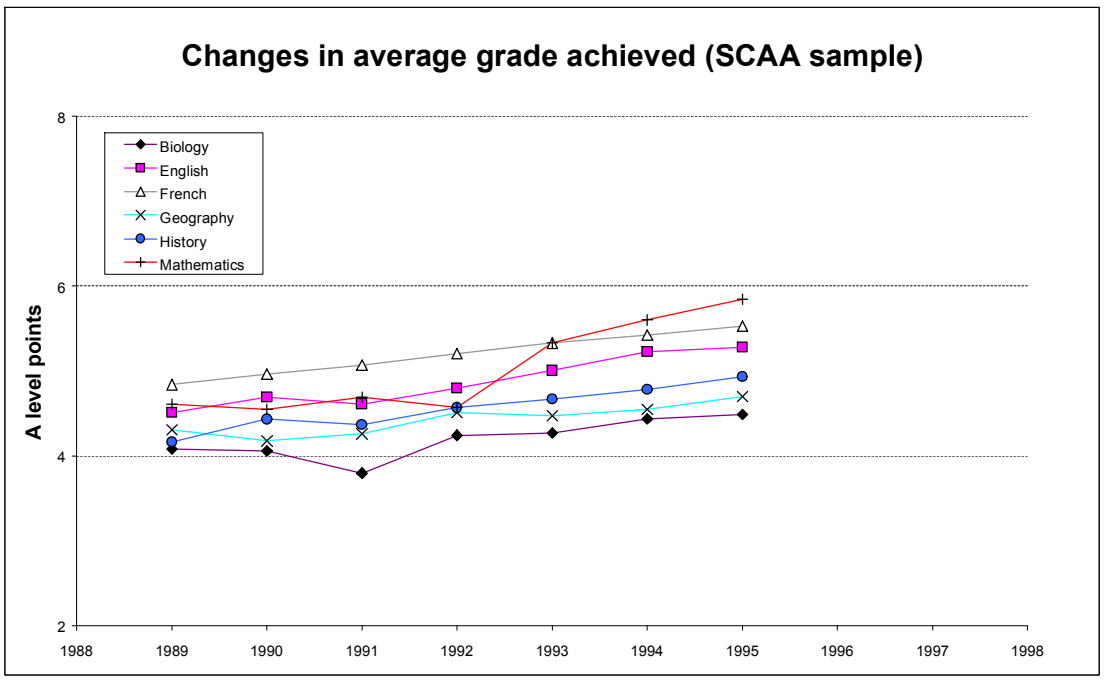
Year	Number	ITDA avg	A grade avg	Correlation	Slope	Const	SD
1998	3093	59.3	5.69	0.38	0.1120	-0.95	3.42
1997	2589	60.9	5.79	0.38	0.1100	-1.10	3.36
1996	2182	61.3	5.66	0.40	0.1200	-1.68	3.32
1995	370	63.0	5.14	0.34	0.1200	-2.18	3.77
1994	1415	65.1	5.59	0.42	0.1270	-2.68	3.33
1993	2822	63.6	5.35	0.38	0.1180	-2.15	3.22
1992	625	64.7	4.25	0.40	0.1440	-5.07	3.65
1991	1771	61.6	3.94	0.26	0.0850	-1.30	3.89
1990	926	67.8	4.38	0.36	0.1194	-3.72	3.70
1989	800	69.0	4.14	0.44	0.1380	-5.38	3.52
1988	562	72.3	3.78	0.48	0.1640	-8.08	3.56

Figure 3



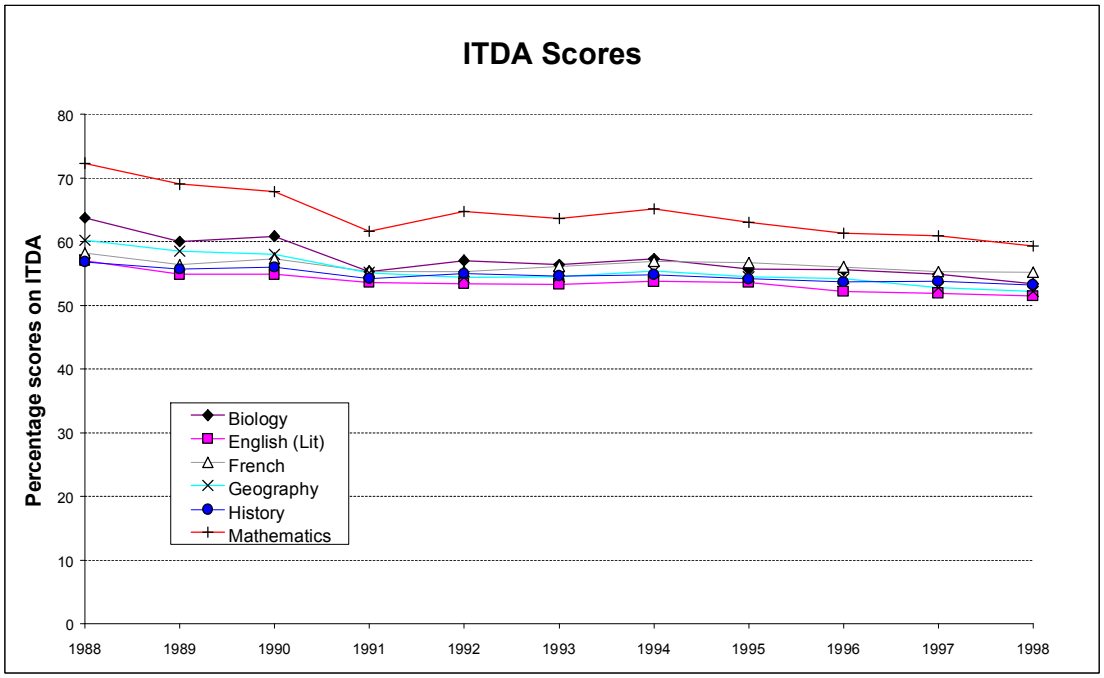
The changes in grades achieved by the sample are shown in Figure 3. In all subjects there was a steady rise. For comparison, the average grades achieved in these subjects for all A level entries in England and Wales are shown in Figure 4 (derived from SCAA, 1996b). Although there are some differences in the average achievements of the two samples for particular subjects and years, the overall trend is very similar, and, generally speaking, the match is fairly good. However, it could certainly be argued that the ITDA sample is not representative of A level entries as a whole.

Figure 4



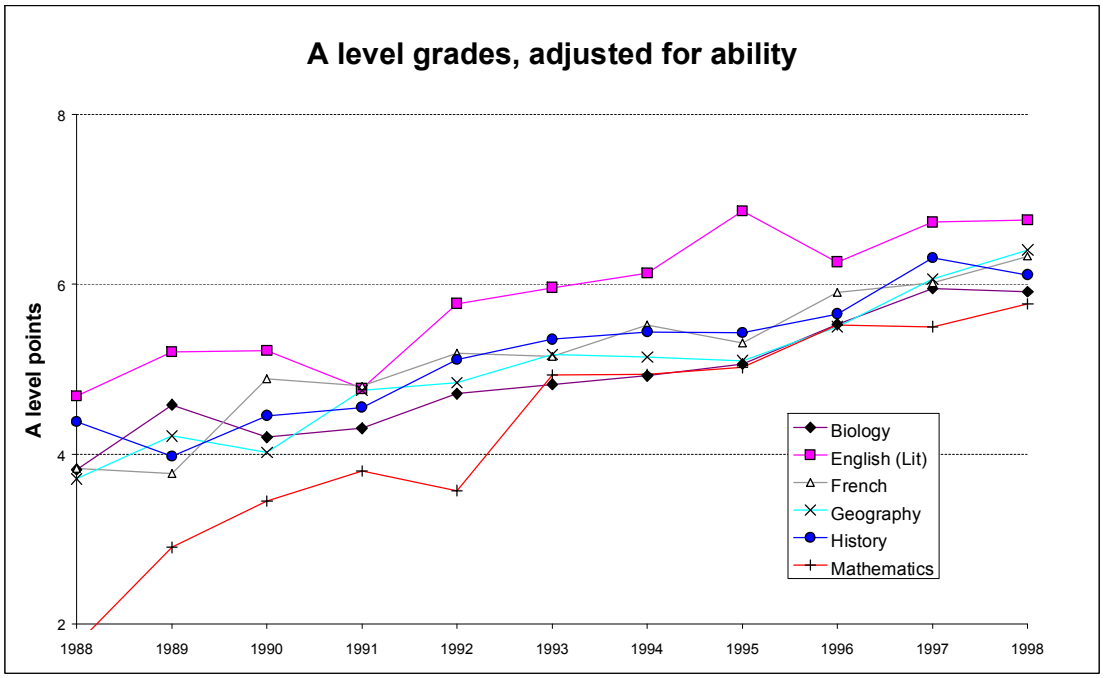
The ITDA scores of students in the ALIS sample for each subject and year are shown in Figure 5. It can be seen that there has been a small but steady decline in the ability of the sample in all subjects.

Figure 5



The combined effect of increasing grades achieved and decreasing ability can be seen using the regression equations shown in Table 1. A typical ITDA score of 60 was fed into the equation for each subject in each year to calculate an estimate for the average grade achieved by students who had that ability score. The results are shown in Figure 6. A fair summary would be to say that A level candidates across a range of subjects achieved over a grade higher in 1998 than candidates of the same ability had done in 1988.

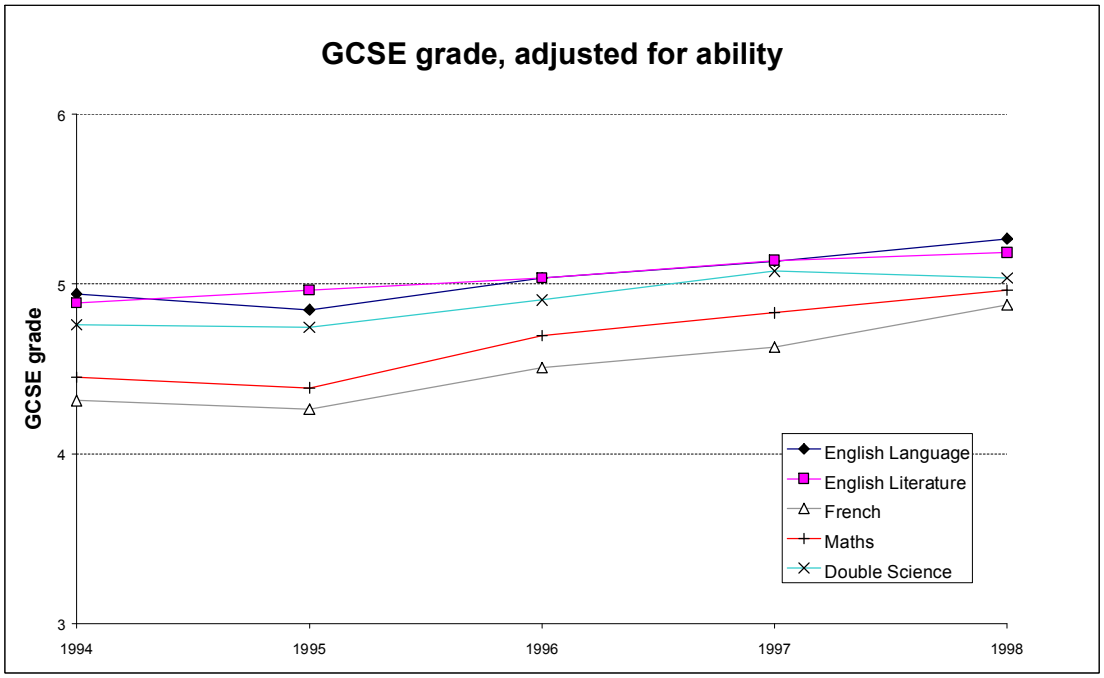
Figure 6



YELLIS data and GCSE grades

The equivalent graph for GCSE is shown in Figure 7. Once again the regression equations were used to estimate the average achievement of students with the same score on the YELLIS test. Once again the trend is for the grades achieved by students of the same ability to iincrease.

Figure 7



Discussion

Methodology

There are a number of criticisms that could be made of the methodology employed in this analysis

Appropriateness of the reference test

The use of a reference test such as the ITDA or YELLIS test as a fixed point from which to measure change in grade standards depends crucially on a number of assumptions. If what is measured by the test does not overlap sufficiently with the graded examination then it is simply an irrelevance. The fact that scores on a measure that has changed each year have declined relative to scores on another that has remained constant is of no significance if their content does not have at least some overlap. This is a particular problem when the same reference test is used for a range of different subjects and when the content and style of examination within a given subject may have changed appreciably over the period in question.

To some extent this question is addressed by considering the correlation between the two measures. The correlations between the ITDA and A level can be found in Table 1. Although they vary somewhat, they are generally in the range 0.3 to 0.5 and for most subjects most years are of the order of 0.4. This is high enough to indicate overlap between the two measures, but is perhaps not high enough to give real confidence in the validity of the ITDA as a base-line. In the case of the YELLIS test, correlations with GCSE grade are generally in the region of 0.7, sometimes as high as 0.8, so it would be much harder to argue that this test was inappropriate.

Interpretations

There are a number of (not exclusive) possible interpretations of the results presented here, including the following:

- **Teaching and learning have improved.** This is argued every year by teachers and their associations, and, indeed, may well be true. It is, however, hard to see how the claim could be convincingly substantiated. OFSTED purports to evaluate the quality of teaching and learning, but its judgements have little scientific credibility.
- **Examination performance has improved.** It could be that modern phenomena such as the increasing pressures of league tables and the frantic setting of targets might lead to an increased focus on aspects of examination preparation, without necessarily having any effect on learning. Tactics such as paying closer attention to syllabus content, endlessly practising previous years' papers or being more selective about who is actually allowed to sit the exams could all have this effect.
- **Changes in assessment have made the same levels of competence easier to demonstrate.** Changes such as the introduction of assessed coursework or the use of modular examinations could arguably have had this effect. The quality of work presented for examination may well be equal to or better than that of candidates in previous years, and therefore standards could not really be said to have fallen. However, given identical conditions, today's candidates might nevertheless be unable to match the performance of their predecessors.

- **Demographic changes have facilitated increased academic achievement.** Because of the increasing proportion of the population that can be classified as belonging to the higher socioeconomic groups, and the well know correlation between socioeconomic status and academic achievement, it is sometimes argued that this can explain the year on year rises in grades awarded.
- **The reference tests used in this analysis are inappropriate.** If performance in the ITDA or YELLIS test is as irrelevant to examination performance as shoe size, then it makes no difference how the comparison has changed.
- **The content and style of GCE and GCSE examinations have changed too much to make valid comparison possible.** Making a judgement about changes in grade standards over time requires that examinations at different times are measuring broadly the same thing. It may well be true that today's candidates would perform very poorly on an examination of ten years ago. However, if it is also true that candidates prepared for the older exam would struggle with today's papers then it is hard to argue that either is really harder; they are simply different.
- **Grade standards have slipped.**

Implications

Whether or not grade standards have been maintained will no doubt continue to be debated each year, particularly now that the Secretary of State has staked his reputation on the results. No doubt too, the debate will be politically highly charged, vigorously argued and inconclusive. The issue is a complex one, and all the explanations listed above are to some extent true; it is not possible to say definitively either that standards have slipped or that they have not. However, there may be one important sense in which it is now true to say that 'the same is worth less'.

Most of the consumers of examination grades – that is to say people who use them to make judgements, or prescribe them as requirements in some way – are not really interested in the precise knowledge, skills or understanding they certify. An 'A' grade in mathematics, for example, is not significant for its demonstration that a person, on a certain day in June, solved a particular equation in a matter of a few minutes, but because it provides evidence of qualities such as a general numerical ability, the potential to learn and the capacity to apply oneself to an extended and difficult project. The grade itself is therefore most often interpreted as a proxy measurement of these abilities. If that is the case, then it matters less whether the award of each grade represents the same achievement than whether it signifies the same level of general ability in the candidate. The evidence from the ALIS and YELLIS samples seems to be that it does not.

References

SCAA (School Curriculum and Assessment Authority) (1996a) *GCSE Results Analysis: an analysis of the 1995 GCSE results and trends over time*. London: SCAA.

SCAA (School Curriculum and Assessment Authority) (1996b) *GCE Results Analysis: an analysis of the 1995 GCE results and trends over time*. London: SCAA.

SCAA (School Curriculum and Assessment Authority)/OFSTED (Office for Standards in Education (1996c) *Standards in Public Examinations 1975 to 1995*. London: SCAA.

SCAA (School Curriculum and Assessment Authority) (1996d) *Seminar on A level Mental Mathematics*. London: SCAA.