

The pitfalls and positives of pop comparability

Nicky Rushton, Matt Haigh and Gill Elliott Assessment Research & Development Comparability Team

Introduction

During recent years the media debate about standards in public examinations has become something of an August ritual. The debate tends to be polarised with reports of 'slipping standards' at odds with those claiming that educational prowess has increased (Lebus, 2009). Some organisations have taken matters into their own hands, and have carried out their own studies investigating this. Some of these are similar to academic papers; others are closer in nature to a media campaign. In the same way as 'pop psychology' is a term used to describe psychological concepts which attain popularity amongst the wider public, so 'pop comparability' can be used to describe the evolution of a lay-person's view of comparability. Studies, articles or programmes which influence this wider view fall into this category and are often accessed by a much larger audience than academic papers. In this article, five of these studies are considered: Series 1 of the televised social experiment "That'll Teach 'em", conducted by TwentyTwenty Television; The Five-Decade Challenge, mounted by the Royal Society of Chemistry; the *Guardian's* and the *Times's* journalists (re)sitting examinations in order to experience their difficulty first-hand; a feature by the BBC Radio 4 programme, 'Today' (2009), which asked current GCSE students to examine and discuss exam papers from 1936; and a book of O level past papers and an associated newspaper article which described students' experiences of sitting the O level exams.

Experiments like these are largely unreported amongst the academic community, but they are influential within the popular press. This article explores the strengths and weaknesses of the studies, questions whether they should be taken into greater account by the academic community, and investigates the extent to which they help or hinder public perceptions of standards of qualifications in schools.

"That'll Teach 'em"

"That'll Teach 'em" was a television series which achieved worldwide success. Versions of the format were developed in Holland, Germany, Belgium, France, Norway and Spain. There have been three series of the show in the UK: the first airing in 2003 (recreating a 1950's grammar school and featuring academically high-achieving pupils); the second in 2004 (a 1960's secondary modern, focused upon vocational skills); and the third in 2006 (a grammar school again, this time focusing on single-sex classes). Series 1, which will be the focus of discussion in this article, was watched by 3.25 million viewers.

The purpose of the programme was to provide both entertainment and an investigation of examination standards in the UK. Thirty students who had just finished sitting their GCSE examinations undertook to board at the 'school' set up by the programme makers. They had four weeks of 1950's style lessons as well as experiencing the living conditions, food

and discipline of the era. At the end of the experiment they sat a partial GCE 'O' level exam in four subjects (Maths, English, English Literature and History) that was marked to the standards of the 1950s.

The experiment addressed a number of features which are often unrecognised when long-term standards over time (i.e. when comparisons span a large number of intervening years) are addressed in the media. Students were:

- removed from their usual environment and placed into a situation resembling that of the period of history being compared as closely as possible;
- taught a 1950's curriculum for a period of four weeks;
- taught according to the 1950's style for a period of four weeks;
- fed 1950's food.

This being a television show, there were additional concerns above and beyond the social experiment – the programme needed to make interesting viewing and to be accessible for a wide audience. Thus, the actual televised episodes would have been edited with this in mind, which might have detracted from the explanation and investigation of standards over time. Also, whilst the students were experiencing the teaching style and living conditions of the 1950s they were also being followed by a camera crew, which may have caused distraction.

The strengths of the programme included debating the topic of standards over time in a public context in a way in which the context of changes in society in the corresponding time were not only acknowledged, but put into the heart of the debate. It was not just about how well students might fare when given question papers from the era, but about what the whole experience of education was like. Much of the discussion was not about how 'standards' differ, but about how experiences differ, and that is a crucial distinction when considering long-term standards over time.

The major limitations of this study as an exercise in investigating standards over time were that: (i) the student sample was small, so it was not possible to draw a great deal from the ultimate examination results; (ii) the experiment was conducted for purposes of entertainment as well as investigation, so a more 'academic' report on its findings was not commissioned; and (iii) although the programme makers went much further than many other commentators in engaging with the social context of the time in question, the students were still modern students experiencing a previous culture, rather than truly representative of the previous era. So, although the students 'experienced' a 1950's curriculum and, to an extent, the lifestyle of the time, their underlying knowledge of the school equipment, teaching styles, and home and school environments pertaining to the 21st century would inevitably have influenced their learning and behaviour during the course of the experiment. The limited time scale of the experiment was also a drawback – students were not undertaking a two-year course, and were

only taking shortened versions of O level papers. Also their motivation for undertaking the experiment and all that it entailed would have been very different from their motivation towards their high-stakes GCSE examinations.

The students' exam results (from the partial O level set at the end of the study) showed a relationship with their GCSE results (which had been taken just before the experiment). In all four subjects there was a trend for students who attained a grade A* at GCSE to score higher on average in the partial O level examination than students who achieved an A at GCSE, who themselves scored higher on average than students who achieved B at GCSE. This trend held for English Literature, English Language and Maths at GCSE grades A*–D (no students scored lower) and in History at grades A* to B. These statistics were based upon very small samples of students, and hence were highly unreliable, but nevertheless showed a reassuring trend.

The Five-Decade Challenge

This study was carried out by the Royal Society for Chemistry (RSC) in 2008. Over a thousand students sat an examination paper containing numerical and analytical chemistry questions drawn from O level and GCSE exams spanning 1965 to 2005. The average scores from the questions in each decade were used as a measure of standards over time. The results are given in Table 1.

Table 1: Five-Decade Challenge results

Decade	Average score
1960s	15.4%
1970s	18.2%
1980s	22.2%
1990s	34.9%
2000s	35.3%
All questions	25.5%

The report concluded that:

Performance against each decade showed a remarkably steady step-wise progression, with the average scored for the 1960's questions being 15%, rising to 35% for the current 2000's decade. Changes to the syllabus and to the language used in examinations since the 1960s may partially explain this progression, but are unlikely to provide a complete explanation. (Royal Society of Chemistry, 2008, p.2).

It is interesting to note that the report did not make any explicit reference to declining standards, but focused its recommendations on changing curriculum content to emphasise quantitative and analytical science skills. However, the headline writers used the report to make significant pronouncements on science standards:

- *The proof that science exams are easier now* (Daily Mail, 2008)
- *Dumbed-down science is 'failing a generation'* (Daily Telegraph, 2008)
- *School science standards are slipping, says study* (Guardian, 2008)
- *GCSE students flunk past papers in experiment that exposes decline in standards* (Independent, 2008)

The study had strengths in its sample size and addressed issues of student motivation by offering financial reward for top scores. However, the method for selecting questions for the test may have meant that the items from each decade were not representative of the 'standard' at the time. It can also be argued that the use of average score to represent a standard was meaningless without being referenced to the overall 'pass-mark' at that time. For example, it could be that in the 1960s a score of 15% represented a 'pass' but in the 2000s a score of 40% represented a 'pass'. If that were the case, the results in Table 1 would have to be interpreted very differently with respect to the relative difficulties of the exam questions.

The study did acknowledge the potential influence of curriculum changes on the outcomes. Usefully, the report identified which questions were part of the current Chemistry curriculum. Restricting analysis to this set, which was not done in the original report, provides the alternative results in Table 2.

Table 2: Five-Decade Challenge results restricted to questions in the current curriculum

Decade	Average score
1960s	51.5%
1970s	27.5%
1980s	39.6%
1990s	34.2%
2000s	35.2%

Although the restricted analysis is based on fewer questions, the outcomes are very different. If average score on questions testing the current curriculum is used as a proxy for standards over time, then science exams in the 1960s were much easier than they are today.

The overt lobbying for curriculum change contained in the report and the lack of peer review probably make the findings less credible for academic researchers. However, the method and ideas contained within the study could provide a stimulus for further research.

The report provided helpful commentary to aid the public perception of standards; however, the media representation radically oversimplified and made judgements on standards not supported by evidence in the report. The RSC seemed to suggest that comparing standards over time is worth pursuing, but acknowledged the complexity of confounding factors. It is interesting to note the government response contained in the *Daily Telegraph* article reporting on the study:

... exam standards are rigorously maintained by independent regulators and we would rather listen to the experts whose specific job it is to monitor standards over time. (Daily Telegraph, 2008)

Commentators re-sitting examinations

Some smaller scale experiments reported in the popular press have involved journalists sitting A level papers in order to draw conclusions about whether the exams are as easy as has popularly been claimed. The two instances in the past few years, reported in the *Times* and the *Guardian*, both used this method, but differed slightly in their approach.

Journalists at the *Times* were given a choice of subjects, and were then allowed to pick which A level paper they sat within that subject. In contrast, the journalist at the *Guardian*, whilst given a free choice of

subject, sat all the papers in order to take the full A level. Unsurprisingly, most of the journalists picked subjects that they were familiar with – either subjects that they had studied at university, or subjects that were related to their current jobs. Sometimes the links between subject and job were obvious, such as the Berlin correspondent choosing German or the political correspondent choosing politics. Other journalists were slightly more adventurous in their choice: the linguist who chose English Literature having always wanted to study it at A level or the writer who chose Critical Thinking as it was something that should relate to journalism. Inevitably, subject choice affects the results of the studies. A levels are not intended for those who already have a degree in the subject; therefore (as pointed out at the end of the *Times* article) you would expect the journalists to do well in the subject areas that they worked in. Even the linguist did not have experience representative of an 18 year old's, as she had clearly encountered literary study as part of her university degree.

Another feature of these studies is often the short amount of time that is given to prepare for and sit the examinations. The journalist at the *Guardian* took A level English (including AS) in one year, whilst the *Times* journalists appeared only to have been given a few days. The argument seems to be that if an A level can be studied successfully in such a short period of time, then it cannot be worthwhile, or it must have been devalued. These arguments ignore the experience which journalists bring to their examinations. They talk of exam techniques such as "quoting authorities", timing, and choosing "bluffable subjects", in addition to their already considerable writing skills. "Making a plausible argument is something that I have been paid to do for the past 25 years" (journalist taking critical thinking paper – Mary Ann Sieghart). This makes their experience rather different to that of the average 18 year old. Nor do they have to cope with the difficulty of learning three, possibly four, new subjects at once.

Perhaps the most useful output of these studies is the insight that they give journalists into the experiences of those taking the exams. None of them reported that they found the experience easy, with several of them experiencing the same nervousness that they had when taking A levels in the past. Whilst some questioned whether they deserved their grades, none of them concluded that A levels are easier now. In fact one started by saying, "The one I sat was as demanding as any I tackled in the mid-1960s". The power of these studies is that they make the general public realise that gaining good marks on an A level paper is not as easy as the press sometimes claims that it is.

Discussions of/commentary on historic examinations by current students

Exam results and standards are frequently discussed in the media. Every year, on results day, there is the predictable round of stories of higher results than ever, taken to imply 'dumbing down' of the A level system:

- *GCSEs hit new high as experts criticise tests* (*Daily Telegraph*, 2010)
- *So easy a five-year-old has passed and a seven-year-old got an A star: Record numbers achieve GCSE top marks* (*Daily Mail*, 2010)

These headlines represent a common misunderstanding of the relationship between results and standards. Whilst the percentage of students achieving particular grades may have increased, this does not automatically imply that standards have fallen, or that the exams have

got easier. All the students achieving a grade will have met the standard required for it.

Often, as in the examples above, news stories are based on results alone, but occasionally discussion is informed by the inclusion of additional evidence. One such discussion was aired on BBC Radio 4's 'Today' programme in 2009, where the discussion included extracts of pupils comparing examination papers from the 1930s with today's papers.

This particular method is useful as it uses pupils to make the comparisons. As they are of the relevant age, they are arguably better able to make a judgement about how difficult they would have found the papers. These pupils have not studied beyond the level expected in the papers, so do not have the issues of adults' additional knowledge and skills which could make papers seem easier.

In this instance the discussion of papers was directed towards similarities and differences in the papers, the skills required by the papers, and the purposes of them. The pupils identified differences in skills such as the need to learn the text in the 1930's paper versus needing the skills to analyse it in the papers today; however, they found it difficult to agree on the difficulty of the papers. Some pupils thought the memorisation required would make the 1930's paper easy, whilst others thought that would make it much harder.

A drawback of this sort of study is that only small extracts from the discussion were reported in the programme. That makes it difficult to know whether the extracts were representative of the discussion as a whole, or whether they were chosen to illustrate particular points that the editors wanted to highlight. Whilst this programme concluded that some pupils preferred the old exam, it did not say what proportion of the pupils this represented, nor did it go into detail about their reasons for this preference. In addition, as the discussion only took place in one classroom, it is impossible to generalise that the pupils' experiences of the papers would be the same for pupils in all schools. These small discussions are not able to produce firm conclusions about the difficulty of papers, but they are useful in drawing attention to the differences in style and purpose of the papers for the general public.

O level papers

Comparisons using past exam papers are not limited to radio programmes. In 2008 'The O Level Book' (Anon, 2008) was published, containing a collection of past O level papers from 1955 to 1959. Readers were challenged to attempt the papers in a variety of subjects and compare their answers to those provided by experts in the subject. The book formed the basis for an article in the *Times* (Griffiths, 2008).

The book is useful for such comparisons as it contained complete papers, rather than selections of one or two questions. However, closer inspection reveals that the so-called 'complete papers' were actually a collection of questions taken from different years. Whilst they probably retained the structure of the original papers, they may not have been representative of the real challenge. The foreword and editor's notes made reference to the challenge of the O levels, describing them as a "...stinkingly hard, fact-based exam...", and the questions as "...doable, if tough." This suggests that there may have been deliberate selection of difficult questions.

There was an attempt to account for the differences between O levels and today's exams. In the foreword, an interesting comment was made

that at O level it was the number of subjects that mattered, not the grades achieved, which was contrasted with the situation today. The editor's note also commented on changes to the context and content of the exams, drawing attention to changes in teaching and examining, as well as the more obvious changes in content for subjects such as science and history. These are important observations in the context of comparing standards, as all these things will affect the experience of students sitting the exam.

In the *Times* article by Griffiths mentioned above, two of the examination papers in the book were used to test the claim that exams have been dumbed down. Five GCSE pupils sat English and Mathematics O level papers taken from the book in examination conditions just after they had completed their GCSE exams. The teenagers quoted in the article seemed to suggest that the O level papers in mathematics were more challenging, and this was backed up by their results. All of them were predicted Bs and above in their GCSEs, yet only two pupils achieved a pass mark in the Mathematics exam. Their reactions to the difficulty of the English papers was mixed, but none obtained the highest grade in English, despite several of them being predicted As and A*s at GCSE.

These results might seem to confirm that GCSEs are easier than O levels, but there are other factors influencing the results. Whilst the students had studied the subjects recently, they had not received any teaching or preparation for the O level papers. The style of the questions was not the same as a GCSE's, nor were the tasks required of them identical. In English they were asked to summarise passages and explain the meanings of different parts of speech (pronoun, conjunction). Whilst the students did not comment on the content of the mathematics exam, they did mention that they could not use calculators. Their deputy headmistress, who was quoted in the article, acknowledged these differences and several other contributing factors, but nevertheless concluded that the O level papers were harder.

The book of O level questions provides an interesting resource for comparison, but in our opinion there are too many varying factors of unknown effect for a conclusion to be drawn about the relative difficulty of exams today and in the 1950s.

Discussion

There are two key strengths to the type of study discussed in this article. First, they are often able to reach a much larger and broader audience than academic papers can. Secondly, they encourage debate in these areas, which is important. However, these sorts of studies also have weaknesses – the most crucial of which is that, just like academic studies of long-term standards over time, it is not possible to control for all the changes in social context. With the exception of the RSC study, they rely on 'case study' approaches: these have advantages in extracting a rich description of the issues, but are a shaky foundation on which to make generalised statements about national exam standards.

The academic community needs to find a better way of reaching more people, and a way of describing comparability in a clear way. Some of the studies described in this paper are useful in illuminating issues which might be overlooked in more academic research. For example, the depiction of teaching methods from the 1950s in "That'll Teach 'em", brought alive the differences in context in a way that would be difficult to achieve in an academic paper. On the basis of these studies, attracting a wider audience seems to rely on the use of a broad range of media, and

on a simplification of the issues. The former is likely to be more readily accepted by the academic community than the latter.

"That'll Teach 'em", probably helped public perceptions because it illustrated the issue of contextualisation in a dramatic way. Readers of the pieces by journalists who re-sat qualifications also gained a greater insight into the complications of the issue and of the fact that long-term comparisons of standards over time are not straightforward, either to conceptualise or to interpret.

One issue, particularly evident in the RSC study, is the relationship of newspaper headlines to the outcomes of a study. There is clearly a tension between a headline accurately representing the content of a report in a handful of words and the need for a headline to sell a story. It can be the case that a headline is far removed from the data on which it is based, and in these cases there is a danger that the benefits from gaining readership are then lost in the misrepresentation of the research.

Comparing standards over time is beset by limitations – the effects of changes in technology, social expectations, culture, educational priorities and knowledge all have to be taken into account when making these types of comparison. The five examples used in this article show that clearly, as do studies from elsewhere within the educational research community. If there is a common theme to be found running through all the studies it is that there is more to standards of time research than at first meets the eye.

These five examples highlight how the 'standards over time' debate has been taken up by television, professional associations, newspaper and radio. It can be argued that the motivation for these studies has moved from the 'contribution to knowledge' of academic research towards viewing and listening figures, newspaper sales, and government lobbying. All acknowledge the complexities of the standards issue, but perhaps rely too much on over-simplification in order to reach a wider audience.

References

- Anon (2008). *The O level book: genuine questions from yesteryear*. St Ives: Michael O'Mara Books Ltd.
- Griffiths, S. (2008). GCSEs v O-levels, let the battle begin. *The Times*, 6 July [online] <http://www.timesonline.co.uk/tol/news/uk/education/article4275054.ece> (Accessed 25 May 2011)
- Groskop, V. (2010). Why I took an A-level at 37. *The Guardian*, 17 August [online] <http://www.guardian.co.uk/global/2010/aug/17/took-a-level-at-37?INTCMP=SRCH> (Accessed 24 May 2011)
- Lebus, S. (2009). Are A levels getting easier? Speech made to the Global Student Education Forum (GSEF). Wednesday 20th May 2009. http://www.dailymotion.com/video/xghrr7_gsef-talk-q-a-simon-lebus_news, accessed on 12.2.
- Riddell, P., Sieghart, M. A., Boyes, R., & Wighton, K. (2008). A levels: put to the test. *The Times*, 14 August, [online] <http://www.thetimes.co.uk/tto/life/families/article1758015.ece> (Accessed 24 May 2011)
- Royal Society of Chemistry (2008). The Five-Decade Challenge. Research report. Retrieved from http://www.rsc.org/images/ExamReport_tcm18-139067.pdf on 19th April 2011.
- The Daily Mail (2008). The proof that science exams are easier now. *Daily Mail*, 27 November 2008, p.12.
- The Daily Mail (2010). So easy a five-year-old has passed and a seven-year-old got an A star: Record numbers achieve GCSE top marks. 25 August, [online] <http://www.dailymail.co.uk/news/article-1305746/GCSE-results-2010-Students-future-uncertain-university-rejects-snap-college-places.html#> (Accessed 25 May 2011)

The Daily Telegraph (2008). Dumbed-down science is 'failing a generation and will end in catastrophe' Experts warn of 'catastrophic' science teaching. *Daily Telegraph*, 27 November 2008, p.16.

The Daily Telegraph (2010). GCSEs hit new high as experts criticise tests. 23 August, [online] <http://www.telegraph.co.uk/education/educationnews/7958673/GCSEs-hit-new-high-as-experts-criticise-tests.html> (Accessed 25 May 2011)

The Guardian (2008). School science standards are slipping, says study. *Guardian*, 27 November 2008, p.14.

The Independent (2008). 0% What this year's top science pupils would have got in 1965. GCSE students flunk past papers in experiment that exposes decline in standards. *Independent*, 27 November 2008, p.45.

The Today Programme (2009). BBC Radio 4. Thursday 31st April.