



- 1 **Foreword** : Tim Oates
- 2 **Editorial** : Tom Bramley
- 3 **100 years of controversy over standards: an enduring problem** : Gill Elliott
- 9 **A guide to comparability terminology and methods** : Gill Elliott
- 20 **A level pass rates and the enduring myth of norm-referencing** : Paul Newton
- 27 **Subject difficulty – the analogy with question difficulty** : Tom Bramley
- 34 **Comparing different types of qualifications: an alternative comparator** : Jackie Greatorex
- 42 **Linking assessments to international frameworks of language proficiency: the Common European Framework of Reference** : Neil Jones
- 48 **The challenges for ensuring year-on-year comparability when moving from linear to unites schemes at GCSE** : Mike Forster
- 52 **The pitfalls and positives of pop comparability** : Nicky Rushton, Matt Haigh and Gill Elliott

If you would like to comment on any of the articles in this issue, please contact Tom Bramley.
Email:

researchprogrammes@cambridgeassessment.org.uk

This issue and previous issues of *Research Matters* are available on our website:

[http://www.cambridgeassessment.org.uk/ca/](http://www.cambridgeassessment.org.uk/ca/Our_Services/Research)

[Our_Services/Research](http://www.cambridgeassessment.org.uk/ca/Our_Services/Research)

Research Matters: Special Issue 2: Comparability

A CAMBRIDGE ASSESSMENT PUBLICATION

Foreword

Comparability is an area beset by assumptions, trammelled by methodological dispute, and regarded, by some, as a bankrupt pursuit. For the public, authentication of claims that the standards, in the 'same' qualification in different years, and in different qualifications which claim equivalence, is vital. But this need for confidence frequently is accompanied by ill-grounded assumptions. It may feel self-evident that standards should, in all circumstances, be maintained over time, but standards do NOT have to be maintained over time, with precision, in *all* assessments. They may have to be in the case of qualifications such as A levels. Here, different candidates, who have gained their qualifications in different years, are competing for university places in the same year. In these admissions processes, an 'A' grade from different sessions is treated as if it means the same thing – so it needs to mean the same thing. Contrast STEP assessment – a maths admissions test. STEP is a purely norm-referenced assessment. The standard is not exactly the same each year. But in each year it remains a highly effective instrument for rank ordering candidates – and thus helps universities identify the top performing pupils. But the consequence of this is that a pupil cannot use a score gained in 2009 in 2011 admissions processes – if applying in 2011, they need to take the test again (not a bad idea in fact...have they retained their mathematical understanding?). Yet another example calls many assumptions into question: there are instances of standards in qualifications deliberately being varied, over time, in order to achieve change in education systems. Here, comparability can be used not to secure equivalence, but to understand the magnitude of change. Is comparability work always concerned principally with standards? No; in some cases comparability can be focused on qualitative comparison of the *focus, structure or purpose of different qualifications*. So the assumptions in comparability need clarification.

So too the methods. Methods are hotly contested and, again, beset by assumptions. Some studies have not considered the quality of work produced by candidates. Others consider candidate work but take so long and are so expensive that the results of the analysis come too late to inform action, and draw resource away from more pressing research. Yet others yield interesting results, but appear compromised by assumptions that teaching has not improved, that successive cohorts of candidates have remained equally motivated, and so on. Perhaps comparability should indeed be seen as an analogue of the pursuit of the Holy Grail – ultimately fruitless in attaining the assumed goal, but the real worth is in striving to attain it. In other words, the fact that comparability work is unlikely to yield flawless conclusions of ultimate precision does not mean that we should not strive ceaselessly to improve our techniques.

This edition emphasises the need to be clear about definitions of comparability, to specify precise objectives, to be discriminating in respect of choice of method, and to understand the utility and limitations of findings. Our conclusion is that comparability is NOT a bankrupt activity. It is complex, demanding (both theoretically and practically), and frequently produces indicative rather than definitive findings. But it remains a vital part of both management of, and research on, qualifications and assessments. For as long as we and others make claims about equivalence, we need to strive to put in place robust, practical, and cost-efficient processes for understanding and authenticating those claims. Dark art or science? If science involves review, critique and refinement of method, alongside constant reconceptualisation and redefinition, then yes, it's science.

Tim Oates *Group Director, Assessment Research & Development*