



- 1 **Foreword** : Tim Oates, CBE
- 1 **Editorial** : Tom Bramley
- 2 **To “Click” or to “Choose”?**  
**Investigating the language used in on-screen assessment** : Rushda Khan and Stuart Shaw
- 9 **Articulation work: How do senior examiners construct feedback to encourage both examiner alignment and examiner development?** :  
Martin Johnson
- 15 **Characteristics, uses and rationales of mark-based and grade-based assessment** : Joanna Williamson
- 22 **Is comparative judgement just a quick form of multiple marking?** :  
Tom Benton and Tom Gallacher
- 28 **How have students and schools performed on the Progress 8 performance measure?** : Tim Gill
- 38 **Research News** : Karen Barden

If you would like to comment on any of the articles in this issue, please contact Tom Bramley, Director, Research Division. Email: [researchprogrammes@cambridgeassessment.org.uk](mailto:researchprogrammes@cambridgeassessment.org.uk)

This and all previous issues of *Research Matters* are available to download, in full and by individual article, from our website: [www.cambridgeassessment.org.uk/research-matters](http://www.cambridgeassessment.org.uk/research-matters)

# Research *Matters* / 26

A CAMBRIDGE ASSESSMENT PUBLICATION

## Foreword

Observing the ebb and flow of contemporary international discussion about curriculum, it sometimes is hard to see scientific discussion rather than political discourse. We should continue to strive to lay down the principles and practices of the science of measurement, particularly as technologically driven change creeps into day-to-day practices. Globally, accountability systems are becoming more evident, and where exam and test results are available, these become a natural focus of attention. The UK now has nearly three decades of experience of using examinations outcomes as a principal basis of public accountability, and it is right that we continue to try to bring measurement science to this highly politicised area.

The Progress 8 measure, which is the focus of Tim Gill's research, is a genuine attempt by the UK Government to introduce both greater validity and sensitivity into target setting and performance measurement. It rightly aims to remedy the serious defects of the simple performance measures introduced in the 1990s, and which persisted despite known severe limitations and highly adverse 'backwash' effects. The research findings – that schools with lower attaining intake are less likely to perform well on the Progress 8 measure – has led to calls from various organisations to abandon Progress 8 or indeed, abandon accountability measures based on assessment outcomes. Many of the comments miss a key point in measurement science – that specific measures of curriculum attainment (which also function indirectly as weak measurement of engagement and motivation) cannot measure everything. Not all the factors which give rise to the findings in the research are necessarily outside of schools' control, or immune from the impact of other aspects of supportive government policy. For example, Daniel Hamlin and Paul E. Peterson's work in the USA shows that hard-bitten schools in deprived areas are beginning to attract higher performing teachers, by changing their ethos and curriculum principles. We need to keep the science in with the politics.

**Tim Oates, CBE** *Group Director, Assessment Research and Development*

## Editorial

In the first article of this issue of *Research Matters*, Rushda Khan and Stuart Shaw give a detailed discussion, illustrated by real examples, of some of the issues that need to be considered when preparing on-screen versions of exam questions. I was particularly interested by their observations about the metaphors of working on-screen (files, folders, notepads) and the implications for instructions to examinees such as “write” or “type”. The second article by Martin Johnson, based on his PhD thesis, uses a detailed analysis of inter-examiner communications to show what a tricky task is faced by Team Leaders when giving feedback that potentially criticises the marking of their team members (who are often experienced teachers and examiners themselves).

In the third article, Joanna Williamson considers the advantages and disadvantages of directly grading assessments (as opposed to marking them and then defining cut-points on the mark scale that identify the different grades). This work is particularly relevant and timely as T Levels<sup>1</sup> are introduced in 2020 – decisions about how their various components are to be assessed could have a big impact on whether they achieve their intended purpose.

There has been a lot of recent interest in using comparative judgement (CJ) instead of traditional marking to assess students' work in certain domains, such as essay writing. Because CJ differs from marking in several important respects, it is difficult to compare them on a like-for-like basis. In the fourth article, Tom Benton and Tom Gallacher argue that much of the apparent advantage of CJ (in terms of its value in predicting scores on other assessments) comes from the simplicity of the marking task combined with using a statistical model to iron out differences in severity among markers. One implication is that traditional marking can be just as effective as CJ if it is combined with simpler mark schemes, multiple marking, and statistical scaling of examiners' marks.

Finally, Tim Gill uses statistical modelling to explore the relationship between Progress 8 scores and variables at school and student level. He concludes that many relevant variables are beyond the control of schools, and hence that it is questionable whether Progress 8 should be used as a basis for comparing them.

**Tom Bramley** *Director, Research Division*

1. See <https://www.gov.uk/government/publications/introduction-of-t-levels/introduction-of-t-levels>