



- 1 **Foreword** : Tim Oates, CBE
- 1 **Editorial** : Tom Bramley
- 2 **Accessibility in GCSE Science exams – Students' perspectives** : Victoria Crisp and Sylwia Macinska
- 10 **Using corpus linguistic tools to identify instances of low linguistic accessibility in tests** : David Beauchamp and Filio Constantinou
- 17 **A framework for describing comparability between alternative assessments** : Stuart Shaw, Victoria Crisp and Sarah Hughes
- 23 **Comparing small-sample equating with Angoff judgement for linking cut-scores on two tests** : Tom Bramley
- 27 **How useful is comparative judgement of item difficulty for standard maintaining?** : Tom Benton
- 37 **Research News** : Anouk Peigne

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* / 29

A CAMBRIDGE ASSESSMENT PUBLICATION

## Foreword

'Improved accessibility' has been vigorously pursued in contemporary assessments in England, and has featured in other many nations' discussions of fairness and bias. Perhaps it would better be described as 'removal of sources of construct-irrelevant score variation'. This better describes the relation between item quality and item purpose. Increasing assessment quality for those with sight or other impairment is essential where irrelevant features of items (font size, font type, colour, etc.) can readily be adjusted to improve the measurement properties of items and assessments. But some work on accessibility can impact adversely on the measurement properties of items and assessments. For example, Isaac Physics – its leading developers recently rewarded with a prestigious gold medal at the 2019 Institute of Physics awards – has highlighted how some efforts to improve accessibility (providing a diagram where none was expected before; breaking a question down into steps; providing equations) can materially impact on standards of demand, with negative washback into learning. 'Improving accessibility' is not some discrete and pre-eminent concern in assessment, since pursuing some accessibility aims can have a very specific, adverse impact on standards of demand. As is so often the case in assessment, complex things are entwined in complex ways. The best policy scenario is that the tension between enhanced accessibility and maintenance of standards is held in careful balance. The worst scenario is that the tensions lead to oscillations in priorities, and the resulting pendulum swings create precisely the kind of constant change in qualifications which educational professionals and candidates find disruptive and frustrating. The tensions will never go away; but sound and careful management means that adverse effects can be minimised.

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

## Editorial

Writing good exam questions is a difficult art. We need questions that elicit responses that demonstrate the relevant knowledge, skills and understanding (KSU). We want to avoid anything that hinders or prevents examinees with the relevant KSU from demonstrating this, so we should make the questions as accessible as possible without reducing their subject-related demands. The first two articles in this issue are about accessibility. The first, by Victoria Crisp and Sylwia Macinska, describes students' reactions to 'before and after' versions of exam questions that had various features modified in ways that were hypothesised to affect accessibility. The second, by David Beauchamp and Filio Constantinou, explores the potential of automated tools to give insights about the linguistic complexity of the words and sentences in exam questions.

In the third article, Stuart Shaw, Victoria Crisp and Sarah Hughes describe a rigorous but practical approach that could help practitioners to investigate the comparability of alternative assessments. Their framework distinguishes different kinds of standard and helpfully recognises that an overall informed judgement is required about the extent to which differences in comparability matter.

The final two articles, by me and Tom Benton respectively, explore an issue that is of perennial interest to assessment developers – namely the extent to which expert judgement about the difficulty of exam questions can give useful information about the relative difficulty of two exams as experienced by the examinees. The conclusions are somewhat pessimistic, but no doubt this will not be the last word on this topic!

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