



- 1 **Foreword** : Tim Oates
- 1 **Editorial** : Sylvia Green
- 2 **An investigation into marker reliability and some qualitative aspects of on-screen marking** : Martin Johnson and Rita Nádas
- 8 **How effective is fast and automated feedback to examiners in tackling the size of marking errors?** : Dr Elizabeth Sykes, Dr Nadežda Novaković, Dr Jackie Creatorex, John Bell, Rita Nádas and Tim Gill
- 16 **Mark scheme features associated with different levels of marker agreement** : Tom Bramley
- 23 **Thinking about making the right mark: Using cognitive strategy research to explore examiner training** : Dr Irenka Suto, Dr Jackie Creatorex and Rita Nádas
- 32 **Capturing expert judgement in grading: an examiner's perspective** : Peter King, Dr Nadežda Novaković and Dr Irenka Suto
- 34 **Investigation into whether z-scores are more reliable at estimating missing marks than the current method** : Peter Bird
- 43 **'Happy Birthday to you'; but not if it's summertime** : Tim Oates, Dr Elizabeth Sykes, Dr Joanne Emery, John F. Bell and Dr Carmen Vidal Rodeiro
- 45 **Cambridge Assessment Parliamentary Research Seminar Series – Better training: Better teachers?** : Sylvia Green
- 47 **Research News**

If you would like to comment on any of the articles in this issue, please contact Sylvia Green.

Email: [researchprogrammes@cambridgeassessment.org.uk](mailto:researchprogrammes@cambridgeassessment.org.uk)

The full issue and previous issues are available on our website:  
[www.cambridgeassessment.org.uk/ca/Our\\_Services/Research](http://www.cambridgeassessment.org.uk/ca/Our_Services/Research)

# Research Matters : 8

A CAMBRIDGE ASSESSMENT PUBLICATION

## Foreword

One key question keeps recurring: '...do you do fundamental work or is it all instrumental...?'. My own answer: '...both... at different times and sometimes entirely wrapped up together...'. An instance of fundamental work not tightly linked to operational work in the Group is the literature review on Birthdate Effect reported here under the title 'Happy Birthday to you – but not if it's summertime'. A seemingly-humble and increasingly unfashionable literature review, it cuts through the restrictions of cross-sectional studies – which look at individual phases of education and training – to illuminate the shocking persistence of the effect across the system as a whole. But much of the work in this volume exemplifies fundamental and operational work which is 'wrapped up together'. All too often, there exists an assumption that there is a contradiction between fundamental research and operational research. But in the most parochial of matters in assessment there lurk fundamental matters. And fundamental work can – and should – be used to drive improved 'evidence-based' practice. Perhaps we should take a lead from medical research – clinical practice of the most applied kind (ranging from surgical techniques to genetic counselling) – which is progressed by sound fundamental work. I characterise it as the 'janus-headed' nature of assessment research – looking both to enhance the canon of fundamental, generalisable knowledge and simultaneously improve the operation of complex public systems. Whilst tensions might arise in terms of issues such as '...onto the next project or do I disseminate well the outcomes of my existing work...?', the parallels with medical research (and with aeronautical engineering, meteorology, etc) suggest that those that preach an absolute distinction between fundamental and applied work will starve assessment systems of assessment of knowledge essential to their improvement. Look to the studies in this volume as examples of being 'janus-headed' in the best possible way.

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

## Editorial

Most of the articles in this issue relate to how judgements are made and the factors that impact on those judgements. In the first article Johnson and Nádas consider how on-screen essay marking is affected by mode. Reliability is also a focus in the article from Sykes *et al.* which reports on an investigation into the effectiveness of potential procedures for providing fast and automated feedback to examiners. Bramley's article concentrates on marker agreement at item level rather than at candidate level. He reports on a study that explored the features of question papers and mark schemes associated with higher and lower levels of marker agreement. Suto, Creatorex and Nádas consider the benefits of, and variations in, training procedures for examiners, drawing together research on examiner training and on the nature of the judgements entailed in the marking process. In the article from King, Novaković and Suto we learn how judgements are made using rank ordering, traditional awarding, and Thurstone pairs. The article focuses on the perspective of an examiner who took part in the research and his views are extremely valuable in informing the design of future research. Peter Bird, a member of OCR's operational research team, compares two methods of estimating missing marks and highlights issues and differences in the accuracy of the process. The article on the effect of birthdate on performance by Oates *et al.* outlines the findings of a research review which provides robust evidence from around the world that, on average, the youngest children in their year group at school perform at a lower level than their classmates. The review detailed in this article was released to the press in February 2009. It was widely reported in England and has also received attention in other countries, including China. At the same time it was submitted as evidence to the Rose Review of Primary Education which, as part of its interim report, had recommended that all children should start formal schooling at the age of four (rather than five, as is currently the case). Sylvia Green then reports on the latest in the Cambridge Assessment Parliamentary Research Seminar Series, hosted by Barry Sheerman MP, Chair of the Children, Schools and Families Select Committee.

**Sylvia Green** *Director of Research*