



- 1 **Foreword** : Tim Oates
- 1 **Editorial** : Sylvia Green
- 2 **Exploring non-standard English amongst teenagers** : Beth Black
- 11 **The evolution of international History examinations: An analysis of History question papers for 16 year olds from 1858 to the present** : Stuart Shaw and Gillian Cooke
- 19 **The reliabilities of three potential methods of capturing expert judgement in determining grade boundaries** : Nadežda Novaković and Irenka Suto
- 24 **How do examiners make judgements about standards? Some insights from a qualitative analysis** : Jackie Greatorex
- 33 **'Key discriminators' and the use of item level data in awarding** : Tom Bramley
- 39 **Statistical Reports** : The Statistics Team
- 40 **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

The full issue and previous issues are available on our website:
www.cambridgeassessment.org.uk/ca/Our_Services/Research

Research Matters : 9

A CAMBRIDGE ASSESSMENT PUBLICATION

Foreword

Once again, *Research Matters* is testimony to the breadth of focus and method which is present in research across Cambridge Assessment. The editorial rehearses the detail of the wide variety of approaches taken but what readers might miss is the crucial nature of the statistical reports listed at the rear of this issue. These are all available on the Cambridge Assessment main website and, in my view, make an essential contribution to the transparency of the examinations system. For a few years, QCA published time series data on GCSE and A level/AS results – the grade profile for almost all subjects, across all boards, in aggregate and also broken down by gender. QCA stopped publishing these in late 2006. These simple data – the numbers gaining the respective grades each year – are important enabling exploration of so many key issues: trends in the numbers getting A and A* grades; the low number of females doing some subjects (notably physics); and so on. Simple yet vital as we need to understand trends, and have comparable year-on-year data. Of course, they highlight some difficult issues and thus enable some hard questions to be asked. This, I believe, is a good thing. A better 'base level' of information allows journalists to explore issues with greater precision and understanding, and I would hope that young people, parents and other groups will be able to understand better the complex system we call 'education and training' and engage with the qualifications system on a more informed basis. The removal of these time series data from the QCA website marked a retrograde step in public accountability. By restoring these data to the public domain and presenting them in a clear and simple format, Cambridge Assessment has sought to increase transparency and make a genuine contribution to public accountability.

Tim Oates *Group Director, Assessment Research and Development*

Editorial

In this issue the themes range from the awareness and usage of non-standard English among sixteen year olds to examiner judgements in awarding. In the opening article Beth Black adds to the empirical research base on non-standard written English among young people at GCSE level. A questionnaire/assessment instrument was used to explore students' awareness of non-standard English and to investigate differences between school types, gender differences and regional differences.

Shaw and Cooke take us from 1858 to the present day with an analysis of history question papers for 16 year olds. They also used a variety of archive material to show more general developmental changes to the curriculum throughout the period. This article gives an interesting perspective on changes over time in both the structure and language of the papers as well as the marking schemes and processes of assessment.

The next three articles focus on expert judgements and the methods that impact on them. Novaković and Suto investigate the reliabilities of three potential methods for capturing expert judgement. These include traditional awarding (currently used), Thurstone pairs and rank ordering. A three-way comparison of the intra-method and inter-method reliabilities of all three methods was conducted in the context of setting grade boundaries. This research provides some interesting insights into the different judgemental methods, one of which is the current method used and two which could be useful in the future. In her article on how examiners make judgements about standards using different methods Greatorex used 'think aloud protocols' which involved Principle Examiners verbalising their strategies. The qualitative think aloud data were analysed using a framework designed for the purpose. This article offers many insights into what Principle Examiners attend to when they make judgements about grading standards.

In the final article Bramley discusses the theoretical rationale for using item level data in awarding. He presents some possible formats for displaying data and suggests ways in which the data could be used in practice. Data on individual questions or question parts are collected automatically as examination papers are scanned and marked on screen. These new processes provide a wealth of data that can be used to investigate how items function and how key discriminators can be used in awarding processes.

The Statistical Reports Series provides statistical summaries of a range of information using national-level examination data. The Factsheets are designed to make our research accessible to a wider audience 'headlining' main findings. Full reports can be found in the 'Conference Papers' section of our web-site.

Sylvia Green *Director of Research*