very clearly the thinking behind the decisions that were made. Tracing the outcomes of those decisions, through the legislation which followed, and into policy and practice can inform current educational debates, particularly in instances where consideration is being made of similar initiatives to those which have gone before. It is in these instances that it is possible to be informed by hindsight.

References

Aldrich, R. & Gordon, P. (1989). *Dictionary of British Educationalists*. London: Woburn.

Apprenticeship, Skills, Children and Learning Act, 2009. (c.22). London: HMSO.

Board of Education (1911). *Report of the Consultative Committee on Examinations in Secondary Schools*. London: HMSO.

Children, Schools and Families Committee (2008). 'Testing and Assessment' HC (2007–08) 169–1.

UK Parliamentary Services (2009). Hansard (1803–2005). Available at http://hansard.millbanksystems.com accessed on 14.2.2011.

Hartley, C. (Ed.) (2003). *A Historical Dictionary of British Women*. London: Europa Publications Ltd.

Report of the Consultative Committee (1923). *Differentiation of the Curriculum for Boys and Girls Respectively in Secondary Schools*. Reproduced in D. Gillard, (2006), *The Hadow Reports: an introduction*. www.educationengland.org.uk/articles/24hadow.html

The Conservative Manifesto (2010). Available at http://media.conservatives.s3.amazonaws.com/manifesto/cpmanifesto2010_lowres.pdf accessed on 14.2.2011

The Importance of Teaching – The Schools White Paper, 2010. (CM 7980). London: HMSO.

Transactions of the Royal Historical Society (1945). *Back Matter*. Fourth Series, 27, 97–106. Published by: Royal Historical Society.

Watts, A. (2008). Cambridge Local Examinations 1858–1945. In: S. Raban (Ed.), *Examining the World. A History of the University of Cambridge Local Examinations Syndicate*. Cambridge University Press: Cambridge.

ASSURING QUALITY IN ASSESSMENT

Evaluating Senior Examiners' use of Item Level Data

Hannah Shiell and Nicholas Raikes Research Division

Many of CIE and OCR's written examination scripts are now scanned and marked on screen by examiners working on computers. One benefit arising from on-screen marking is that the marks are captured at item or question-part level and are available for analysis in Cambridge within hours of being submitted by examiners. Cambridge Assessment now routinely analyses these item marks and provides subject staff and senior examiners with reports containing Item Level Data (ILD) for nearly all examinations marked on screen. In this article we present findings from an evaluation of senior CIE and OCR examiners' use of these Item Level Data reports.

Background

Historically, CIE and OCR's written examinations were marked on paper and usually only the total marks were captured electronically. Consequently, if item marks were to be analysed they first had to be keyed in from a sample of written scripts, and this constrained the availability of item level data. With the introduction of on-screen marking, however, marks are now routinely captured at item level for a large and growing number of CIE and OCR's written examinations.

In addition to introducing on-screen marking, Cambridge Assessment has made a major investment in infrastructure to provide research and evaluation staff with:

- a data warehouse providing easy access to operational data, including item marks;
- statistical analysis and reporting tools;
- automation tools (for automating and scheduling analysis and reports);
- an Intranet Portal for publishing statistical reports and data to colleagues across the organisation.

This new infrastructure has enabled us to start routinely producing ILD reports for most CIE and OCR examinations marked on screen. An indication of the scale of this activity is that during peak periods last summer (2010) we analysed 60 million marks per night across nearly 600 examinations.

The nature of the Item Level Data provided

Previous work in Cambridge Assessment identified the kinds of Item Level Data and presentation most useful to subject staff and senior examiners (Johnson, Gill, Elliot and Black, 2006).

We now produce ILD reports on two occasions: firstly during marking, then again after grade boundary marks have been set and candidates' grades are known. The first set of reports are provided to assist subject staff and senior examiners with tasks relating to the current examination, such as providing reports on the candidature's performance and recommending grade threshold marks. The second set of ILD reports, provided once marks have been finalised and candidates' grades determined, are to assist with post-hoc evaluations of the examinations to help identify any improvements that can be made in future examinations. ILD reports are made available as web pages on our Intranet Portal and as documents in pdf format. Few senior examiners

have access to our Intranet, so electronic copies of the pdf reports are sent to them; they may also be shown ILD when attending meetings at our offices.

The following types of output are produced during marking (all updated nightly):

- item statistics (omit rate, facility overall and by quartile, correlation between item marks and overall marks excluding the item);
- item curves (plots of facility by quartile);
- item mark distributions;
- overall internal reliability (Cronbach's alpha);
- overall mark distribution and summary statistics (mean, standard deviation, minimum and maximum mark, all presented overall and by quartile).

OCR generally sends all this information to senior examiners, but CIE initially only sends the item statistics, supplying other information on request since it can amount to many pages of output.

Similar output is produced once candidate grades are known, but this time including grade distributions and breaking information down by grade, sex and, for CIE, country. Sample output can be seen in Figure 1, a screenshot of the item curve and mark distribution chart for one item for one CIE country (details identifying the county and examination have been redacted).

Benefits Review

As part of a wider review of the benefits realised from routinely producing Item Level Data, we solicited feedback from senior examiners.

A short online questionnaire was developed following discussion with subject staff responsible for working with senior examiners using ILD. The questionnaire was reviewed by a panel of researchers not previously involved in the study prior to being used.

We emailed the following senior examiners and invited them to complete the online questionnaire about their use of ILD:

- Principal Examiners and Setters of examinations marked on screen in June 2010 (OCR) and November 2010 (CIE).
  NB: in many cases the same individual was both setter and principal examiner, i.e. he or she both set the question paper and led the marking. The roles are not necessarily combined, however.

Questionnaire findings

The response rate was 71% for CIE (58 responses from 82 invitations) and 59% for OCR (159 responses from 269 invitations).

Some 86% of CIE respondents and 82% of OCR respondents reported that they used ILD.

When asked to assess how helpful they found ILD overall, 78% of CIE respondents and 79% of OCR respondents reported that they found it helpful or very helpful (ratings were made on a five point scale: 2 = very helpful, -2 = very unhelpful). The actual numbers of respondents in each category are shown for CIE in Figure 2 and for OCR in Figure 3.

The senior examiners were also asked to provide feedback on specific uses of ILD. The majority of respondents from both CIE and OCR found ILD helpful or very helpful:
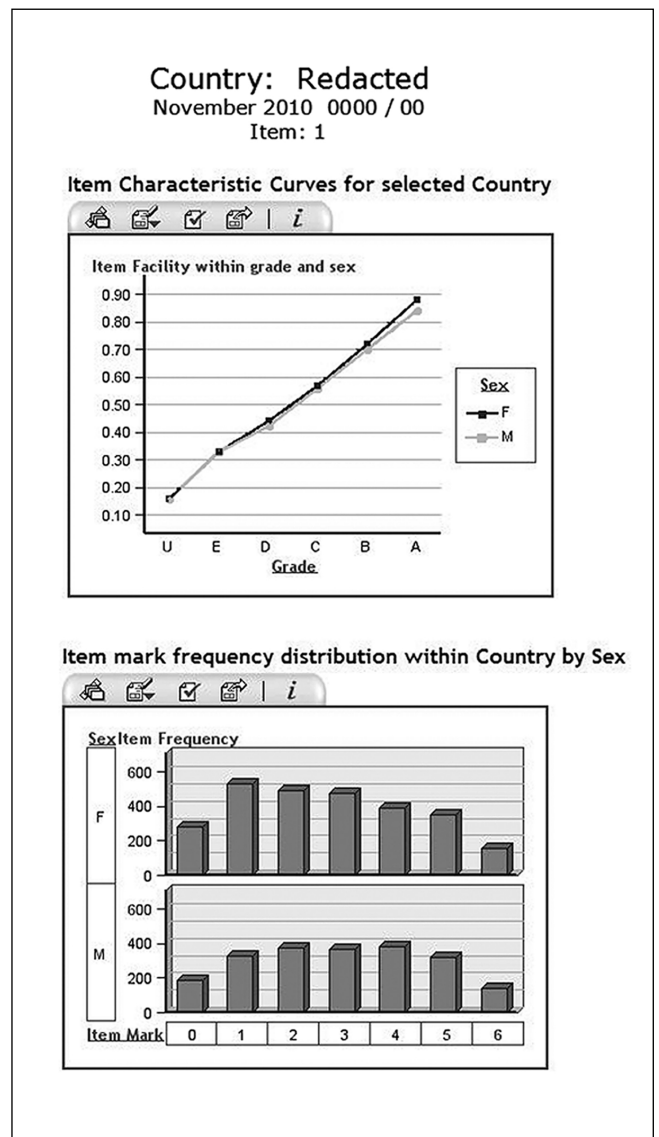

Figure 1: Sample CIE output produced once candidate grades are known

- when writing reports to teachers on candidate performance (see Figure 4 and Figure 5 for the exact number of respondents from CIE and OCR in each category);
- when filling in their 'SRS forms' (Figure 6 and Figure 7). These are the forms on which Principal Examiners make their initial recommendations on where grade boundary marks should be set;
- when identifying items which were harder or easier than expected (Figure 8 and Figure 9), or which did not discriminate as expected between candidates of different 'ability' (as indicated by candidates' total marks) – see Figure 10 and Figure 11.

When asked whether they felt adequately supported in their use of ILD, 76% of CIE respondents and 74% of OCR respondents answered "yes". Given that ILD are relatively novel to many of our senior examiners, this finding is encouraging, though clearly there is scope to improve the support provided. Our current support centres on written documentation explaining each part of the ILD, presentations at meetings, and individual support from subject staff. Improvements suggested by respondents included provision of a separate quick reference glossary of the statistical terms, together with additional written documentation giving examples of use.

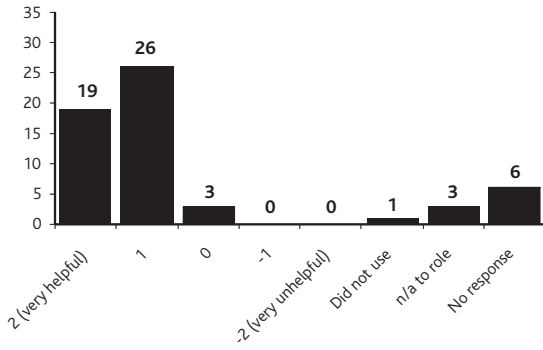In terms of your overall use of ILD, how helpful or unhelpful did you find this data?

Figure 2: Overall usefulness of ILD – CIE Respondents

In terms of your overall use of ILD, how helpful or unhelpful did you find this data?
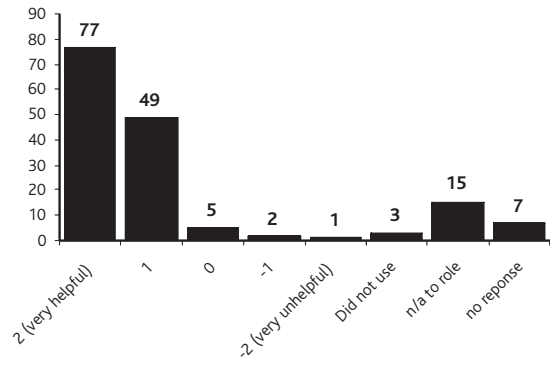

Figure 3: Overall usefulness of ILD – OCR Respondents

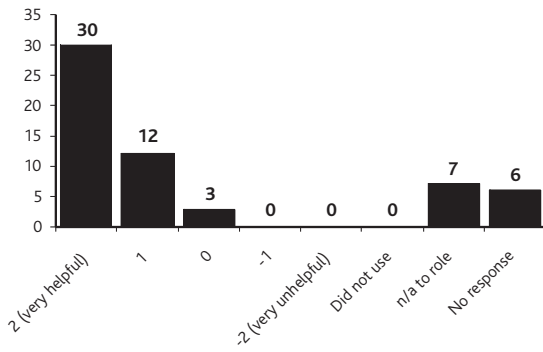To inform writing the PE report to teachers/centres

Figure 4: Use of ILD when writing reports to teachers/centres – CIE respondents
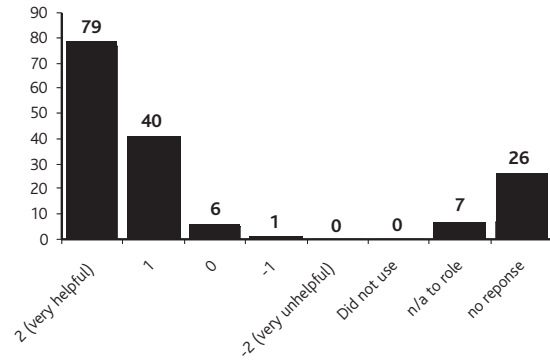
To inform writing the PE report to teachers/centres

Figure 5: Use of ILD when writing reports to teachers/centres – OCR respondents

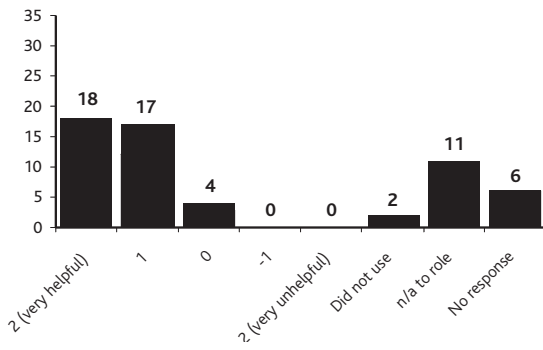To help make PE recommendations on the SRS form

Figure 6: Use of ILD when completing the SRS form relating to recommending grade boundary marks – CIE respondents

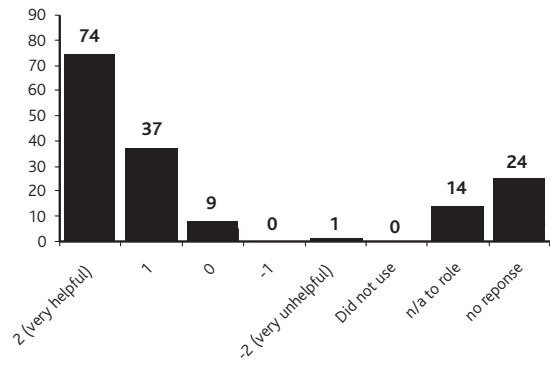To help make PE recommendations on the SRS form

Figure 7: Use of ILD when completing the SRS form relating to recommending grade boundary marks – OCR respondents

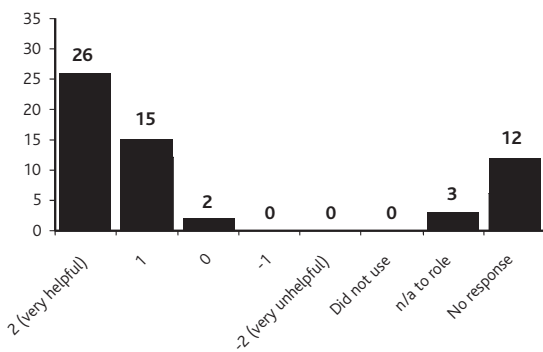Identifying questions that were easier or harder than expected

Figure 8: Use of ILD for investigating question difficulty – CIE respondents

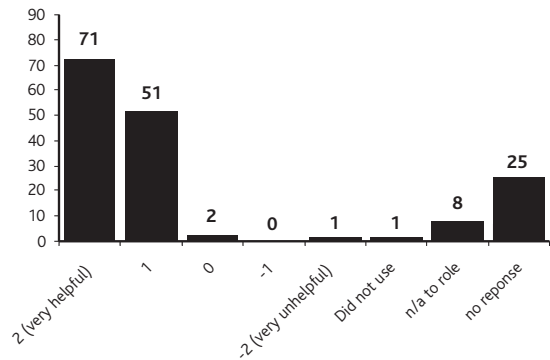Identifying questions that were easier or harder than expected

Figure 9: Use of ILD for investigating question difficulty – OCR respondents
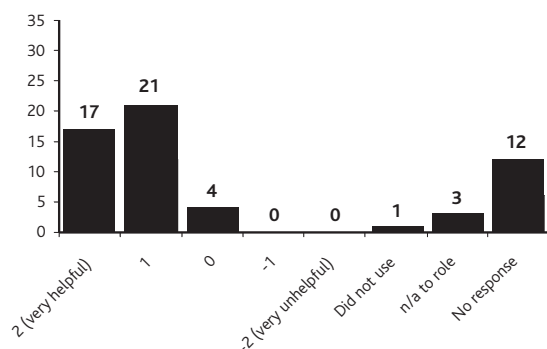
Identifying questions that did not disciminate as expected

Figure 10: Use of ILD for investigating question discrimination – CIE respondents
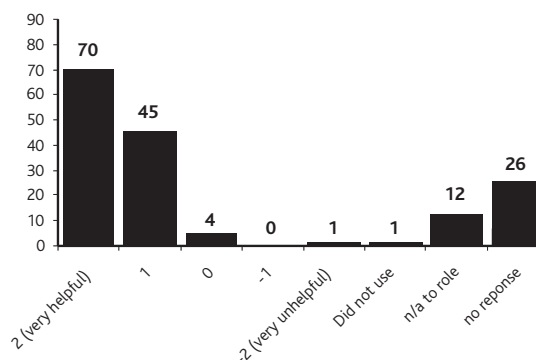

Figure 11: Use of ILD for investigating question discrimination – OCR respondents

Conclusion

The questionnaire findings provide evidence that Cambridge Assessment has successfully introduced routine reporting of Item Level Data to senior CIE and OCR examiners, and that the reports provide helpful information that is widely used. Further work would be required to probe exactly how the information is used. The main limitations of the study are those which generally affect questionnaire-based studies, principally an unquantifiable self-selection bias arising from examiners deciding whether to complete the questionnaire, and the degree to which participants were willing to be open with us and provide accurate and complete answers.

Reference

Johnson, N., Gill, T., Elliott, G. and Black, B. (2006). *Opportunities for evaluating assessments using item level data. Report on 2005–06*. Unpublished internal Cambridge Assessment report.

EXAMINATIONS RESEARCH

Practical issues in early implementation of the Diploma Principal Learning

Victoria Crisp and Sylvia Green Research Division

This short article reports on some of the findings from an interview study conducted in the first year of implementation of the 14–19 Diplomas. The Diplomas were introduced by the Labour government as part of wider educational reforms (DfES, 2005a, 2005b). They were designed to prepare young people for the world of work or for independent study and are intended to combine theoretical and applied learning, to provide different ways of learning, to encourage students to develop skills valued by employers and universities, and provide opportunities for students to apply skills to work situations in realistic contexts. They are also intended to contribute to ensuring that a wide range of appropriate learning pathways are available to young people, thus facilitating increased participation and attainment. The Diplomas are available at Levels 1, 2 and 3 and rather than being taught by an individual school or college, they are available through consortia consisting of a small group of schools and/or colleges working collaboratively.

The Diploma is a composite qualification which is made up of the following elements: principal learning; generic learning; additional and specialist learning. The current research focused on the Principal Learning (PL). The Principal Learning components are specific to a domain or 'line of learning'. Learning through experience of simulated or real work contexts, through applying and practically developing skills, as well as theoretical learning, is emphasised. The PL components are assessed predominantly via assignments which are internally marked and externally moderated. Teaching of Diplomas in the first five 'lines of learning' began in September 2008 with a further five beginning in September 2009 and four in September 2010.

Several initial evaluations of Diploma implementation and other sources have already provided some insights on various issues. One publicly prominent point has been that the uptake of the Diploma was initially lower than expected. The uptake of any course is likely to be strongly affected by whether learners and teachers have a good understanding of that course in order to make informed choices. McCrum *et al.* (2009) interviewed Year 11 students and found that many had limited or incorrect knowledge about Diplomas and that it tended to be