How can mark scheme design support reliable and valid school-based assessment?

Conference Paper Abstract

Joanna Williamson & Simon Child

Author contact details:

Joanna Williamson & Simon Child
Assessment Research and Development,
Research Division
Cambridge Assessment
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA
UK

williamson.j@cambridgeassessment.org.uk
child.s@cambridgeassessment.org.uk

http://www.cambridgeassessment.org.uk

As a department of Cambridge University, Cambridge Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

Abstract

Research background and research questions

Previous research has identified ways in which mark scheme design can support the reliability and validity of examiners' assessment judgements, and, specifically, improve the level of agreement between examiners (Ahmed & Pollitt, 2011; Black, Suto, & Bramley, 2011; Pinot de Moira, 2011, 2013, 2014). Attending to mark scheme design is, therefore, one important way in which assessment quality can be supported and safeguarded in high-stakes national qualifications.

Alongside external assessments such as examinations, national qualifications in many countries make use of school-based assessment, in which assessment tasks are marked, and sometimes set, by students' own schools or colleges (Burdett, 2013; Dufaux, 2012). School-based assessment has the potential to complement examination assessment in a number of useful ways (Vitello & Williamson, 2017). Not least, it is widely recognised that certain constructs – such as the skill of drafting and re-drafting a piece of writing over an extended period of time – are not amenable to assessment by written examination, but can be effectively assessed through well-designed school-based assessment (Johnson, 2013; Ofqual, 2013).

Despite the acknowledged capabilities of school-based assessment, its use within high-stakes national qualifications is contested (Barrance, 2018; Torrance, 2018). Concerns exist about the difficulty of standardising marking and assessment practice across different schools, and the potential for malpractice. In England, recent reforms by the Department of Education have reduced the proportion of school-based assessment used in national qualifications, by mandating higher proportions of external assessment (DfE, 2017), reflecting concerns about the rigour and quality of school-based assessment. Despite this, school-based assessment remains present in national qualifications in England and many other countries, because of the acknowledged need to assess constructs that cannot be effectively assessed in examinations. Supporting the quality of such school-based assessment therefore remains an important priority, particularly where national qualifications have high stakes for both students and schools.

The contribution of this research was to investigate how mark scheme design can be used to support reliable and valid school-based assessment. The distinctive characteristics of school-based assessment mean that it is not immediately clear how far principles of good design for examination mark schemes would apply to school-based assessment mark schemes. The research questions addressed were the following:

1. How are the distinctive characteristics of school-based assessment likely to affect marking task demand and marker agreement, as conceptualised by Black et al. (2011)?
2. What recommendations can be made for school-based assessment mark scheme design?

The theoretical framework for the research was the evidence-based model of marking task demand proposed by Black et al. (2011). This model proposes that marking task demand

and marker expertise are the two overall groupings of factors influencing marker agreement. Where a marking task has higher demand (i.e., is more difficult for the marker to carry out) or markers have lower levels of expertise, there is likely to be a lower level of marker agreement – that is, more variation between the assessment judgements made by different markers. In this model, mark scheme design, along with task features and candidate response features, are the three core factors that influence marking task demand. Mark scheme design is therefore understood to influence marking task demand, and subsequently marker agreement, through interaction with other factors, rather than in isolation.

Methods

The first stage of the research reviewed the characteristics of school-based assessment in comparison with examination assessment. The factors considered were all those in Black et al's (2011) marking task demand model *other* than mark scheme design, and those relating to the context of assessment (e.g., relationship between marker and candidate). The research focused on school-based assessments used in national qualifications for students aged 14-19 in England, such as GCSEs (general academic qualifications) and Cambridge Technicals (applied qualifications). The results of this stage were used to create an adapted model of marking task demand in school-based assessment, incorporating the factors and inter-relations between factors specific to the school-based assessment context.

The second stage of the research reviewed both theoretical and empirical evidence on the impact of mark scheme features, for example, any consistently beneficial effect on marking outcomes as measured by marking reliability. The research considered empirical findings from both external and internal assessment contexts, and evidence was reviewed against the adapted model of marking task demand. In this way, findings that derived from research into examination mark schemes could be critically examined for their likely applicability to a school-based assessment context. This critical examination was necessary since factors that are related to mark scheme design in the marking task demand model, such as task features, differ significantly between examination and school-based assessment contexts. Hence, the relationship between mark schemes and marking task demand, and consequently recommendations for mark scheme design, may also differ. Further evidence that could support or challenge the proposed mark scheme effects was sought from areas of cognitive psychology.

Findings

The first stage of this research resulted in a model of marking task demand in school-based assessment contexts. Distinctive features include the affordance for a degree of control over task design by markers, which can mediate marking task demand, and the potential for markers to develop between-session familiarity that can increase their marking expertise. Year-on-year, feedback from moderation and internal standardisation activities can build up markers' expertise in applying the mark scheme accurately for the same particular assessment task. The unique social characteristics of school-based assessment contexts also have the potential to influence marker agreement. The closeness of marker and candidate provides enhanced opportunities for the marker to understand what is being evidenced by the candidate, and the underlying contributing processes, supporting marking

decisions. However, there is also the potential for marker bias, particularly in the context of strong external pressure to pass students.

The second stage of the research identified a number of recommendations for mark scheme design in school-based assessment, supported to varying degrees by evidence from empirical studies in examination marking, empirical studies in school-based assessment, and theoretical accounts. The available recommendations for each aspect of mark scheme design were summarised, and ordered within the following categories: mark scheme type, structure and layout, mark scheme content, exemplification, formatting, and supporting resources . The summary documentation includes the nature and strength of the available evidence for each recommendation. It is intended that this output can be used by assessment professionals to help evaluate design decisions for school-based assessment mark schemes, as part of an ongoing process of mark scheme improvement to support assessment quality.

References

Ahmed, A., & Pollitt, A. (2011). Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy & Practice, 18*(3), 259-278.

Barrance, R. (2018). *The Fairness of Internal Assessment in National Qualifications*. Paper presented at the ECER 2018, Bolzano, Italty. https://eera-ecer.de/ecer-programmes/conference/23/contribution/44104/

Black, B., Suto, I., & Bramley, T. (2011). The interrelations of features of questions, mark schemes and examinee responses and their impact upon marker agreement. *Assessment in Education: Principles, Policy & Practice, 18*(3), 295-318.

Burdett, N., Houghton, E., Sargent C. and Tisi J. (2013). *Maintaining Qualification and Assessment Standards: Summary of International Practice*. Slough: NFER.

DfE. (2017). *Technical and applied qualifications for 14 to 19 year olds. Key stage 4 and 16 to 18 performance tables from 2020: technical guidance for awarding organisations*. London: Department for Education.

Dufaux, S. (2012). *Assessment for Qualification and Certification in Upper Secondary Education: a Review of Country Practices and Research Evidence (OECD Education Working Papers, No. 83)*. Paris: OECD Publishing.

Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education, 28*(1), 91-105.

Ofqual. (2013). *Review of Controlled Assessment in GCSEs* (Ofqual/13/5291).

Pinot de Moira, A. (2011). *Effective discrimination in mark schemes*. Manchester: AQA.

Pinot de Moira, A. (2013). *Features of a levels-based mark scheme and their effect on marking reliability*. Centre for Education Research and Policy paper. Manchester: AQA.

Pinot de Moira, A. (2014). *Levels-based mark schemes and marking bias*. Manchester: AQA.

Torrance, H. (2018). The Return to Final Paper Examining in English National Curriculum Assessment and School Examinations: Issues of Validity, Accountability and Politics. *British Journal of Educational Studies, 66*(1), 3-27.

Vitello, S., & Williamson, J. (2017). Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England. *London Review of Education, 15*(3), 536-548.