

Using assessment and response times data to evaluate a digital mock exams service

Conference Abstract

Carmen Vidal Rodeiro

Tim Gill

Sarah Hughes

AEA-Europe conference, 6-9 November 2024

Author contact details:

Carmen Vidal Rodeiro & Tim Gill
Assessment Research and Development
Research Division
Shaftesbury Road
Cambridge
CB2 8EA
UK

carmen.vidalrodeiro@cambridge.org
tim.gill@cambridge.org
<https://www.cambridge.org/>

As a department of the university, Cambridge University Press & 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.

Cambridge University Press & Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: [Research Division](#)
If you need this document in a different format [contact us](#) telling us your name, email address and requirements and we will respond within 15 working days.

How to cite this publication:

Vidal Rodeiro, C.L., Gill, T. & Hughes, S. (2024, November 6-9). *Using assessment and response times data to evaluate a digital mock exams service*. [Conference presentation]. AEA-Europe 2024, Cyprus. <https://2024.aea-europe.net/>

Abstract

In recent years, there has been a rapid digital transformation in education and, as a result, the need to rethink how students are assessed has become crucial. New digital tools can have implications for high-stakes assessments, not only for how they are delivered (potentially offering higher personalisation, accessibility, and user engagement), but also because the delivery platforms can capture and store students' activity (e.g., time per question, answer-changing, movement through questions). The use of such "process data" can help quality assurance of tests/items, help understand test-takers' behaviours, engagement and motivation, support operational processes, help malpractice detection, and improve reliability of assessments (e.g., Maddox, 2022).

Cambridge University Press & Assessment has recently launched a Digital Mocks Service for GCSE and A-Level qualifications (UK and international), paving the way for high-stakes digital exams in the near future. Schools in England and around the world can sign up for the service and deliver digital mocks, based on live papers, to their students. To date, on-screen mock assessments have been offered in the following subjects: English Language, Chemistry, History, Global Perspectives & Research and Computer Science.

This work made use of process data to evaluate the Digital Mocks Service and to check that the digital assessments were working as expected. In particular, the delivery platform allowed us, for each subject, to collect data on students' individual responses and performance, as well as total time on the assessment and time taken to engage with each question.

Assessment data based on students' responses and performance (e.g., reliability coefficients, omit rates, facility values, discrimination indices, item marks frequency distributions) helped understand how the items and the assessment worked when delivered digitally. As the majority of the assessments in the Digital Mocks Service were taken live (on paper) previously, it was possible to compare performance in the mocks to the performance in live exams. Response times data was used to check that the time limits set for the digital assessments were appropriate. If candidates do not use a big proportion of the time allowed, maybe the time limit for the assessment is too generous. On the contrary, if candidates do not complete items at the end of the assessment, they might be running out of time. At the item level, response times were used to look for an indication of speededness: if response times diminish with later item positions in the assessment, it could imply that candidates are rushing their responses. Short response times alongside declining performance as students' progress through the assessment might also be associated with students' disengagement. Finally, response times data allowed us to investigate where, within the assessment, students spent most of their time, which helps understand students' test-taking strategies.

The outcomes of this research show that for most of the subjects available in Digital Mocks Service, reliability was in line with values for reliability in large-scale assessments, that is, between upper 0.80s and lower 0.90s, and they were quite close to the reliability for the equivalent live paper-based assessments. The item level statistics indicated that items in the digital assessments worked, mostly, as expected. In particular, levels of difficulty of the items in the mock assessments ranged from difficult to easy in a similar way they did in the live assessments. Furthermore, the correlations between facilities in the mock and live assessments were generally high (between 0.85 and 0.97), indicating that the difficulty rank order did not change with the move to digital. Item discriminations indices were acceptable

for most items (higher than 0.25), although a bit lower in the mock than in the live assessments. However, as for the facilities, the correlations between discrimination indices in the mock and live assessments were fairly high (between 0.77 and 0.98).

Analysis of response times data showed that although there was some variation between the different subjects available in the Digital Mocks Service, there was no evidence of candidates running out of time: the average test duration ranged from 72% to 95% of the available total time. Furthermore, time spent in the items did not decline as the students progressed through the assessment. This implies that students were not rushing their responses when attempting the later items, which could have been an indication of speededness or students' disengagement.

The capture and use of process data has the potential to significantly improve the quality and reliability of digital assessments. Making sure we are aware of what data is available from the Digital Mocks Service and, whenever possible, exploit it, is key to evaluating the service. Findings from the evaluation can elicit recommendations, leading to improvements ahead of a future delivery of assessments, either mock assessments or digital high-stakes examinations.

References

Maddox, B. (2022). *The uses of process data in large-scale educational assessments*. OECD Education Working Paper No. 286. Paris: Organisation for Economic Co-operation and Development.