CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Evaluation of the Cambridge International Digital Mock Exams Service

Conference Abstract

Martina Kuvalja

AEA-Europe conference, 1–4 November 2023

Author contact details:

Martina Kuvalja
Assessment Research and Development,
Research Division
Shaftesbury Road
Cambridge
CB2 8EA
UK

martina.kuvalja@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:

Kuvalja, M. (2023, November 1–4). *Evaluation of the Cambridge International Digital Mock Exams Service*. [Paper presentation]. Annual conference of the Association for Educational Assessment – Europe (AEA-Europe), Malta. https://2023.aea-europe.net/

Abstract

Cambridge International are now offering a Digital Mocks Service for some international IGCSE and A-level exams in preparation for these live exams. The mock exams are taken on screen via a testing platform, and the marking is completed by Cambridge examiners. After the marking, teachers have access to various exam reports. After we launched our first live Digital Mocks Service across 6 countries in January 2023, we were interested to evaluate the service in order to input into the next series' development.
As part of the iterative design approach, we run a range of research activities throughout the product's lifecycle which includes the discovery stage, testing, build and delivery/launch of a product. After our first launch in January 2023, we were keen to evaluate the delivery of our first live Digital Mocks Service, and we used a particular methodological approach for this purpose. We collected (1) user data from the test platform and (2) user experience (UX) data, and we conducted (3) validity research.

(1) Test platform data

Test platform data is the data that is accessible from the test platform itself. For example, assessment (or Item Level data) and user log data are readily available for exporting and analysis. The purpose of collecting this data is to get validation evidence but also to understand test-takers' and examiners' activity and interactions with the on-screen assessment (user log data). User log data is also used to compare different designs (Dumais, Jeffries, Russell, Tang, & Teevan, 2014).

(2) User experience (UX) data

By UX data, we mean any data we collect about our customers' experience with the entire digital mock service; from the experience of booking a test through the booking application, to the test-taking experience and reporting of results. In this instance, we were interested in the experiences of the following groups of customers:

- the school staff in charge of booking and invigilating the test, and accessing (and interpreting) results;
- test-takers logging into the test platform and sit digital mock exams; and
- examiners who marked digital mock exams in the testing platform.

We propose a range of techniques to get this UX data; a log of issues/incidents raised by customers, observations (unmoderated remote usability testing sessions), post-delivery remote interviews (think-aloud), and surveys (online questionnaires and intercept surveys). We also suggest running some post-delivery targeted usability testing sessions with think-aloud to compare the presentation of test content on the screen (e.g., testing the accessibility of several design solutions, and comparing them).

(3) Validity research

In terms of establishing the validity of the uses of digital mock exams, and in addition to investigating ILD measures we used two additional approaches; a method used by Fishbein and colleagues (Fishbein, 2018; Fishbein et al., 2018) and expert judgment gathered from subject expert examiners (Shaw, Crisp & Hughes, 2020). Fishbein's approach aims to compare test features on paper and screen and test-takers' responses to questions on paper and screen, and expert judgment is used to establish if the constructs are interpreted and demonstrated as intended.

The findings that are produced as a result of the data collection and analysis can then be used to demonstrate the extent to which the service's purpose was met, and to inform the continuous development and improvement of the service based on this evidence. The data collection plan required a collaborative approach with different parts of the business involved in the Digital Mocks Service carrying out the proposed research activities as part of the iterative design process. This evaluation methodology aims to give the teams a structure for the evaluation so that the purpose of each data collection strand and the research questions are clear and linked to the objectives of the Digital Mocks Service.

References

Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding user behavior through log data and analysis. *Ways of Knowing in HCI*, 349-372.

Fishbein, B. (2018). Preserving 20 years of TIMSS trend measurements: Early stages in the transition to the eTIMSS assessment. Unpublished Doctoral Dissertation, Boston College.

Fishbein, B., Martin, M.O., Mullis, I.V.S. & Foy, P. (2018). The TIMSS 2019 Item Equivalence Study: examining mode effects for computer-based assessment and implications for measuring trends. *Large Scale Assessments in Education, 6*(11), 1-23.

Shaw, S., Crisp, V., & Hughes, S. (2020). A Framework for Describing Comparability between Alternative Assessments. *Research Matters, 29*, 17-22.