CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

Assessment mode effects and their relationship with item characteristics

Conference Abstract

Carmen Vidal Rodeiro
Carmen H. J. Lim

AEA-Europe conference, 5–8 November 2025

Author contact details:

Carmen Vidal Rodeiro
Research Division
Shaftesbury Road
Cambridge
CB2 8EA
UK

carmen.vidalrodeiro@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., & Lim, C.H.J. (2025, November 5–8). *Assessment mode effects and their relationship with item characteristics* [Paper presentation]. AEA-Europe 2025, The Hague. https://2025.aea-europe.net/

Abstract

Introduction

The digitalisation of high-stakes exams is becoming a reality. Globally-recognised school-leaving examinations, such as Cambridge International AS levels, will be available on-screen from June 2026, and awarding bodies in the United Kingdom are planning to introduce digital exams within the next few years. This transition is driven by the benefits on-screen assessments offer. For example, digitalisation has the potential to make exams more accessible and inclusive. Furthermore, it allows assessment to be more authentic, enabling, for instance, more effective testing of 21st-century skills such as programming.

Digital exams are likely to be gradually rolled out and be mostly delivered alongside paper versions. However, offering digital and paper versions simultaneously raises comparability concerns: will the same test items administered in different modes assess the same intended constructs and function similarly? It is, therefore, vital to ensure that students are not disadvantaged by the assessment mode and everyone benefits equally from technological innovation.

Cambridge University Press & Assessment (CUP&A) launched a Digital Mocks Service in 2023 to pave the way for the introduction of digital high-stakes examinations and to support schools and students during the transition by offering digital mocks. Our research took advantage of the data available in the Mocks Service to investigate the comparability between paper and on-screen assessments. In particular, the focus was on Differential Item Functioning (DIF) by assessment mode and whether certain types of items were more likely to drive mode effects.

Data

The digital mocks offered by CUP&A are based on exams delivered on paper in previous live sessions. In this research, we used item level data from the mock assessments and their paper versions (in subjects such as Computer Science, English Language and Global Perspectives), which were obtained from CUP&A data systems.

Methods

As a first step, Partial Credit Models, estimated using Winsteps, were employed to identify DIF across modes (i.e., mode effects).

The next step aimed to identify item characteristics that might drive DIF. To do that, we compared items in both the digital and paper versions to check for changes in presentation and classified them based on stimulus type, response format, response strategy, and tariff. This approach allowed us to systematically examine whether certain item types were more likely to exhibit mode effects (e.g., does a higher percentage of drag-and-drop items have DIF compared to other item types?). To avoid confirmation bias, we reviewed and classified all items (not just those showing differential performance) before the analyses.

Our classification framework was adapted from work by Crisp and Ireland (2022), with additions such as reading skills required when interacting with the texts (e.g., retrieve vs. reflect), which was considered by Harrison et al. (2023) when examining mode effects in PISA tests.

Findings

Among the 237 items investigated, we found that about 1 in 6 items showed DIF by assessment mode, with more items being harder on paper than on-screen. However, the type of subject (e.g., English Language, Computer Science) did not correlate with the number of items exhibiting DIF and items within each subject were not found to be consistently harder on-screen or on paper.

On item characteristics, how students interacted with text passages appeared to influence mode effects. Specifically, items requiring candidates to "access and retrieve" information from a text were about twice as likely to exhibit DIF compared to items involving other types of interactions. Additionally, items requiring students to provide their workings as part of their answers seemed to be harder on-screen than on paper. Other items that exhibited relatively higher levels of DIF included numeric or mathematical answers, required navigation to read the entire question, and had presentation differences between the digital and paper versions that could potentially impact reading.

However, it was challenging to determine whether the above characteristics were the sole reason driving the mode effects, as multiple factors could simultaneously affect item functioning. Furthermore, our research was limited by the availability of data within each item type and the comparability of the cohorts of students who took the digital and paper versions of the tests.

Conclusions

The findings suggest that, in the CUP&A qualifications included in this research, mode effects exist but are not extensive. More importantly, items within each subject were not found consistently harder on-screen or on paper, indicating that students are unlikely to be systematically disadvantaged by the assessment mode. The research also highlights several item characteristics that might drive DIF and should be investigated further.

References

Crisp, V. and Ireland, J. (2022). *A structure for analysing features of digital assessments that may affect the constructs assessed*. Cambridge University Press & Assessment.

Harrison, S., Kroehne, U., Goldhammer, F., Lüdtke, O. and Robitzsch, A. (2023). Comparing the score interpretation across modes in PISA: an investigation of how item facets affect difficulty. *Large-scale Assessments in Education*, 11(1): 8.