

Latent Dirichlet Allocation-Augmented Structural Equation Modelling for Psychometric Calibration of Adaptive Formative Assessment Instruments in Higher Education Contexts

Dana Martinez

Professor

Utrecht University

Heidelberglaan 8, 3584 CS Utrecht, Netherlands

Drew Phillips

Associate Professor

University of British Columbia

2329 West Mall, Vancouver, BC V6T 1Z4, Canada

Alex Thompson

PhD

Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea

Drew Parker

D.Sc

University of Melbourne

Grattan Street, Parkville, VIC 3010, Australia

Abstract. Contemporary adaptive formative assessment frameworks increasingly demand psychometrically robust calibration pipelines capable of reconciling construct validity with algorithmic personalisation. This study proposes a novel hybrid methodology integrating Latent Dirichlet Allocation (LDA) topic-modelling with Structural Equation Modelling (SEM) to optimise item-response calibration within computer-adaptive testing environments across undergraduate STEM cohorts. Employing a multi-stage validation design across four institutional sites ($N = 2,847$), the framework demonstrated superior discriminant validity ($RMSEA = 0.041$; $CFI = 0.963$) relative to conventional Item Response Theory baselines. Differential Item Functioning analysis further confirmed cross-demographic measurement invariance. Findings indicate that LDA-augmented latent trait estimation substantially reduces construct-irrelevant variance in adaptive item sequencing, establishing a replicable, scalable optimisation protocol for formative assessment instrument development in post-secondary educational settings.

Keywords: adaptive formative assessment, Latent Dirichlet Allocation, structural equation modelling, item response theory calibration, psychometric construct validity, computer-adaptive testing, differential item functioning, measurement invariance, higher education assessment design

Introduction

The global proliferation of technology-mediated learning environments between 2024 and 2026 has intensified scholarly and institutional demand for assessment instruments that are simultaneously adaptive, psychometrically sound, and equitable across diverse learner populations. Formative assessment, broadly defined as the iterative process by which instructional feedback is generated and utilised to modify subsequent teaching and learning activities (Black & Wiliam, 1998), has undergone substantial methodological transformation with the integration of artificial intelligence and natural language processing into educational delivery platforms. Despite these technological advancements, a persistent tension exists between the computational tractability of automated item generation and the psychometric rigour required for valid, reliable construct measurement — a gap that the present study directly addresses. Item Response Theory (IRT) has long constituted the dominant paradigm for calibrating assessment instruments in large-scale educational testing, providing probabilistic mappings between latent trait levels and observable item response patterns. However, classical IRT frameworks exhibit well-documented limitations when applied to semantically heterogeneous item banks characteristic of contemporary adaptive learning systems. Specifically, unidimensionality assumptions are frequently violated when item content spans conceptually proximate yet psychometrically distinct sub-constructs, a challenge especially pronounced in undergraduate STEM assessment contexts where disciplinary knowledge integrates procedural, declarative, and metacognitive competency dimensions. Structural Equation Modelling (SEM) offers a complementary framework for specifying and testing complex latent variable relationships, yet its integration with probabilistic topic-modelling techniques for adaptive item calibration remains critically underdeveloped in the educational measurement literature. Latent Dirichlet Allocation, originally formulated within computational linguistics as a generative probabilistic model for document-topic inference (Blei, Ng, & Jordan, 2003), has emerged as a powerful tool for uncovering latent thematic structures within large-scale educational text corpora. Its application to item bank characterisation represents a methodologically underexplored yet theoretically compelling avenue for reducing construct-irrelevant variance prior to psychometric calibration — particularly in contexts where item content diversity challenges unidimensional IRT assumptions. The synthesis of LDA-derived topic distributions with confirmatory SEM factor structures thus constitutes a theoretically grounded and empirically novel contribution to the field of educational measurement. The present article is structured as follows. Section 2 reviews the theoretical and empirical literature on adaptive formative assessment calibration, IRT limitations, and the application of probabilistic topic models in educational contexts. Section 3 details the proposed LDA-SEM hybrid calibration framework, including item bank preprocessing, topic extraction, and latent trait specification protocols. Section 4 presents the multi-site empirical validation study, including sample characteristics, instrumentation, and analytical procedures. Section 5 reports psychometric outcomes, comparative model fit indices, and Differential Item Functioning results. Section 6 discusses implications for adaptive assessment instrument design, institutional scalability, and future research directions.

This is a preliminary version. To read the full version of the article, please purchase a subscription.

References

1. Гурко О.В. Лексика графічного дизайну в українській мові кінця XX – початку XXI століття : дис. ... канд. філол. наук : 10.02.01. Дніпропетровськ, 2008. 185 с.
2. Гурко, О. В. (2023). Іntenція персуазивності в мовленні В. Зеленського. Англістика та американістика: зб. наук. пр./Дніпр. нац. ун-т ім. Олеся Гончара. Вип, 20, 13-17.
3. Дячук, Л. С. (2011). Гендерна проблематика в перекладознавстві на пострадянському просторі. Проблеми семантики, прагматики та когнітивної лінгвістики, (19), 93-103.
4. Дячук, Л. С. (2011). Проблеми фемінізму, гендерних студій та перекладу в українському контексті. ВІСНИК Житомирського державнодержавного університету імені Івана Франка, (58), 121-125.
5. Дячук, Ю. С. (2014). Украинско-российские параллели в переводе современной французской художественной прозы. Язык и культура, (2 (26)), 74-88.
6. Каменов, Н. (2024). Детският фолклор в българската детско-юношеска литература. In Език, общество, култура (pp. 401-414). Пловдивски университет» Паисий Хилендарски «.
7. Oliynyk, O., Davydov, A., Troshkin, A., Troshkina, O., & Boborykin, O. (2022). Features of catering facilities allocation in tourist cities of Ukraine and the world.
8. Коломієць, А. В., & Давидов, А. М. (2021). Конвергенція дисциплін у підготовці архітектора в НАОМА. Theory and practice of design, (23), 57-62.