Contribution ID: 233

Investigating the potential of large language models to streamline psychometric test development

Wednesday 23 July 2025 08:30 (15 minutes)

Oral presentation

Investigating the potential of large language models to streamline psychometric test development

Author

Meltem Ozcan, Hok Chio (Mark) Lai

Affiliation

University of Southern California

Abstract

Despite extensive research, accessible resources, sophisticated tools, and clear guidelines for the development and use of psychological scales, researchers often bypass critical steps in this process-such as measurement invariance (MI) testing-due to the complexity and time demands of these procedures. Questionable measurement practices, such as failing to test for MI, modifying scales without proper justification, or constructing scales without conducting necessary psychometric evaluations, are commonplace in research and compromise the validity of inferences (Flake, Pek, & Hehman, 2017; Maassen et al., 2023), highlighting the need to refine and streamline existing methods to increase adherence to best practices. Large language models (LLMs), with their high capacity for pattern recognition and human-like text generation capabilities, offer many new possibilities to address these challenges. While initial studies have primarily focused on using LLMs for item generation, their potential to streamline other aspects of test development (e.g., in identifying potentially biased items or by harnessing linguistic cues to supplement statistical evidence when data are limited) remains largely unexplored. In this talk, I discuss the challenges of conventional test development, review emerging applications of LLMs in psychometrics, and present findings from a systematic investigation of whether, when, how, and to what extent LLMs may be leveraged during the highly resource-intensive and iterative test development process. This work explores and highlights LLMs' potential to enhance and complement current practices to reduce researcher burden, improve the validity and fairness of psychometric measures, foster greater accessibility for researchers in fields with limited resources, and enable a more widespread adoption of rigorous methodological practices.

Keywords

scale development, large language models

Primary author: OZCAN, Meltem (USC)Co-author: LAI, Hok Chio (Mark) (USC)Session Classification: Session 2 : "Innovative methods in Measurement and Evaluation"

Track Classification: Measurement: Measurement