International Electronic Journal of Elementary Education
March 2025, Volume 17, Issue 2

# Editorial for the Special Issue on Large Scale Assessment: Challenges and Innovations

Ummugul Bezirhan, Murat Doğan Şahin

Received : 01 March 2025
Revised : 01 March 2025
Accepted : 19 March 2025
DOI : 10.26822/iejee.2025.381

[a] Ummugul Bezirhan, TIMSS & PIRLS International Study Center, Boston College, USA.
E-mail: bezirhan@bc.edu
ORCID: https://orcid.org/0000-0002-8771-4780

[b] Murat Doğan Şahin, Faculty of Education, Anadolu University, Eskişehir, Türkiye.
E-mail: muratdogansahin@gmail.com
ORCID: https://orcid.org/0000-0002-2174-8443

Copyright ©
www.iejee.com
ISSN: 1307-9298



## Abstract

This editorial introduces the IEJEE's Special Issue on Large Scale Assessment: Challenges and Innovations, highlighting emerging themes and methodological advancements in educational measurement. The selected studies focus on process data utilization to examine test-taker behavior, innovations in psychometric modeling for assessment, classification, and the influence of social-emotional learning on academic achievement. This editorial discusses the contributions of the included studies, their implications for future research, and the evolving role of AI, machine learning, and digital assessment technologies in shaping the future of large-scale assessments.

## Introduction

Educational testing is transforming dramatically as digital platforms, data analytics, and machine learning reshape assessment practices. These advancements provide deeper insights into test-taker behavior, enhance psychometric modeling techniques, and expand our understanding of the cognitive and non-cognitive factors influencing student achievement. As educational systems worldwide embrace digital testing and artificial intelligence (AI)-driven methodologies, researchers must navigate both the opportunities and challenges presented by these innovations.

One of the most profound shifts in large-scale assessment research is the increasing reliance on process data to capture real-time student interactions during testing. Process data allows researchers to analyze patterns of engagement, test-taking strategies, and response modifications, providing a richer picture of student performance (e.g., Bezirhan, 2021; Goldhammer et al., 2014; Ulitzsch et al., 2020; Wise 2017), insights that were largely inaccessible in paper-based assessment environments (Kane & Mislevy, 2017). Additionally, well established methodologies such as latent class analysis (LCA) and latent profile analysis (LPA), offer powerful tools for identifying unobserved subgroups of test-takers, allowing researchers to refine student classifications (Williams & Kibowski, 2016). LCA has been explored as a data-driven alternative for setting proficiency classifications in assessments (e.g., Templin & Jiao, 2012; Binici & Cuhadar, 2022). Similarly, recent applications of LPA in educational measurement have been particularly effective in analyzing test-taking engagement (e.g., Anghel et al., 2025) and variations in student problem solving strategies (e.g, Teig, 2024) across different populations utilizing process data.

Beyond the technical innovations, the role of non-cognitive factors in student achievement has garnered attention as well. Traditional assessments have long focused on cognitive abilities and content knowledge, but emerging research highlights the importance of social-emotional learning (SEL), test-taking motivation, and engagement as key predictors of achievement (OECD, 2021). These factors not only shape student performance but also raise important considerations for fairness and equity in assessment design. This holistic approach acknowledges that academic achievement is shaped by complex interactions between content knowledge, test taking strategies, and non-cognitive factors such as perseverance, self-regulation, and social awareness (Farrington et al., 2012).

This special issue brings together a collection of studies balancing established methodologies with emerging advancements to address challenges faced in large-scale assessments. The included articles explore the intersection of process data, psychometric innovation, and non-cognitive influences on learning outcomes. By addressing both theoretical and practical implications, this issue offers fresh perspectives on how assessment research can evolve to meet the demands of contemporary education.

**Overview of the Special Issue**

As large-scale assessments continue to evolve, researchers explore novel approaches to address fundamental challenges in educational measurement. The studies featured in this special issue contribute to this growing body of research by examining innovations in process data, psychometric modelling, and non-cognitive measurement. The research presented here spans a diverse range of assessment contexts, including international assessments such as the Programme for International Student Assessment (PISA) and the Progress in International Reading Literacy Study (PIRLS), national assessments like the National Assessment of Educational Progress (NAEP), and state-level assessments such as the Iowa Assessments and the Cognitive Abilities Test (CogAT). While these studies employ distinct analytical frameworks, they collectively enhance our understanding of how large-scale assessments can be designed, analyzed, and interpreted to better support diverse student populations.

A key area of innovation in this issue is process data and student behavior analysis. The study by Ogut et al. (in this issue) examines extended time (ET) accommodations in the NAEP Grade 8 Mathematics assessment, utilizing a machine learning model (XGBoost) to identify students who may benefit from additional time. Their findings indicate that while a majority of students granted ET do not fully utilize it, nearly a quarter of students without accommodations remain actively engaged when their time expires. This study highlights the potential for predictive models to guide more equitable ET allocation policies, confirming that students with actual needs receive appropriate support. Ni et al. (in this issue) investigate response change behaviors in NAEP constructed response items, developing a novel framework that integrates automated scoring with dimensional response analysis. Their study finds that students who make substantive changes to their responses, particularly those involving conceptual modifications, are more likely to improve their scores. This research underscores the value of process data in understanding student engagement and response strategies, paving the way for more adaptive scoring and feedback mechanisms in digital assessments. Kara (in this issue) explores test-taking disengagement in PISA 2022, using LPA to classify students based on response time, number of actions, and self-reported effort. The findings reveal that disengagement is associated with lower test performance and that process data-based measures such as response time and number of actions are more reliable indicators of engagement than self-reported effort. Gender disparities in disengagement further highlight the need for targeted interventions to improve test-taking motivation across diverse student populations.

Beyond test-taking behavior, two studies focus on improving measurement methodology. Yin et al. (in this issue) introduce an LCA-based approach to setting cut scores for context scales addressing challenges posed by skewed response distributions. By applying their method to PIRLS 2021 data, they demonstrate its potential to enhance the interpretability of context scales and provide a more statistically robust alternative to conventional judgment-based cut-score definitions. Demirkaya et al. (in this issue) examine latent profiles of mathematical skills by comparing student classifications derived from achievement and ability assessments using widely administered state assessments in the United States. Their study reveals substantial differences in the profiles emerging from these two classification approaches, highlighting the importance of using multiple measures to identify students with distinct instructional needs. These findings have direct implications for gifted education, as they suggest that relying on a single measure may overlook students who demonstrate strong cognitive potential despite lower achievement scores.

The final study in this special issue, by Altiner Sert and Arikan (in this issue) explores the relationship between social-emotional learning (SEL) and mathematics achievement, using data from the OECD's 2019 Survey on Social and Emotional Skills. Their findings suggest that emotional regulation and open-mindedness positively predict math performance, while high social engagement is negatively associated with achievement. Notably, SEL skills have a stronger predictive impact on students from lower socioeconomic backgrounds, reinforcing the importance of SEL programs in mitigating educational inequities.

**Concluding Thoughts**

The studies featured in this special issue demonstrate the evolving landscape of large-scale assessments, driven by advancements in data science, psychometric techniques, and a deeper understanding of student behavior. The findings highlight the increasing role of process data in improving assessment validity and fairness, the need for refined measurement models that accommodate diverse student populations, and the growing recognition of non-cognitive factors in shaping academic performance.

Future research should continue to explore AI-driven models for personalizing test accommodations, enhancing test development process and refine process data methodologies to improve engagement detection and response behavior analysis, and further examine the role of non-cognitive skills in educational assessments. As educational systems continue to embrace computer-based assessment practices and AI-driven methodologies, the intersection of assessment technology, psychometrics, and behavioral insights will remain a critical area of research. This special issue aims to inspire further innovation and interdisciplinary collaboration, ultimately contributing to more equitable and insightful large-scale assessments.

**References**

Anghel, E., Wry, E., & von Davier, M. (2025). Affective, behavioral, and cognitive engagement in ePIRLS: a latent profile analysis. *Educational technology research and development*, 1-27.

Bezirhan, U., von Davier, M., & Grabovsky, I. (2021). Modeling item revisit behavior: The hierarchical speed–accuracy–revisits model. *Educational and Psychological Measurement, 81*(2), 363-387.

Binici, S., & Cuhadar, I. (2022). Validating performance standards via latent class analysis. *Journal of Educational Measurement, 59*(4), 502-516.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: the role of noncognitive factors in shaping school performance--a critical literature review*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*(3), 608.

Kane, M., & Mislevy, R. (2017). Validating score interpretations based on response processes. In *Validation of score meaning for the next generation of assessments* (pp. 11-24). Routledge.

Organisation for Economic Co-operation and Development. (2021). Beyond academic learning: First results from the survey of social and emotional skills. *OECD Publishing.*

Teig, N. (2024). Uncovering student strategies for solving scientific inquiry tasks: Insights from student process data in PISA. *Research in Science Education, 54*(2), 205-224.

Templin, J., & Jiao, H. (2012). Applying model-based approaches to identify performance categories. In *Setting performance standards* (pp. 379-397). Routledge.

Ulitzsch, E., von Davier, M., & Pohl, S. (2020). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non response. *British Journal of Mathematical and Statistical Psychology, 73*, 83-112.

Williams, G. A., & Kibowski, F. (2016). Latent Class Analysis and Latent Profile Analysis. In *Handbook of methodological approaches to community-based research: qualitative, quantitative, and mixed methods* (pp. 143-151). Oxford University Press.

Wise, S. L. (2017). Rapid-guessing behaviour: Its identification, interpretation, and implications. *Educational Measurement: Issues and Practice, 36*, 52–61. https://doi.org/10.1111/emip.12165

.