ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Editorial; 2024; 15(Special issue); 260-262

EDITORIAL

Opportunities and Challenges of AI in Educational Assessment

Alper SAHIN* **Nathan THOMPSON**** **Kadriye ERCIKAN*****

Abstract

In the past few years, as artificial intelligence (AI) and large language models (LLM) have rapidly entered our lives, we have witnessed groundbreaking innovations across numerous fields. The rapid pace of these changes has been met with excitement by some and apprehension by others. However, we all agree that they have made tremendous contributions so far and their contributions in the future will reshape our existence. The field of educational assessment is no exception. With this in mind, we issued a call for a special issue themed "Opportunities and Challenges of AI in Educational Assessment." which finally included seven distinguished articles on subthemes of fair and responsible use of AI in educational assessment, learning analytics, automated scoring, and real-life examples of AI and LLM.

Keywords: AI in Educational assessment, Large language models, fair and responsible use of AI and LLM, learning analytics, automated scoring, real-life examples of AI and LLM

For this special issue, we received 23 high-quality article proposals from seven different countries: Canada (1), Germany (1), Türkiye (1), South Africa (4), South Korea (1), the United Kingdom (1) and the United States (12). Of these proposals, we invited 12 authors to submit their articles for the special issue. Ultimately, we received 14 article submissions. After a rigorous blind review procedure and revisions, we are pleased to present seven meticulously selected articles in our special issue.

In line with our theme, our special issue begins with a systematic review, by Sato et al. (2024), of literature related to the fair and responsible use of artificial intelligence in educational assessment. In their article titled "Putting AI in Fair: A Framework for Equity in AI-driven Learner Models for Accessible and Inclusive Assessment", Sato et al. (2024) present us with an extensive literature review covering theoretical, empirical, ethical, and policy documents addressing the role of learner models in K-12 assessment. The authors sought answers to 5 important research questions regarding whether and how these models were used to promote accessibility and inclusivity, and they propose a framework that aspires to influence the equitable and valid of assessment of all students.

Following the opening article in the introduction section, the second section of our special issue includes two valuable articles by Guo et al. (2024) and Cavus & Kuzilek (2024) on *Learning Analytics*, which has the potential to directly impact the quality of education. In their very interesting and useful study titled "*Human-Centered AI for Discovering Student Engagement Profiles on Large-Scale Educational Assessments*", Guo et al. (2024) propose an artificial intelligence-supported model that combines response and process data to better reflect students' knowledge and test-taking processes using multi-source data to reveal their engagement profiles. We strongly recommend that you read this article where they described and tested this model, which is a first in the field that will allow educators to access deeper and more useful information about their students' test performances and knowledge levels. In their article titled "*An Effect Analysis of the Balancing Techniques on the Counterfactual Explanations*

* Assist. Prof. Dr., Department of Basic English, Atilim University, Ankara, Türkiye, alpersahin2@yahoo.com, ORCID ID: 0000-0001-7750-4408

** CEO, Assessment Systems Corporation, Minneapolis, MN, US, nthompson@assess.com, ORCID ID: 0000-0002-3981-7881

*** SVP of Global Research, ETS, Princeton, NJ, US, kercikan@ets.org, ORCID ID: 0000-0001-8056-9165

To cite this article:
Sahin, A., Thompson, N., & Ercikan, K. (2024). Opportunities and challenges of AI in educational assessment. *Journal of Measurement and Evaluation in Education and Psychology, 15,*(Special issue), 260-262. https://doi.org/10.21031/epod.1607441

of Student Success Prediction Models", Cavus & Kuzilek (2024) investigated the effectiveness and feasibility of using various counterfactual explanations to predict students' success to be better understood by students and parents and to increase their trust to these predictions. We believe you will enjoy this article.

The third section of our special issue includes studies by Chan et al. (2024) and Mo Zhang et al. (2024), who have undertaken two important studies on *Automated Scoring*, where AI and LLM have been widely used for a long time. Chan et al. (2024), in their study titled "*Integrating Metadiscourse Analysis with Transformer-Based Models for Enhancing Construct Representation and Discourse Competence Assessment in L2 Writing: A Systemic Multidisciplinary Approach*", address the important but somewhat neglected topic of discourse competence in Automated Essay Scoring. While doing this, the authors, who use Metadiscourse markers (MDM), test what can be done to expand the ability of automated scoring models to identify and classify MDM with 2000 texts at different CEFR levels and share their findings with us, providing a foundation for future research to expand the construct of L2 automated scoring models. In their article titled "*Investigating Sampling Impacts on an LLM-Based AI Scoring Approach: Prediction Accuracy and Fairness*", Mo Zhang et al. (2024) investigated the effects of different sampling methods on the ability of AI to predict the scores given by human raters, and of the stratified sampling method on the fairness of AI's prediction ability, together with other methods. We strongly recommend that you read this article, which is one of the novel studies conducted with newly developed language models and yielded interesting findings.

Finally, in the fourth and final section of our special issue, there are two valuable studies by Bolender et al. (2024) and Ting Zhang et al. (2024), which evaluated the performance of *real-life examples of AI and LLM* that reached the end user. In their article titled "*Generative AI in K12: Analytics From Early Adoption*", Bolender et al. conducted three comprehensive case studies that included real-life use of Finetune's Finetune Generate, developed for item development using natural language models, and Finetune Catalog, developed to tag and align educational content to various standards and frameworks. Ting Zhang et al. (2024) contributed to our special issue with a multi-disciplinary study titled "*Ask NAEP: A Generative AI Assistant for Querying Assessment Information*", which, as the name suggests, includes an evaluation of the performance of the Ask NAEP chatbot, which was developed to provide accurate and comprehensive answers using publicly available National Assessment of Educational Progress (NAEP) data. You will not regret reading it.

All in all, we hope that you will read the articles in this special issue with pleasure and that this special issue will contribute significantly to the field of education assessment. On behalf of all researchers who conduct studies on educational assessment, we would like to thank the authors who have shown interest in our special issue, supported us with their article proposals and articles, and contributed to the publication of this magnificent special issue.

We would also like to express our gratitude to the journal editors and to our expert reviewers who have supported us so that the blind review process of this special issue ran smoothly and without any problems, to the layout editors, and to the journal's editorial team who have worked hard to prepare the articles in the special issue for publication.

References

Bolender, B., Vispoel, S., Converse, G., Koprowicz, N., et al. (2024). Generative AI in K12: Analytics From Early Adoption. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 361-377. https://doi.org/10.21031/epod.1539710

Cavus, M., & Kuzilek, J. (2024). An Effect Analysis of the Balancing Techniques on the Counterfactual Explanations of Student Success Prediction Models. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 302-317. https://doi.org/10.21031/epod.1526704

Chan, S., Sathyamurthy, M., Inoue, C., Bax M., et al. (2024). Integrating Metadiscourse Analysis with Transformer-Based Models for Enhancing Construct Representation and Discourse Competence Assessment in L2 Writing: A Systemic Multidisciplinary Approach. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 318-347. https://doi.org/10.21031/epod.1531269

Guo, H., Johnson, M., Saldivia, L., Worthington, M., et al. (2024). Human-Centered AI for Discovering Student Engagement Profiles on Large-Scale Educational Assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 282-301. https://doi.org/10.21031/epod.1532846

Sato, E., Shyyan, V., Chauhan, S., Christensen, L. (2024). Putting AI in Fair: A Framework for Equity in AI-driven Learner Models and Inclusive Assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 263-381. https://doi.org/10.21031/epod.1526527

Zhang, M., Johnson, M., & Ruan, C. (2024). Investigating Sampling Impacts on an LLM-Based AI Scoring Approach: Prediction Accuracy and Fairness. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 348-360. https://doi.org/10.21031/epod.1561580

Zhang, T., Patterson, L., Beiting-Parrish, M., Webb, B., et al. (2024). Ask NAEP: A Generative AI Assistant for Querying Assessment Information. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special issue), 378-394. https://doi.org/10.21031/epod.1548128