Journal of Applied Testing Technology, Vol 23(Special Issue 1), 1-3, 2022

Editorial
Special Issue on Artificial Intelligence, Machine Learning, and Natural Language Processing Applications in Education

Okan Bulut

Centre for Research in Applied Measurement and Evaluation, University of Alberta, Canada;
bulut@ualberta.ca

1. Introduction

Advancements in Artificial Intelligence (AI) continue to impact the future of almost every sector and industry, and education is no exception. AI technologies have profoundly changed education in many ways, increasing opportunities for personalized learning and enabling new ways for individuals to learn and work together. Researchers and practitioners in educational measurement and psychometrics are also interested in investigating how AI can be used as a supplementary tool (e.g., for scoring essays and constructed-response items automatically). Rather than replacing human intelligence, AI technologies are typically used to perform complex tasks (e.g., building personalized learning paths for thousands of learners) or streamline laborious tasks (e.g., scoring essays and generating test items) in an efficient and cost-effective manner.

Researchers can now conduct innovative research using non-traditional data sources such as process data (Jiao *et al*., 2021; Lindner & Greiff, 2021) and textual data (Flor & Hao, 2021), thanks to emerging AI technologies. Similarly, advances in technology and learning science have prompted researchers to adapt or redefine key concepts in educational measurement and psychometrics (e.g., validity, reliability, and fairness) in the context of digital learning and assessment systems. A recent special issue of the Journal of Educational Measurement, for example, focused on how modern validity arguments can be applied or adapted to AI-enabled innovative forms of assessment, from conception to reporting (Dorsey & Michaels, 2022).

This special issue aims to demonstrate a range of innovative applications and theoretical frameworks in educational measurement and psychometrics using different branches of AI (e.g., machine learning, deep learning, and natural language processing). The topics of interest for this special issue included machine learning applications in educational measurement and psychometrics, innovative assessment formats (e.g., intelligent tutoring systems), psychometric applications involving process data, natural language processing applications (e.g., automated scoring, automatic item and content generation, and automatic speech recognition), and ethics of AI.

In the original call for papers, we asked all interested authors to submit a Letter of Intent (LOI) providing a summary of their proposed contribution. Both empirical and theoretical submissions were considered at this stage. After reviewing the submitted LOIs, we identified a set of submissions aligned with the goals of this special issue and then invited their authors to submit a full manuscript to be considered for inclusion in the special issue. All articles submitted for publication went through a rigorous peer review based on the review standards established by the Journal of Applied Testing Technology.

2. Overview of the Special Issue

This special issue consists of seven articles highlighting a wide range of topics, such as automated essay scoring,

natural language processing for identifying enemy items, pupillometry and eye-tracking data for measuring cognitive effort, and AI for generating feedback. A summary of each article is provided below.

The first two articles in this special issue focus on Automated Essay Scoring (AES) systems based on deep learning algorithms. Shin and Gierl's article (in this issue) compares the performance of traditional and modern AES systems in capturing the coherence-related attributes of writing (e.g., organization, word choice, and narrativity). The authors discuss how modern AES systems utilizing deep learning algorithms can overcome the limitations of traditional AES systems designed to score the overall writing quality. Then, using the Automated Student Assessment Prize (ASAP) dataset, they demonstrate that deep-neural AES systems can predict score attributes related to coherence more accurately than traditional, feature-based AES systems.

Firoozi, Bulut, Demmans Epp, Naeimabadi, and Barbosa (in this issue) examine the effect of fined-tuned word embedding techniques on the accuracy of deep learning-based AES systems. The authors argue that fine-tuning word embedding techniques such as GloVe and Word2Vec can improve the accuracy of the AES system using deep learning structures, such as the Long-Short Term Memory (LSTM) model. For their empirical analysis, Firoozi and colleagues also used the ASAP dataset and found that when word embeddings were fine-tuned, not only did they improve the accuracy of the AES systems, but they also helped increase the computational efficacy of the models in terms of training time.

The next two articles in this special issue present frameworks for identifying enemy item pairs using natural language processing techniques in large item banks. Micir, Swygert, and D'Angelo use a large item bank of 4130 items from a large-scale healthcare specialist certification exam to find enemy item pairs. The authors compute TF-IDF weighted cosine similarity among the items and then build a machine-learning model to predict a final similarity indication based on the similarity indices and item content coding. The results show that their proposed method can detect the enemy item pairs with a very high level of accuracy. Becker and Kao (in this issue) also focus on the use of natural language processing for the identification of the enemy and duplicate items in large item banks. The authors provide a summary of methods for detecting the enemy and duplicate items and then use several item banks to demonstrate the application of cosine similarity to enemy item identification.

Mead and Zhou (in this issue) investigate how accurately the wording of test items can predict Bloom's taxonomy level of the items. The authors apply natural language processing techniques to process a sample of multiple-choice items extracted from online practice exams measuring information technology-related topics. Then, they build a Naive Bayes classifier to predict Bloom's taxonomy level for the items. Their findings show that the accuracy of their automated classification approach is comparable with the reliability of Bloom's classification performed by subject matter experts, although the performance of their approach seems to vary across the cognitive levels.

Thomas (in this issue) investigates whether pupillometry data can be used to quantify cognitive effort and then validate a construct map for graphic literacy based on cognitive effort. Thomas utilizes the eye tracking methodology to collect gaze data from participants who worked through items on an assessment measuring workplace graphic literacy skills. Despite using a small sample of participants, Thomas presents convincing evidence that total cognitive effort from the initial step of viewing the items to the final step of providing an answer is associated with cognitive difficulty. The author also demonstrates how this information can be used for better understanding cognitive performance.

In the final article, Wongvorachan, Lai, Bulut, Tsai, and Chen (in this issue) provide a comprehensive overview of the present and future of AI applications in educational feedback practices. The authors unveil the role of the three branches of AI (i.e., natural language processing, educational data mining, and learning analytics) in feedback applications in education. As they explain the involvement of AI in educational feedback, they also discuss future research directions focused on the integration of AI technologies into educational feedback practices.

3. Concluding Thoughts

In summary, the goal of this special issue of the Journal of Applied Testing Technology is to share innovative ideas and applications for the use of AI in educational measurement and psychometrics. We believe that the articles accepted for publication in this issue are thought-

provoking, provide valuable insights, and can form the basis for further discussion about the role and use of AI for enhancing educational assessments. It should be noted that this special issue does not cover the entire spectrum of topics represented in the original call for papers. Some of these topics included consideration of ethical and privacy concerns related to AI, AI-based decision-making processes, and regulatory frameworks on the use of AI in education. We strongly encourage future research to address the ethical challenges raised by implementing AI and guide researchers and practitioners about how ethical AI can be achieved in educational measurement and psychometrics.

4. Acknowledgment

I would like to thank Dr. Dianne Henderson (ACT) and Dr. Holly Garner (Kaplan) for inviting me as a guest editor to run this special issue in the Journal of Applied Testing Technology. This special issue could not have been created without their help and guidance throughout the design, peer review, and publication processes. Also, I would like to thank anonymous reviewers for their support and insightful comments on the papers submitted to this special issue.

5. References

Dorsey, D. W., & Michaels, H. R. (2022). Validity arguments meet artificial intelligence in innovative educational assessment. *Journal of Educational Measurement, 59*, 267-271. https://doi.org/10.1111/jedm.12331

Flor, M., & Hao, J. (2021). Text mining and automated scoring. In A. A. von Davier, R. J. Mislevy, & J. Hao, (Eds.) Computational psychometrics: New methodologies for a new generation of digital learning and assessment. *Springer*, pp. 245-262. https://doi.org/10.1007/978-3-030-74394-9_14

Jiao, H., He, Q., & Veldkamp, B. P. (2021). Process data in educational and psychological measurement. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.793399 PMid:34925194 PMCid:PMC8677654

Lindner, M. A., & Greiff, S. (2021). Process data in computer-based assessment: Opening the black box. *European Journal of Psychological Assessment, 37*(2), 161-162. https://doi.org/10.1027/1015-5759/a000658