



## Special issue: e-testing from artificial intelligence approach

Masaki Uto<sup>1</sup> · Kazuo Shigemasu<sup>2</sup>

Published online: 31 July 2021  
© The Behaviormetric Society 2021

Artificial intelligence has recently achieved tremendous success in various domains, including data science, computer vision, robotics, and natural language processing. Furthermore, artificial intelligence approaches are already having a big impact on modern educational and psychological measurement technologies including e-testing.

E-testing, a technology that uses ICT to manage and deliver tests for measuring examinee ability, has recently been applied in various settings including entrance and qualification exams and tests in online learning environments. A characteristic feature of e-testing is the automatic assembly of uniform test forms, for which each form has equivalent measurement accuracy for examinee ability but with a different set of test items. The latest e-testing technologies using artificial intelligence approaches have succeeded in assembling a vast number of uniform test forms while controlling how many times each item is used.

Artificial intelligence approaches have also been utilized as a key component of various modern testing technologies. For example, recent advanced probabilistic models have succeeded in improving the reliability of performance assessments, in which human raters assess examinee performance in a practical task, by mitigating the effects of various bias factors in the assessment process. Another representative application of artificial intelligence in the testing domain is automated essay scoring (AES) and automated short answer scoring utilizing natural language processing and machine learning techniques to automatically grade essays and constructed responses.

This special issue of *Behaviormetrika* consists of two invited papers and one review paper on advanced uses of artificial intelligence in modern testing technologies.

The first invited paper, titled “e-testing from artificial intelligence approach” Ueno et al. (2021), begins by giving an overview of state-of-the-art uniform test assembly methods that substantially increase the number of assembled tests while controlling

---

✉ Masaki Uto  
uto@ai.lab.uec.ac.jp

<sup>1</sup> The University of Electro-Communications, Tokyo, Japan

<sup>2</sup> Keio University, Tokyo, Japan

the item exposure frequency. Then, to relax strong assumptions in item response theory (IRT) models, which are probabilistic models that have been widely used to estimate examinee ability in e-testing, the paper introduces an extension of IRT using a deep learning approach called Deep-IRT.

The second invited paper, titled “A multidimensional generalized many-facet Rasch model for rubric-based performance assessment” Uto (2021a), proposes a new IRT model for improving the reliability of performance assessment that uses a scoring rubric consisting of multiple evaluation items. In such assessments, measurement accuracy for examinee ability is known to depend on the characteristics of the rubric’s evaluation items and of the raters. To resolve this problem, the paper proposes a new multidimensional IRT model that can estimate examinee ability while considering the effects of these characteristics.

The review paper, titled “A review of deep-neural automated essay scoring models” Uto (2021b), presents a detailed survey of deep neural network (DNN)-based AES models that have been proposed in the past few years. The paper classifies the AES task into four types and introduces various DNN-based AES models from the literature according to this classification. This review provides a good source of knowledge for both researchers and practitioners in e-testing, suggesting future directions for the next stage of AES.

The editors hope that this special issue will contribute to the further extension and advanced application of artificial intelligence in e-testing.

## References

- Ueno M, Fuchimoto K, Tsutsumi E (2021) E-testing from artificial intelligence approach. *Behaviormetrika*. <https://doi.org/10.1007/s41237-021-00143-x>
- Uto M (2021a) A multidimensional generalized many-facet rasch model for rubric-based performance assessment. *Behaviormetrika*. <https://doi.org/10.1007/s41237-021-00144-w>
- Uto M (2021b) A review of deep-neural automated essay scoring models. *Behaviormetrika*. <https://doi.org/10.1007/s41237-021-00142-y>

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.