

**Editorial: Advancements in Technology-Based Assessment: Emerging Item Formats,  
Test Designs, and Data Sources**

Frank Goldhammer<sup>1,2</sup>, Ronny Scherer<sup>3</sup>, Samuel Greiff<sup>4</sup>

<sup>1</sup>DIPF | Leibniz Institute for Research and Information in Education

<sup>2</sup>Centre for International Student Assessment (ZIB)

<sup>3</sup>University of Oslo, Centre for Educational Measurement (CEMO)

<sup>4</sup>University of Luxembourg

Technology has become an indispensable tool for educational and psychological assessment in today's world. Individual researchers and large-scale assessment programs alike are increasingly using digital technology (e.g., laptops, tablets, and smartphones) to collect behavioral data beyond the mere correctness of item responses. Along these lines, technology innovates and enhances assessments in terms of item and test design, methods of test delivery, data collection and analysis, and the reporting of test results.

The aim of this research topic is to present recent developments in technology-based assessment and in the advancements of knowledge associated with it. Our focus is on cognitive assessments, including the measurement of abilities, competences, knowledge, and skills, but also includes non-cognitive aspects of assessment (Rausch, Kögler, & Seifried, 2019; Simmering, Ou, & Bolsinova, 2019). In the area of (cognitive) assessments, the innovations driven by technology are manifold, and the topics covered in this collection are, accordingly, wide and comprehensive: Digital assessments facilitate the creation of new types of stimuli and response formats that were out of reach for assessments using paper; for

instance, interactive simulations may include multimedia elements, as well as virtual or augmented realities (Cipresso et al., 2019; De-Juan-Ripoli et al., 2019). These types of assessments also allow for the widening of the construct coverage in an assessment; for instance, through stimulating and making visible certain problem-solving strategies that represent new forms of problem solving (Han, He, & von Davier, 2019; Kroeze, van den Berg, Lazonder, Veldkamp, & de Jong, 2019). Moreover, technology allows for the automated generation of items based on specific item models (Shin, Guo, & Gierl, 2019). Such items can be assembled into tests in a more flexible way than what is possible in paper-and-pencil tests and can even be created on the fly; for instance, tailoring item difficulty to individual ability (adaptive testing) while assuring that multiple content constraints are met (Born, Fink, Spoden, & Frey, 2019; Zhang, Wang, Gao, Cai, & Tu, 2019). As a requirement for adaptive testing, or to lower the burden of raters who code item responses manually, computers enable the automatic scoring of constructed responses; for instance, text responses can be coded automatically by using natural language processing and text mining (He, Veldkamp, Glas, & van den Berg, 2019; Horbach & Zesch, 2019).

Technology-based assessments provide not only *response data* (e.g., correct vs. incorrect responses) but also *process data* (e.g., frequencies and sequences of test-taking strategies, including navigation behavior) that reflect the course of solving a test item and gives information on the path towards the solution (Han et al., 2019). Process data, among others, have been used successfully to evaluate and explain data quality (Lindner, Lüdtke, & Nagy, 2019), to define process-oriented latent variables (De Boeck & Scalise, 2019), to improve measurement precision, and to address substantial research questions (Naumann, 2019). Large-scale result and process data also call for data-driven computational approaches in addition to traditional psychometrics and new concepts for storing and managing data (Von Davier, Wong, Polyak, & Yudelso, 2019).

The contributions of this research topic address how technology can further improve and enhance educational and psychological assessment from various perspectives. Regarding educational testing, not only is research presented on the assessment *of* learning, that is, the summative assessment of learning outcomes (Molnár & Csapó, 2019), but a number of studies on this topic also focus conceptually and empirically on the assessment *for* learning, that is, the formative assessment providing feedback to support the learning process (Arieli-Attali, Ward, Thomas, Deonovic, & von Davier, 2019; Blaauw, van der Gaag, Snell, Emerencia, Kunnen, & Jonge, 2019; Csapó & Molnár, 2019; Den Ouden, Keuning, & Eggen, 2019; Kroeze et al., 2019).

Table 1 gives an overview of all the papers included in this research topic and summarizes them with respect to their key features. Reflecting the scope of the research topic, we used four major categories to classify the papers: (1) papers focusing on the use of new data types and sources, (2) innovative item designs, (3) innovative test designs, and (4) statistical approaches. We refrained from multiple category assignments of papers, which was possible, and focused on their core contribution. The papers' key findings and advancements impressively represent the current state-of-the-art in the field of technology-based assessment in (standardized) educational testing, and, as topic editors, we were happy to receive such a great collection of papers with various foci.

Regarding the future of technology-based assessment, we assume that inferences about the individual's or learner's knowledge, skills, or other attributes will increasingly be based on empirical (multimodal) data from less- or non-standardized testing situations. Typical examples are stealth assessments in digital games (Shute, 2015; Shute & Ventura, 2013), digital learning environments (Nguyen, Huptych, & Rienties, 2018), or online activities (Kosinski, Stillwell, & Graepel, 2013). Such new kinds of unobtrusive, continuous assessments will further extend the traditional assessment paradigm and enhance our

understanding of what an item, a test, and the empirical evidence for inferring attributes can be (Mislevy, 2019). Major challenges lie in the identification and synthesis of evidence from the situations the individual encounters in these non-standardized settings, as well as in validating the interpretation of derived measures. This research topic provides much input for these questions. We hope that you will enjoy reading the contributions as much as we did.

## References

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802. doi:10.1073/pnas.1218772110
- Mislevy, R. (2019). On integrating psychometrics and learning analytics in complex assessments. In H. Jiao, R. W. Lissitz, & A. van Wie (Eds.), *Data analytics and psychometrics* (pp. 1-52). Charlotte, NC, USA Information Age Publishing.
- Nguyen, Q., Hupych, M., & Rienties, B. (2018). *Linking students' timing of engagement to learning design and academic performance*. Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge, Sydney, New South Wales, Australia.
- Shute, V. (2015). Stealth assessment. In J. Spector (Ed.), *The SAGE Encyclopedia of Educational Technology* (pp. 675-676). Thousand Oaks, California: SAGE Publications, Inc.
- Shute, V., & Ventura, M. (2013). *Stealth Assessment: Measuring and Supporting Learning in Video Games*: The MIT Press.

Table 1

*Overview of the papers.*

Article	Area(s) of advancement	Data types	Statistical approach	Assessment purpose (of/for learning)	Assessment domains	Key finding and advancement
Focus on new data types and sources						
Blaauw et al. (2019)	Computerized assessment of learning with multiple informants	Survey responses, platform user data	Descriptive approach	FOR	Vocational education	Multi-informant time-series data can inform the success of educational interventions to support students at risk.
De Boeck & Scalise (2019)	Log-file and performance data to assess CoIPS	Actions, response times, correctness of item responses	Confirmatory factor analysis	OF	Collaborative problem solving (PISA 2015)	Dependencies among action, time-on task, and performance indicators do not only exist at the construct but also the item (residual) level.
Lindner et al. (2019)	Time-on task to identify rapid guessing	Correctness of item responses, response times	Latent class analysis	OF	Science achievement	Response times can provide information about rapid-guessing behavior and its relations to cognitive resources and test-taking effort.
Naumann (2019)	Time-on task data of reading	Correctness of item responses, response	Linear mixed modeling	OF	Reading literacy (PISA 2009)	Response times can help identify relations between item difficulties, strategic knowledge, skills, and motivation to ultimately craft a validity

		times				argument.
Simmering et al. (2019)	Assessment of non-cognitive skills	Continuous process data (e.g., behavioral, physiological)	-	-	Non-cognitive skills	Challenges and limitations in using technology-enhanced assessments require consideration.
Von Davier et al. (2019)	Data paradigms for educational learning and assessment systems	Response behavior, test content, instructional content	e.g., machine learning	OF/FOR	Divers	The concept of the “data cube” can be used to label, collect and store data.
Focus on innovative item designs						
Arieli-Attali et al. (2019)	Learning design	Learners’ responses and use of learning support	e.g., hidden Markov modelling	FOR	Divers	The traditional evidence centered design models can be expanded to assess learning.
Cipresso et al. (2019)	Assessment of unilateral spatial neglect	Correctness of item responses	-	-	Unilateral spatial neglect	Complex 3D environments on mobile devices are promising for the ecological assessment of unilateral spatial neglect.
De-Juan-Ripoli et al. (2019)	Assessment of risk taking	Behavioral and physiological responses	-	-	Risk taking	Virtual realities (VR) can be employed to simulate hazardous situations realistically.
Den Ouden et al. (2019)	Computerized dynamic assessment of text	Correctness of item responses	Linear modeling and	FOR	Text comprehension	Computer-based dynamic assessments bear the potential to support students in acquiring

	comprehension skills		MTMM			reading skills.
Horbach & Zesch (2019)	Automated content scoring	Written text	Machine learning	OF	Diverse	Automated content scoring approaches can take into account the variance in learner answers.
Kroeze et al. (2019)	Automated feedback generation	Written text, actions, correctness of item responses	Descriptive approach, linear model	OF/FOR	Scientific inquiry in economics and physics	Automated feedback on scientific hypotheses can agree with human ratings to a great extent, and students who receive it are likely to develop better hypotheses than those who don't.
Focus on innovative test designs						
Born et al. (2019)	Computerized adaptive testing and test equating	Correctness of item responses	Item response theory	OF	-	Equating designs and CAT can be combined through a continuous calibration strategy.
Csapó & Molnár (2019)	Assessment for teaching and learning	Correctness of item responses	Item response theory	FOR	Mathematics, science, and reading	Teaching and learning can be supported on a large scale by online assessment solutions (authoring, assembly, scoring, delivery, feedback).
Molnár & Csapó (2019)	Computerized assessment of cognitive development	Correctness of item responses	Confirmatory factor analysis and structural equation models	OF/FOR	Mathematics, science, and reading competences	Computerized assessments can capture differences in the academic performance on tests in mathematics, science, and reading across grade levels and make visible the psychological dimension of learning.

Rausch et al. (2019)	Embedded experience sampling for assessing non-cognitive skills	Survey responses, correctness of item responses	MTMM, item response theory	OF	Non-cognitive facets of problem solving	Embedded experience sampling provides an approach to assess non-cognitive facets of competences through multiple self-reports.
Zhang et al. (2019)	Computerized adaptive testing of Internet addiction	Survey responses	Item response theory	OF	Internet addiction	A computerized adaptive test of Internet addiction assessed the construct accurately and efficiently, and provided evidence for both the reliability and validity of the resultant test scores.
Focus on statistical approaches						
Han et al. (2019)	Data mining using random forests to predict item performance	Actions, response times, correctness of item responses	Tree-based model	OF	Problem solving (PISA 2012)	A random forest algorithm can generate and select features from the process data that predict students' item responses.
He et al. (2019)	Text mining and item response data to identify PTSD	Written text, survey responses	Item response theory and text classification	-	Posttraumatic stress disorder	Combining text classification and item response theory models provides an efficient approach to estimating the latent trait.
Shin et al. (2019)	Topic modeling for item distractor generation	Written text	Machine learning	OF	Knowledge and skills in biology	Latent topic modeling supports the identification of students' misconceptions in biology and aids the development of distractors.

