Editorial to the Special Issue: Current Innovations in Computer-Based Assessments

Ronny Scherer

University of Oslo, Norway

Samuel Greiff

University of Luxembourg, Luxembourg

Paul A. Kirschner

The Open University of the Netherlands, The Netherlands;

University of Oulu, Finland

Author Note

Ronny Scherer, University of Oslo, Faculty of Educational Sciences, Centre for Educational Measurement at the University of Oslo (CEMO), Postbox 1161 Blindern, 0318 Oslo, Norway; Email: ronny.scherer@cemo.uio.no; Samuel Greiff, Institute of Cognitive Science and Assessment (COSA), University of Luxembourg, 11, Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg, Email: samuel.greiff@uni.lu; Paul A. Kirschner, The Open University of the Netherlands, Heerlen, The Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands; University of Oulu, Learning & Educational Technology Research Unit, Oulu, Finland; Email: paul.kirschner@ou.nl.

Editorial to the Special Issue: Current Innovations in Computer-Based Assessments

Introduction

Assessment matters. Indeed, scientific progress largely depends on the extent to which assessments can provide reliable and valid measures of variables – be it well-defined and observable variables in the natural sciences or complex and unobservable variables in the social sciences (Duckworth & Yeager, 2015). With the rapid development of information and communication technologies, new potentials also arise for assessing complex psychological skills and human behavior (Mayrath, Clarke-Midura, & Robinson, 2012; Shute & Rahimi, 2017). Computer-based assessments (CBAs), for example, now allow researchers to capture complex constructs such as collaborative problem-solving and computational thinking skills that have recently gained importance across domains and contexts (Greiff, Holt, & Funke, 2013; Grover & Pea, 2013; Scherer, 2015), and assess constructs that have been considered essential skills for decades with more innovative and perhaps more authentic item formats (e.g., mathematical, reading, and scientific literacy; OECD, 2016). Besides the core testing purposes of distinguishing between students of different knowledge, skills, and performance levels, CBAs can also be used to assess student learning – without any high-stakes consequences based on a single, final score. In this sense, CBAs are powerful tools for both assessment of learning (i.e., summative) and assessment for learning (i.e., formative assessment; Shute & Rahimi, 2017).

The potential of CBA is widely recognized, especially in the areas of educational and psychological testing (Drasgow, 2016). Even further, international large-scale assessments in education, such as the Programme for International Student Assessment (PISA), the Programme for the International Assessment of Adult Competencies (PIAAC), the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the International Computer and Information Literacy Study

(ICILS), have shifted from paper-and-pencil towards CBA approaches of educationally relevant constructs. These constructs comprise not only "traditional" skills (e.g., mathematical, reading, scientific literacy) but also "new" skills that have become relevant for students in the 21st century (e.g., complex and collaborative problem solving, ICT literacy, computational thinking). The core potential of CBAs lies in the provision of novel, interactive tasks (OECD, 2013), and the possibility to obtain information on test-taking behavior (Goldhammer, Martens, Christoph, & Lüdtke, 2016; Greiff, Wüstenberg, & Avvisati, 2015). Taking an educational measurement perspective, Zenisky and Luecht (2016) summarize the core innovations of computer-based assessment and highlight the assessment and psychometric modeling of complex constructs, the automated scoring and test assembly (Gierl, Latifi, Lai, Boulais, & De Champlain, 2014; Veldkamp, 2015), and the availability of process data to describe not only performance (for example, by the correctness of item responses) but also strategic behavior, sequences, and patterns of actions (Greiff, Niepel, Scherer, & Martin, 2016). It is the designated aim of this special issue to present both the core innovations of CBAs in various domains and contexts and the challenges associated with them.

Objectives, Scope, and Content

Considering the opportunities that come with the technological advancements in assessments, this special issue presents empirical research on current innovations in CBAs of existing and new constructs across various sectors, along with state-of-the-art applications focusing on the use of the resulting data to describe human behavior that go beyond traditional assessment approaches. We expect this special issue to impact future research and practice of using modern technologies as assessment tools, to generate publicity of CBAs, and to create an awareness of their potential in various contexts and disciplines.

In fact, with the increasing technical developments in the world of information and communication technologies is the implementation of computers for assessment inevitable, possibly to the point where CBAs take over assessment in general. Researchers and designers should be prepared to deal with this and can provide guidelines instead of only reacting to students being confronted with sub-optimal design in and use of CBAs. If this is not done, the possibilities of CBAs can be negated and/or even bring with them serious adverse effects. For instance, when introducing CBAs, it is tempting to simply put paper-and-pencil assessments on a computer. This change of medium without adaptation to that medium can cause disadvantages, for example to the processing of information because paper pages that can easily be turned often cannot be revisited on computer pages leading to distinct differences between the two versions (Kirschner, Park, Malone, & Jarodzka, 2016). On the other hand, CBAs can allow for assessment to be more adaptive to the learner. With a large enough database of well-designed items accompanied by a (smart) adaptivity algorithm, it becomes possible to provide different versions of an assessment to different groups of learners in different situations. Hence, the assessment can be adapted to each student's knowledge level and thus, not only be conducted more quickly (by avoiding too difficult and too simple questions), but also be more accurate by carving out the abilities of a student in detail. Hence, proper study – and the bundling of such studies in a special issue – is of utmost importance.

The papers in this special issue present several innovations associated with CBAs. These innovations range from generic assessment approaches to the measurement of highly complex skills (e.g., collaborative problem-solving skills) by innovative task designs all the way through to ways of capturing the complexity of computer-generated data (e.g., log-file and multi-channel data) to describe the processes that underlie these complex skills. Table 1 provides an overview of the topics covered by the nine papers in this special issue along with their core innovations, which fall into four main categories: (1) Assessment of new constructs

or widening the assessment of existing constructs; (2) Use of log-file and multi-channel data; (3) Psychometric models and experiments that inform the measurement of complex skills and task construction; (4) Integration of assessment and learning. Each paper showcases how the potential of CBAs can be exploited to address substantively relevant issues, primarily in educational and psychological measurement. The papers take different perspectives on the assessment of constructs, ranging from task design to data mining strategies and psychometric models (see Table 1).

Concluding Remarks

The papers presented in this special issue feature core innovations in the field of CBAs that not only tap design issues or the development of interactive and perhaps more authentic tasks in comparison to existing assessments, but also present ways of making sense of the resultant data that are stored in log files. These ways comprise both the extraction of relevant information from log-file data (e.g., sequential actions, response times) and the psychometric modeling techniques. In doing so, the contributions that make up this special issue highlight at least two perspectives on CBAs: (1) the need for crafting a validity argument of indicators derived from CBAs; (2) the need for the simultaneous development of assessments and psychometric models that combine performance and behavioral data meaningfully. The nine papers presented in the special issue exemplify the integration of these perspectives. At the same time, these papers point to possible future directions of research and development of CBAs. These directions include the application of CBAs across psychological sub-disciplines, their theoretical and empirical foundation and validation, the integration of assessment and learning, modern test designs, and, perhaps most importantly, the development of a CBA-specific, cognitive theory.

Tables

Table 1

Overview of studies describing innovations in computer-based assessments

Paper	*Topic*	*Core innovation(s)*
Assessment of new constructs or widening the assessment of existing constructs		
Graesser, Cai, Morgan, and Wang (2017)	Assessment with computer agents that engage in conversational dialogues and trialogues with learners	Assessment of collaborative skills and processes with log-file data based on human-agent interaction, including cognitive and non-cognitive aspects
Rowe et al. (2017)	Assessing implicit science learning in digital games	Assessment of implicit learning (i.e., unarticulated knowledge development) based on log-file data and data-mining strategies
von Davier, Hao, Liu, and Kyllonen (2017)	Interdisciplinary research agenda in support of assessment of collaborative problem solving: Lessons learned from developing a collaborative science assessment prototype	Assessment of collaborative problem-solving skills and processes with log-file data based on a statistical definition of collaboration and data-mining strategies
Use of log-file and multi-channel data		
Taub et al. (2017)	Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with "Crystal Island"	Assessment of cognitive and meta-cognitive processes of self-regulation combining eye-movement and log-file data
Vista, Care, and Awwal (2017)	Visualizing and examining sequential actions as behavioral paths that can be interpreted as markers of complex behaviors	Assessment of collaborative problem-solving skills and processes with log-file data based on data-mining strategies (i.e., visualization of behavioral sequences)
Zechner, Yoon, Bhat, and Leong (2017)	Comparative evaluation of automated scoring of syntactic competence of non-native speakers	Assessment of language skills applying automated scoring techniques to spoken item responses

Psychometric models and experiments that inform the measurement of complex skills and task construction		
Engelhardt, Goldhammer, Naumann, and Frey (2017)	Experimental validation strategies for heterogeneous computer-based assessment items	Assessment of ICT skills and information on the effects of task characteristics on construct validity
Veldkamp, Avetisyan, Weissman, and Fox (2017)	Stochastic programming for individualized test assembly with mixture response time models	Potential of response times (obtained from log-file data) and response-time models to inform test development
Integration of assessment and learning		
Nguyen, Rienties, Toetenel, Ferguson, and Whitelock (2017)	Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates	Potential of log-file data to inform test development; Examination of the stability and changes of performance over time (i.e., generation of time-intense log-file data)

References

Drasgow, F. (2016). *Technology and Testing: Improving Educational and Psychological Measurement*. New York, NY: Routledge.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters. *Educational Researcher, 44*(4), 237-251. doi:10.3102/0013189X15584327

Engelhardt, L., Goldhammer, F., Naumann, J., & Frey, A. (2017). Experimental validation strategies for heterogeneous computer-based assessment items. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.02.020

Gierl, M. J., Latifi, S., Lai, H., Boulais, A.-P., & De Champlain, A. (2014). Automated essay scoring and the future of educational assessment in medical education. *Medical Education, 48*(10), 950-962. doi:10.1111/medu.12517

Goldhammer, F., Martens, T., Christoph, G., & Lüdtke, O. (2016). Test-taking engagement in PIAAC. *OECD Education Working Papers*. doi:10.1787/19939019

Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.03.041

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving, 5*(2), 71-91. doi:10.7771/1932-6246.1153

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61*, 36-46. doi:10.1016/j.chb.2016.02.095

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment

of problem solving. *Computers & Education, 91*, 92-105. doi:10.1016/j.compedu.2015.10.018

Grover, S., & Pea, R. (2013). Computational Thinking in K–12. *Educational Researcher, 42*(1), 38-43. doi:10.3102/0013189X12463051

Kirschner, P. A., Park, B., Malone, S., & Jarodzka, H. (2016). Toward a cognitive theory of multimedia assessment (CTMMA). In J. M. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 1-23). Cham: Springer International Publishing AG. doi:10.1007/978-3-319-17727-4_53-1

Mayrath, M. C., Clarke-Midura, J., & Robinson, D. H. (2012). Introduction to Technology-Based Assessments for 21st Century Skills. In M. C. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.), *Technology-Based Assessments for 21st Century Skills* (pp. 1-12). Charlotte, NC: Information Age Publishing.

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.03.028

OECD. (2013). *PISA 2012 Assessment and Analytical Framework - Mathematics, Reading, Science, Problem Solving and Financial Literacy*. Paris: OECD Publishing.

OECD. (2016). *PISA 2015 Assessment and Analytical Framework*. Paris: OECD Publishing.

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., . . . Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.03.043

Scherer, R. (2015). Is it time for a new measurement approach? A closer look at the assessment of cognitive adaptability in complex problem solving. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01664

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer-Assisted Learning, 33*, 1-19. doi:10.1111/jcal.12172

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.01.038

Veldkamp, B. P. (2015). Computerized Test Construction. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 510-514). Oxford: Elsevier.

Veldkamp, B. P., Avetisyan, M., Weissman, A., & Fox, J.-P. (2017). Stochastic programming for individualized test assembly with mixture response time models. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.04.060

Vista, A., Care, E., & Awwal, N. (2017). Visualising and examining sequential actions as behavioural paths that can be interpreted as markers of complex behaviours. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.01.027

von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: lessons learned from developing a Collaborative Science Assessment Prototype. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.04.059

Zechner, K., Yoon, S.-Y., Bhat, S., & Leong, C. W. (2017). Comparative evaluation of automated scoring of syntactic competence of non-native speakers. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.01.060

Zenisky, A. L., & Luecht, R. M. (2016). The future of computer-based testing. In C. Wells & M. Faulkner-Bond (Eds.), *Educational Measurement: From Foundations to Future* (pp. 221-238). New York, NY: Guilford Press.

Acknowledgements

The guest editors would like to thank the journal managers and the journal staff of *Computers in Human Behavior* for their support in all practicalities. The guest editors would also like to thank the authors and reviewers without whom this special issue would have never come to pass.

This research was supported by the FINNUT Young Research Talent Project Grant (NFR-254744 “ADAPT21”) awarded to Ronny Scherer by The Research Council of Norway, and by the Fonds National de la Recherche Luxembourg (ATTRACT “ASKI21”) awarded to Samuel Greiff. All statements expressed in this article are those of the authors and do not reflect the official opinions or policies of the authors’ host affiliations or any of the supporting institutions.