Special Session—Challenges and Opportunities for Assessment in XR

Cindy Ziker
Ziker Research
A Division of Ziker Enterprises, LLC.
San Jose, California, US
cindyziker@yahoo.com

Evelien Ydo
Educational Science and Technology
University of Twente
Enschede, The Netherlands
evelienjy@gmail.com

Diego Zapata-Rivera
Cognitive and Technology Sciences Center
Educational Testing Service
Princeton, New Jersey, US
dzapata@ets.org

Mathew Hillier
Office of the Pro-Vice Chancellor Learning and Teaching
Macquarie University
Sydney, Australia
m.hillier@uq.edu.au

Michael Casale
STRIVR
Los Angelas, California, US
casale@strivr.com

***Abstract*—This Special Session will examine the opportunities and challenges associated with developing and implementing authentic assessments in XR to inform teachers, faculty, researchers, and industry trainers interested in advancements in this field. Methods for conducting real-time assessments in XR environments will be demonstrated and include audience participation. A brief introduction regarding the state of assessment in XR will be presented by moderator Dr. Cindy Ziker, followed by four presenters and an interactive discussion on assessment in XR that will include topics posed by the audience. Evelien Ydo will present on performance-based assessment (PBA) and simulation-based assessment (SBA) using virtual reality. Dr. Diego Zapata-Rivera's presentation will focus on conducting adaptive, embedded assessments that leverage the virtual presence of experts, teachers, and observers, and interactions with virtual agents and virtual objects. Dr. Mathew Hillier will present on the "Transforming Assessment Project", that includes the use of unique software tools and augmented reality. Dr. Michael Casale will present on innovations in work-based assessment and industry training.**

Index Terms*—authentic-assessment, evidence-centered design, XR, virtual reality, augmented reality, simulation-based assessments, performance-based assessment, stealth assessment, work-based assessment

I. Introduction

XR environments have the potential to provide authentic assessment experiences that allow learners to demonstrate their knowledge and skills in innovative and meaningful ways. However, typically assessment of knowledge gained during XR activities involves the use of external instruments that are not embedded in the virtual experiences of the learner (e.g., multiple choice tests administered after the XR experience). Implementing assessments in XR that have the capacity to elicit evidence of what learners know and can do with precision requires close alignment between instructional content, the assessment tasks, and the expected outcomes. This is especially true of work-based assessments that are used to determine readiness for performance in job-related tasks. Challenges to this include construct irrelevant variance, a lack of familiarity with navigating the virtual environment, and weaknesses in scoring methods. One established method of assessment design and development that can alleviate some of these issues is the use of evidence-centered design (ECD), which can support the validity of the assessment argument during the design stage [1]. Here we address the opportunities for novel assessment strategies provided by the affordances of XR, in order to inform practitioners, researchers, and industry trainers interested in advancements in this field. This effort examines authentic, performance-based, simulation-based, embedded, dialogue-based, and work-based assessment methods that use a variety of cutting-edge virtual tools and platforms. The future of assessment in XR is also addressed.

II. Challenges and Opportunities for Simulation-Based Assessment of Technical Procedures in VR

Currently, performance-based assessments (PBA) are heavily relied on by vocational education [2]. The biggest disadvantage of PBAs is that they are prone to measurement errors which cause lower validity and reliability (e.g. [3]). Assessing performance through the use of simulation-based assessment (SBA) in VR could offer great opportunities for vocational education. Adding SBA to the current assessment program of vocational education could help lower the measurement errors, uncover particular knowledge, skills or abilities and assess students in uncommon or dangerous situations as well as motivate students more, reduce test anxiety by stealth assessment and create the possibility to capture process and product data (e.g. [2]–[7]). The challenges of SBAs involve finding the relevant constructs and attributes of an assessment and measuring them correctly with a coherent and complete psychometric model. Processing, analyzing, and giving meaning to large amounts of data is a big challenge as well. In this session, I will discuss my current research project concerning creating a simulation-based assessment in VR of a technical procedural training for vocational education. We will dive into the data opportunities of SBAs in VR, and how to tackle challenges that

978-1-7348995-0-4/20/$31.00 ©2020 Immersive Learning Research Network

6th International Conference of the Immersive Learning Research Network (iLRN 2020)
Online, June 21-25, 2020.

arise when developing them. Furthermore, I will elaborate on how the principles of evidence-centered design [8] can be used to enhance the development of a SBA.

III. Opportunities and Challenges for Embedded assessments in XR

XR environments offer many opportunities and challenges for assessment. Opportunities for assessment include the collection a large variety of process and response data, supporting the virtual presence of experts, teachers, or observers who could "drop-in on" a virtual scene to observe or interact with students, supporting the interaction with virtual objects (e.g., virtual lab instruments and materials), and allowing for opportunities to implement adaptive, embedded assessment through the interaction with virtual agents and virtual objects. Challenges involve processing and analyzing large amounts of data, making sense of these data, implementing adaptive features that support assessment and learning, and addressing accessibility concerns. In this session I will elaborate on our experiences applying Evidence-Centered Design [8] to the development of adaptive learning and assessment systems for English language learning [9], immersive dialogue-based assessments for science inquiry skills [10], English language skills [11], mathematics and vocabulary [12], and collaboration and communication skills [13]. Work at ETS on hybrid assessment systems [14], ECD for learning [15], embedded assessments [16], game-based assessments [17], validity of technology-based assessments [18], accessibility for intelligent tutoring systems [19], and tools for analyzing process data and log files [20] provide insights on how to achieve the potential of assessment in XR environments and successfully deal with its challenges.

IV. Transforming Assessment

The use of XR technologies has shown promise in providing innovative student assessment [21]–[23], yet face a number of hurdles before widespread adoption can be realized. The growing digital divide due to resource constraints, costs, expertise and available connectivity in regional and developing areas is a significant issue [24]. In many developed regions of the world internet connectivity continues to be an issue. A case in point is Australia where a fiber to the home national broadband network started construction to only be trashed by a change of government leaving a developed nation with what could only be described as "fraud-band". This has left many areas including those in the nation's largest cities unable to engage bandwidth intensive applications. This has come to a head with students facing obstacles undertaking live, remotely invigilated online exams given that the upload requirement is above that provided by the base plans offered by the national broadband network wholesaler. In terms of ease of deployment for assessment purposes, most XR technologies require a high level of technical expertise and time to be able to effectively develop suitable assessments within XR environments. This means that the development of assessment tasks is costly and inaccessible when compared to existing assessment approaches such as an online quiz. The COVID-19 pandemic has seen a flurry of activity with assessment platform vendors going into overdrive promoting their wares as a panacea that can address the closure of on-campus learning. Yet many offer relatively mundane capabilities that lack the means for students to demonstrate authentic, twenty first century capabilities. There appears to be tension between the development of assessment platforms that enable the design of authentic assessment tasks and the tendency for vendors of assessment systems to focus on scalability and security. However, there are a small number of projects looking to resolve this tension [25]. One such example is Australian Government funded project "Transforming Exams" [26]. The project sought to deliver a scalable, secure, reliable, cost-effective assessment platform that would enable a pathway from paper equivalent assessments to post-paper, authentic task design. It leverages the digital workflow efficiencies of an LMS, is robust to network outages and allows students to use a range of sophisticated 'software tools of the trade' in constructing complex responses. A demonstration of using augmented reality within the secured platform was also shown [27]. The session will outline an Australian perspective on the obstacles and progress made in the use of XR for student assessment.

V. The Use of Immersive Technologies to Enhance Workplace-Based Assessments

Workplace-based assessments (WBAs) are meant to provide direct evidence of proficiencies of interest in professional training and assessment settings (i.e., what learners will ultimately do in professional practice). While many WBAs have demonstrated that they can be used to predict real world performance (e.g., [28]), the effectiveness of current WBAs techniques have been challenged [29], [30], particularly in the area of leadership assessment. A recently developed and promising research area for tracking workplace behavioral performance - "stealth assessment" - may alleviate some of the challenges of current WBAs. The core idea is that stealth assessments are embedded into virtual learning environments that allow evaluators to track students' knowledge, skills and attributes in an ecologically valid way [31]. Moreover, stealth assessment methods allow evaluators to assess various behaviors related to particular skills and attributes providing indirect real-time evaluations [8]. To make valid inferences from the assessments, the stealth method is supported by evidence-centered design (ECD), a conceptual reference framework that can be used to develop assessment models [5]. In this session I will cover how companies are incorporating immersive technologies like virtual reality (VR) in order for more possibly valid insights that can then be used in conjunction with training protocols. In particular, I will discuss how data generated from VR can be used to provide novel and more valid representations of workplace behaviors, including a use case from a large retailer that employed VR for workplace behaviors. The results from this initial research effort demonstrated that VR is able to provide a highly valid representation of critical workplace behaviors.

References

[1] R. G. Almond, L. S. Steinberg, and R. J. Mislevy, (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment,* I(5).

[2] S. Klerk,(2015). Multimedia-Based Performance Assessment in Dutch Vocational Education. PhD thesis University of Twente, Enschede, the Netherlands. doi: 10.3990/1.9789036539975

[3] R. J. Shavelson, G. P. Baxter, and X. Gao, (1993). Sampling variability of performance assessments. Journal of Educational Measurement, 30(3), 215-232.

[4] S. De Klerk, P. Van Dijk, and L. Van den Berg,(2015). Voordelen en uitdagingen voor toetsing in computersimulaties [Advantages and challenges of assessment in computer simulations]. Examens, 12(1), 11-17.

[5] V. J. Shute, (2011). Stealth assessment in computer-based games to support learning. In S. Tobias and J.D. Fletcher (Eds.), Computer Games and Instruction (pp. 503-524).Charlotte, NC: Information Age Publishing.

[6] V. J. Shute, M.Ventura, M. I. Bauer, and D. Zapata-Rivera, (2009). In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), The Social Science of Serious Games: Theories and Applications.

[7] J. Hedberg, and S. Alexander, (1994). Virtual reality in education: defining researchable issues. Educational Media International, 31, 4, 214–220. DOI: 10.1080/0952398940310402.

[8] R. J. Mislevy, L. S. Steinberg, and R. G. Almond, (2003) On the structure of educational assessments. Measurement: Interdisciplinary Research and Perspectives. 1, 3–62.

[9] D. Zapata-Rivera, T. Jackson, L. Liu, M. Bertling, M. Vezzu, and I. R. Katz, (2014) Science Inquiry Skills using Trialogues. 12th International conference on Intelligence Tutoring Systems. 625-626.

[10] D. Zapata-Rivera, W. Vanwinkle, V. Shute, J. Underwood, and M. Bauer, (2007). English ABLE. In Artificial Intelligence in Education - Building Technology Rich Learning Contexts That Work. Luckin, R., Koedinger, K., & Greer, J. (Eds.) Vol. 158. 323 – 330.

[11] Y. So, D. Zapata-Rivera, Y. Cho, C. Luce, and L. Battistini, (2015). Using trialogues to measure English language skills. In J. García Laborda, D. G Sampson, R. K. Hambleton and E. Guzman (Eds.) Journal of Educational Technology and Society. Special Issue: Technology Supported Assessment in Formal and Informal Learning. 21-32.

[12] A. A. Lopez, C. Luce, D. Zapata-Rivera, and C. Forsyth, (2017). Using formative conversation- based assessments to support students' English language development. IEEE Technical Committee on Learning Technology Bulletin, 19(1), 6-9.

[13] J. Hao, D. Zapata-Rivera, A. C. Graesser, Z. Cai, X. Hu, and B. Goldberg, (2018) Towards an Intelligent Tutor for Teamwork: Responding to Human Sentiments. In R. Sottilare, A. Graesser, X. Hu, and A. M. Sinatra (Eds.). Design Recommendations for Intelligent Tutoring Systems: Volume 6 - Team Tutoring. Orlando, FL: U.S. Army Research Laboratory. ISBN 978-0-9977257-4-2. 151–160.

[14] D. Zapata-Rivera, L. Liu, L. Chen, J. Hao, and A. von Davier, (2016) Assessing Science Inquiry Skills in Immersive, Conversation-based Systems. In B. K. Daniel (ed.), Big Data and Learning Analytics in Higher Education, Springer International Publishing, 237-252.

[15] E. Hansen, and D. Zapata-Rivera, (2010) Evidence Centered Design (ECD) for the Design of Learning-and-Assessment Environments for Diverse Learners. Paper presented at the annual meeting of the American Educational Research Association (AERA).

[16] D. Zapata Rivera, (2012) Embedded Assessment of Informal and Afterschool Science Learning. Summit on Assessment of Informal and After-School Science Learning Retrieved from https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072564.pdf

[17] A. Oranje, B. Mislevy, M. I. Bauer, and G. T. Jackson, (2019). Summative game-based assessment. In D. Ifenthaler & Y. J. Kim (Eds.), Game-Based Assessment Revisited (pp. 37–65). Springer International Publishing. https://doi.org/10.1007/978-3-030-15569-8_3.

[18] I. R. Katz, M. M. LaMar, R. Spain, D. Zapata-Rivera, J. Baird, and S. Greiff, (2017). Validity Issues and Concerns for Technology-based Performance Assessments. In R. Sottilare, A. Graesser, X. Hu, and G. Goodwin (Eds.). Design Recommendations for Intelligent Tutoring Systems: Volume 5 - Assessment Methods. Orlando, FL: U.S. Army Research Laboratory. ISBN 978-0-9893923-9-6. 209–224.

[19] H. G. Hansen, D. Zapata-Rivera, and J. White, (2018). Framework for the design of accessible intelligent tutoring systems. In S. D. Craig (Ed.). Tutoring and Intelligent Tutoring Systems (pp. 69-101). New York, NY: Nova Science Publishers.

[20] J. Hao, and R. J. Mislevy, (2018). The evidence trace file: a data structure for virtual performance assessments informed by data analytics and evidence-centered design. (Research Report No. RR - 18 - 28). Princeton, NJ: Educational Testing Service.

[21] G. Crisp, M. Hillier, and S. Joarder, (2010a). Assessing students in Second Life – some options. Australasian Society for Computers in Learning in Tertiary Education conference 5-8 December, Sydney pp. 256–261.

[22] G. Crisp, M. Hillier, and S. Joarder, (2010b). Assessing students in Second Life with scripted chatbots. Australian Technology Network Assessment Conference, Assessment Sustainability Diversity and Innovation, 18-19 November, Sydney, pp. 1–3.

[23] J. C. Munoz-Carpio, M. A. Cowling, and J. R Birt, (2019). Enhancing Learning of System Modelling through 360 Virtual Reality Video and 3D Printing. Australasian Society for Computers in Learning in Tertiary Education conference, 2-5 Dec, Singapore, pp. 378–382.

[24] M. Hillier, (2018) "Bridging the digital divide with off-line e-learning", Distance Education, 39(1), pp. 110-121. DOI: 10.1080/01587919.2017.1418627

[25] A. E. Fluck, (2019). An international review of eExam technologies and impact. Computers & Education, 132, 1–15. DOI: 10.1016/j.compedu.2018.12.008

[26] M. Hillier, S. Grant, and M. Coleman, (2018) Towards authentic e-Exams at scale: robust networked Moodle, Australasian Society for Computers in Learning in Tertiary Education conference, 25-28 Nov, Geelong, Australia, pp. 133-143

[27] M. Cowling, M. Hillier, and J. Birt, (2018) Integrating mixed reality spatial learning analytics into secure electronic exams, Australasian Society for Computers in Learning in Tertiary Education conference, 25-28 Nov, Geelong, Australia, pp. 332-336

[28] J. Norcini, & V. Burch, (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. Medical teacher, 29(9-10), 855-871.

[29] S. Haines, (2009). Bankrupt leadership development? Training, 46:64.

[30] M. Beer, M. Finnström, and D. Schrader, (2016). Why Leadership Training Fails and What to Do About It. Harvard Business Review, (October). Available at: https://hbr.org/2016/10/why-leadership-training-fails-and-whatto-do-about-it

[31] V. J. Shute, (2009). Simply assessment. International Journal of Learning and Media, 1(2), 1-11.