Assessment in Immersive Virtual Environments: Cases for Learning, of Learning, and as Learning

Jillianne Code¹, Nick Zap²

¹University of British Columbia // ²University of Victoria //

jillianne.code@ubc.ca // nickzap@uvic.ca

ABSTRACT

The key to education reform lies in exploring alternative forms of assessment. Alternative performance assessments provide a more valid measure than multiple-choice tests of students’ conceptual understanding and higher-level skills such as problem solving and inquiry. Advances in game-based and virtual environment technologies are creating new possibilities for learning and assessment. One such possibility is through the use of immersive 3D technologies that aim to situate students in an environment that promotes problem solving and inquiry and sets the context for assessment. This paper describes three on-going research projects that are using immersive 3D virtual environments as a platform in which to enable the summative, self-, and formative assessments of student learning.

Keywords

Virtual environments; formative assessment; immersive learning; learning analytics
Introduction

The key to education reform lies in exploring alternative forms of assessment. The roles that assessments play in educational reform often coincide with the introduction of new philosophies of learning, driving major themes in policy change (Linn, 2000). Regardless of philosophical paradigms, advances in technology create new possibilities for learning and assessment (Behrens, 2009; Pellegrino, Chudowski, & Glaser, 2001). In September 2009, the US National Research Council held a workshop on games and simulations. The white papers from this research conference urge further research to determine the full potential of collaborative, immersive simulations to support assessment (Quellmalz & Pellegrino, 2009), as well as virtual worlds that integrate assessment while engaging learning (Clark, Nelson, Sengupta, & D’Angelo, 2009). Coinciding with the rapid development of technology capable of enabling collaboration and assessment in learning technologies, a quick search of the PsycINFO database reveals that research in 3D virtual environments over the past 10 years has grown at a rate of over 450%. Current research aims to explore this potential through the use of immersive 3D technologies that aim to situate students in an environment that promotes inquiry and sets the context for assessment (i.e. Clarke-Midura, Code, Zap, & Dede, 2012). Immersive environments can potentially support student experimentation and scientific reasoning in a virtual context by allowing students the ability to walk around an environment, giving students the opportunity to take on the identity of an avatar, a virtual personae, that can explore and interact within the 3D environment by making observations, gathering data, and solving a scientific problem in context. In addition, current research is actively exploring whether immersive virtual environments coupled with learning analytics can enable the automated, invisible, and non-intrusive collection of students’ actions and behaviours during the act of learning so that the capture and assessment
of learning is in situ (Clarke-Midura, Code, & Dede, 2011; Clarke-Midura & Dede, 2010; Gobert, Sao Pedro, Baker, Toto, & Montalvo, 2012; Gobert, Sao Pedro, Raziuddin, & Baker, 2013). The purpose of this paper is to provide a case study of three innovative projects that are all actively utilizing immersive virtual environments as situated spaces for the assessments of learning, assessments as learning, and assessments for learning.

**Assessment of Learning**

Assessment of learning, or summative assessment, refers to those strategies designed to confirm what students know at a particular moment in time. These strategies usually come in the form of paper-and-pencil tests implemented by teachers as a snapshot of student achievement for grading and reporting purposes. Summative assessments also serve to demonstrate whether or not students have met curricular outcomes or the goals of individualized student programs to certify proficiency and make decisions about future programs or placements. In short, summative assessments take place at certain intervals when achievement has to be reported as it relates to the progression in learning against a defined criteria in the form or curricular standards often set by public bodies (Martone & Sireci, 2009). As a further means of assessing program and curricular viability, summative assessments mandated by public bodies as high-stakes tests are designed to provide evidence of achievement to parents, other educators, the students themselves, and to outside groups (e.g. Manitoba Education, 2006). Ongoing arguments in the literature suggest that most valid and reliable summative assessments are in the form of multiple-choice or numeric response questions that can be machine scored. Efforts by researchers have led to significant gains in the use of authentic performance based assessments along side the high-stakes paper-and-pencil tests (Palm, 2008).
Performance Assessment

Alternative performance measures provide a more valid measure than multiple choice tests of students’ conceptual understanding and higher-level skills such as problem solving (Lane & Stone, 2006). Research findings indicate that alternate assessments are more transparent to the content being measured (Lane & Stone, 2006) and valuable for providing formative, diagnostic feedback to teachers about ongoing student attainment. However, there are several limitations to their use as summative assessments in accountability settings: First, hands-on performance assessments (HoPAs) in accountability settings are cost-prohibitive when compared to multiple choice tests (Pellegrino et al., 2001). Second, HoPAs often have issues around task sampling and occasion sampling variability (Cronbach, Linn, Brennan, & Haertel, 1997; Shavelson, Baxter, & Gao, 1993). Finally, HoPAs suffer technical, resource, and reliability problems that undercut both their validity and their practicality (Cronbach et al., 1997; Shavelson et al., 1993; Webb, Schlackman, & Sugrue, 2000). Current research explores the feasibility of using technologies, such as immersive virtual environments, as an answer to scalability, validity, and reliability issues. The Virtual Performance Assessment project is one such research project.

Virtual Performance Assessments (VPA)

The purpose of the Virtual Performance Assessment project (VPA; http://vpa.gse.harvard.edu) is to develop assessments that complement rather than replace existing standardized measures by assessing skills not possible via paper-pencil, multiple-choice or performance-based measures (Clarke-Midura, Code, Mayrath, & Dede, 2012; Clarke-Midura, Code, Zap, et al., 2012; Code, Clarke-Midura, Zap, & Dede, 2011b, 2011c). The VPA project identifies several potential benefits of using immersive virtual environments for performance assessment. First, using
immersive environments alleviate the need for extensive training for teachers to administer performance tasks with students as research demonstrates that it is difficult to standardize the administration paper-based performance assessments as extensive training is often required to administer the tasks. Immersive virtual assessments can help ensure standardization of task and delivery. Second, immersive virtual assessments alleviate the need for providing materials and kits for hands-on tasks as everything student’s need is inside the virtual environment. Third, scoring is all done automatically behind the scenes—there is no need for raters or training of raters. Finally, immersive virtual assessments alleviate safety issues and inequity due to lack of resources.

For each of the three assessments developed in the VPA project, the main goal is to assess students problem solving capabilities by allowing students to make choices that advance the hypothesis and theory that they are attempting to build (Clarke-Midura, Code, Zap, et al., 2012). The three assessments are built around two different paradigms of learning: instructivist and constructivist. The first assessment, *Save the Kelp!*, uses an instructivist paradigm where students were guided through the science inquiry process. The second and third assessments, *There’s a New Frog in Town* (HGSE, 2013) and *Silence of the Bees* (HGSE, 1997), utilized more of a constructivist paradigm where students were able to freely explore the environment and practice inquiry though in-world actions and interactions. Regardless of philosophical paradigm in all three assessments the measure of a student’s science inquiry performance is based on their in-world actions (see Figures 1 and 2; Code, Clarke-Midura, Zap, & Dede, 2013). As students make choices, in-world actions and choices are captured and given a range of scores and weightings that contribute to an on-going student model of science inquiry. In this model, student’s actions are temporally evaluated based on past, present, and future actions. In other
words, a choice is evaluated in terms of the previous actions, their actual choice within the context of the available choices, and the outcome of their choice that sets the stage for the next set of actions. For example, if a character asks a student what they think the problem is and the student responds that they think the mutant frog is a result of pollution, the character will ask the student to provide evidence for their claim. The evidence that a student gives is weighted and evaluated based on their prior actions (data that they have previously collected) and by what they choose to present as evidence. Although both assessment paradigms evaluate the same learning outcomes and that are tracked, measured, and scored in the same way, the user experience is qualitatively and markedly different.

![Figure 1. Presentation of competing hypotheses in a VPA environment.](image1)

![Figure 2. Data collection (e.g. water samples and tadpoles) in the VPA environment.](image2)

Results from the research published to date using the second and third VPA assessments, reveal that the most successful students a) did a lot more external research in all areas and b) did a better job of supporting their claim with evidence (McCall & Clarke-Midura, 2011, 2013). Whereas unsuccessful students a) did not visit the Internet kiosk as much (external research), b) tend to engage in confirmatory activities, and c) were not skilled in reasoning from evidence (McCall &
Further, in relation to the engagement and in-world game behavior, results suggest that game behavior is not a good predictor of claim correctness as some students with high engagement and good reasoning chose the incorrect claim, and some people with low engagement and poor reasoning chose the correct claim. Although research is ongoing, one of the most salient features that is lacking in this approach is the one-way interaction between the student and the environment. This two-way interaction is something that students are expecting more of considering the technology this kind of environment is built upon (Code et al., 2011b, 2013). In the next section, a case study of another project involves the students more intimately in their immersive assessment experience, engaging them in assessment as learning.

**Assessment as Learning**

The focus of self-assessment by students is not common practice, even amongst those teachers who take assessment seriously (Black & Wiliam, 1998). Despite significant evidence in the research that suggests that self-assessment improves persistence, self-efficacy, self-direction, achievement, and enhances metacognition, it is an underutilized tool in the classroom (Black, McCormick, James, & Pedder, 2006). Self assessment involves students in making judgments about their own achievement and learning processes giving them an active role in decisions about action for further progress in learning (Sebba et al., 2008). In a systematic review of the research on student self and peer assessment, Sebba et al. (2008) reveal that there is evidence that self assessment improves: a) student attainment across a range of subject areas, b) student self-esteem, and c) students ability in learning to learn such as with goal setting, clarifying objectives, taking responsibility for learning, and increased confidence. Sebba et al. (2008) also found that self and peer assessment are more likely to impact on student outcomes when there is a move from a dependent to an interdependent relationship between teachers and students which enables
teachers to adjust their teaching in response to student feedback. Thus “require[ing] learners to exercise a degree of autonomy from the teacher as the assessor and judge of quality” (Black et al., 2006, p. 128). Ultimately, self-assessment means involving students in ‘co-designing’ the curricula for evaluation helping them to develop a better grasp of their own strengths and weaknesses. From the perspective of self-assessment, one such project placing students at the center of learning is the Transforming Engagement of Students in Learning Algebra project.

**Transforming Engagement of Students in Learning Algebra (TESLA)**

The purpose of the Transforming Engagement of Students in Learning Algebra project (TESLA; HGSE, 2013; Figures 3 and 4; http://tesla-project.org) is to investigate the relationship between specific technology-based motivational activities and students’ interests in Science, Technology, Engineering, and Mathematics (STEM) careers along a developmental span (Chen, Zap, & Dede, 2012). Utilizing Expectancy-Value theory, TESLA explored how immersive virtual environments can be designed to promote students’ interest and motivation. The TESLA environment accompanies a four-day blended experience that integrates technology-based activities with classroom-based instruction for 5th grade students in Mathematics. The immersive virtual activities the students are involved in (see Figures 1 and 2 as examples) enable them to take on the identity of a STEM professional to solve a number of puzzles. Each of the puzzles must be solved correctly before the students can continue on in the environment.
One of the critical features in the research design of TESLA project is that it is hypothesized that students attribute their successes and failures to factors that they could control thus highlighting that successful completion of the puzzles is not dependent on some innate mathematical intelligence that is completely out of their personal control (Chen et al., 2012). Although the initial intention of the TESLA project was not necessarily aimed to serve specifically as a self-assessment, in effect it does just that, through the use of scaffolds assisting students in the solving of mathematical puzzles embedded within the curriculum. This highlights that the use of scaffolds as a means of formative feedback is a critical factor for student success enabling the use of assessment as an embedded learning experience.

**Assessment for Learning**

Current assessment approaches are inadequate at diagnosing how students develop critical thinking, problem solving, and sophisticated scientific reasoning which are all key 21st century skills (BC Premiers Technology Council (BCPTC), 2010; Partnership for 21st Century Skills (P21), 2009). Formative assessments carried out during instruction can be used to help teachers
tailor instruction; aid in deepening students’ understanding, and enable students to self-regulate their learning. Formative assessments however, are often too time consuming for teachers to do on a regular basis (Black & Wiliam, 1998). Teacher-made assessments tend to replicate multiple choice and open-response tests designed for benchmark or accountability settings and rarely tell teachers what they need to know about their students’ thinking (Black & Wiliam, 2009). Such benchmark tests are also typically designed based on psychometric models that are better suited for assessing curricular program outcomes rather than based on a cognitive model of how students learn. This further contributes to the failure to formatively assess students’ learning at the classroom level leading to a decrease in critical thinking and problem solving scores on a national scale. For example, on the 2010 Pan-Canadian Assessment Program (PCAP) of Mathematics, Science and Reading, middle school students from provinces and territories outside of Ontario and Quebec scored significantly below the national average on problem solving ability (Council of Ministers of Education Canada (CMEC), 2010). Similarly, on the 2009 Program for International Student Assessment (PISA), the average score of Canadian fifteen year olds on the use of elaboration and summarization strategies (critical skills for problem solving, science inquiry and the self-regulation of learning) is significantly below the international average of participating OECD nations (CMEC, 2011). In looking at the literature, there is a direct link between student use of self-regulated learning (SRL) strategies (such as elaboration and summarizing) and the development of complex cognitive reasoning and problem solving specifically in science (e.g. Winters & Azevedo, 2005). The fact that students in Canada are behind their international peers on the use of self-regulated learning strategies (i.e. elaboration and summarizing) is alarming and suggests that this will have a significant impact on their ability to conduct meaningful scientific inquiry.
Formative assessments have the potential to provide important feedback to both teachers and students. This feedback is critical in helping teachers adapt instruction so students can overcome any misconceptions they have in moving along a learning progression. Teachers need tools to adequately identify, measure, and evaluate what individual students know and do not know during learning. Without the aid of technology, formative assessment is difficult to accomplish regularly during classroom instruction. Research findings suggest that formative assessments delivered using technology have the capability to record very detailed observations of students’ actions not possible via paper-and-pencil tasks or through anecdotal teacher observations, as well as to adapt to students’ responses in real time (Pellegrino et al., 2001). “Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited” (Black & Wiliam, 2009, p. 9). A limitation of current teacher developed assessments include an over-emphasis on rote learning and recall of facts which has led to a focus on grades and class level learning instead of individual learning that emphasize completing tasks as opposed to taking risks and engaging in cognitive activities. The quality of interactive feedback in formative assessment is a critical feature in determining the quality of overall teaching and learning activity, and is therefore a central feature of pedagogy (Black & Wiliam, 2006). Ironically, research on the success of feedback is mixed because it is often not used correctly (Shute, 2008). In her comprehensive literature review on feedback, Shute (2008) claims that formative feedback should address the accuracy of a learner’s response to a problem or task and may touch on particular errors and misconceptions. Since research establishes that immersive environments have considerable potential to assess and scaffold science inquiry learning (Code, Clarke-Midura, Zap, & Dede, 2011a; Code et al., 2011c, 2013), the Assessment
The Assessment for Learning in Immersive Virtual Environments (ALIVE) project is a research program in the pilot stage (see Figures 5 and 6) that aims to examine how three-dimensional immersive virtual environments (3IVEs; \thrīvz\) can enable student success by (a) providing formative feedback embedded during the process of learning; (b) visualizing students’ cognitive models of problem solving in situ; and (c) scaffolding students’ self-regulated learning. Self-regulated learning is an active, effortful process in which learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behaviour (Pintrich, 2000). Specifically, this project aims to investigate ways that the 3IVE technology, designed around a model of science inquiry, enables individual students to (a) utilize formative feedback to adapt their learning behavior; (b) engage with log-file data-based visualizations; and (c) monitor their 3IVE activities to regulate their cognitive and behavioral processes to maximize learning success.
The ALIVE project integrates and builds upon TESLA and VPA by utilizing 3IVE technology as a means to visualize and formatively assess science inquiry enabling a student to self-regulate their learning. The specific goals of the ALIVE project are to:

1. Focus on the use of 3IVE technologies to scaffold science inquiry and self-regulated learning strategies through the use of formative assessment and feedback.
2. Explore whether 3IVEs can enable the automated, invisible, and non-intrusive collection of students’ actions and behaviors during the act of learning that will enable the capture and assessment of science inquiry in situ.
3. Use well established models of scientific inquiry to examine the use 3IVEs to assess learning processes as they happen and how students use this feedback to regulate their learning.
4. Design the 3IVE technology to non-intrusively collect log-file data of students’ actions and behaviors during the act of learning and present this data as a means to provide feedback that helps students regulate their learning and teachers regulate their teaching.
It is the intention of the ALIVE project to design and develop 3IVEs to use as a formative assessments to research whether this feedback can help address students’ misconceptions in science inquiry and help students regulate their learning and teachers adequately address what students know and do not know about science inquiry processes. The 3IVE will enable the capture of rich sets of student behavioural observations (log-file or trace data) in a database, which can then be analyzed using various learning analytic techniques with the intention of diagnosing student misconceptions in scientific inquiry.

**Directions for Future Research**

Formative assessment provides feedback that can be used by teachers to regulate instruction, by students to regulate their learning, and by classroom peers to collaboratively regulate, monitor and overcome any misconceptions in complex reasoning. Research needs to specifically explore how teachers can use these measures to: 1) clarify and share learning intentions and criteria for success; 2) engineer effective classroom discussions and other learning tasks that elicit evidence of student understanding; 3) provide feedback that moves learners forward; 4) activate students as instructional resources for one another; and 5) empower students as the owners of their own learning (Black & Wiliam, 2009). Authentic assessment requires students to apply knowledge and reasoning to situations similar to those they will encounter in the world outside the classroom (NRC, 1996; Palm, 2008). Since existing assessment frameworks do not provide information on how learning processes develop, a cognitive model is necessary to examine these processes in situ. Projects such as ALIVE aims to use established cognitive models to examine whether immersive virtual environments can assess learning as it happens and how students utilize feedback to regulate their learning goals.
References


the National Academies Board on Science Education Workshop on Learning Science: Computer Games, Simulations, and Education, Washington, DC.


