Chapter 5
Assessment for Learning in Immersive Environments

Valerie Shute, Seyedahmad Rahimi and Benjamin Emihovich

Abstract Immersive Environments (IEs) hold many promises for learning. They represent an active approach to learning and are intended to facilitate better, deeper learning of competencies relevant for success in today's complex, interconnected world. To harness the power of these environments for educational purposes (i.e., to support learning), we need valid assessments of the targeted competencies. In this chapter we focus on how to design and develop such valid assessments, particularly those providing an ongoing, unobtrusive collection and analysis of data as students interact within IEs. The accumulated evidence on learning thus provides increasingly reliable and valid inferences about what students know and can do across multiple contexts. This type of assessment is called "stealth assessment" and is applied toward the real-time measurement and support of learning in IEs—of cognitive and non-cognitive variables. The steps toward building a stealth assessment in an IE are presented through a worked example in this chapter, and we conclude with a discussion about future stealth assessment research, to move this work into classrooms for adaptivity and personalization.

Keywords Augmented reality · Diagnostic assessment · Immersive environments
Stealth assessment · Digital games · Virtual reality

V. Shute (✉) · S. Rahimi · B. Emihovich
Educational Psychology and Learning Technologies, Florida State University, Stone Building, 1114 West Call Street, 3064453, Tallahassee, FL 32306-4453, USA
e-mail: vshute@fsu.edu
URL: http://myweb.fsu.edu/vshute/
URL: http://sahmadrahimi.wixsite.com/mysite

S. Rahimi
e-mail: sr13y@my.fsu.edu
URL: http://sahmadrahimi.wixsite.com/mysite

B. Emihovich
e-mail: bemihovi@gmail.com

D. Liu et al. (eds.), *Virtual, Augmented, and Mixed Realities in Education*,
Smart Computing and Intelligence, DOI 10.1007/978-981-10-5490-7_5

5.1 Introduction

In this chapter, we examine immersive learning environments (e.g., virtual reality, augmented reality, and digital games) and techniques toward the measurement and support of knowledge and skills therein. Our premise is that immersive environments (IEs) represent an active approach to learning and should thus facilitate better, deeper learning of competencies relevant for success in today's increasingly complex world. Such environments also permit the application and practice of competencies in relatively safe and authentic spaces. Moreover, well-designed IEs that incorporate theoretically-grounded learning principles (authentic problem solving, rules/constraints, challenge, control, ongoing feedback, and sensory stimulation—see Shute, Ventura, Kim, & Wang, 2014) can be intrinsically motivating and therefore engaging; and student engagement is a key component of learning (Dede, 2009).

The IEs on which we focus are based on learning through experiencing, and understood through the theoretical lenses of constructivism (Piaget, 1973) and situated learning (Lave & Wenger, 1991). These theories emphasize active learners who construct meaning (Vygotsky, 1978). Constructivism states that effective learning environments are interactive places where learners achieve learning goals by collaborating with tools, information resources, and with others. Situated learning views cognition as being nestled within the activity, context, and culture in which it is developed. The learner is active in the learning process, where "doing" is more important than listening, and the learner determines the pace of learning.

Constructivism and situated learning are not, however, solely cognitive in nature as affect and cognition are complementary processes within all forms of learning. For example, cognitive complexity theory predicts that well-designed IEs facilitate learning by simultaneously engaging students' affective and cognitive processes (Tennyson & Jorczak, 2008). Affective processes are dependent on how environmental stimuli engage the student.

Similarly, flow theory (Csikszentmihalyi, 1990) argues that flow—a positive experience associated with immersive environments—is an optimal learning state induced by intrinsic motivation, well-defined goals, appropriate levels of challenge, and clear and consistent feedback. Attaining a state of flow involves motivation, effort, and sustained attention thus there is a convergence between the core elements of well-designed IEs and the characteristics of productive learning (Shute et al., 2014).

The purpose of this chapter is to describe how to design and develop valid assessments to support learning in immersive environments, particularly in well-designed digital games. The basic idea is that learners' interactions within such environments generate large amounts of data—cognitive and non-cognitive—which may be captured in log files and analyzed to yield cumulative estimates of current states of targeted competencies (Shute, Leighton, Wang, & Chu, 2016a). The results of the ongoing analyses can be used as the basis for feedback and other types of learning support, such as adapting the environment to fit learners' needs.

In the following sections of this chapter, we review the relevant literature on IEs and their effects on learning, examine the role of diagnostic assessment in immersive learning environments by introducing stealth assessment, provide an example of stealth assessment within a well-designed game, and discuss next steps in this research. Our overarching thesis is that: (a) learning is at its best when it is active, goal-oriented, contextualized, and motivating; and (b) learning environments should thus be interactive, provide ongoing feedback, capture and hold attention, and have appropriate and adaptive levels of challenge. Advances in technology, the learning sciences, and measurement techniques help to support these features through the design of IEs with deeply embedded assessment of targeted competencies.

5.2 How Does Immersion Improve Learning?

In this chapter, immersion refers to the subjective impression one experiences when interacting with a realistic, digitally-enhanced environment (Dede, 2009). Immersion may be experienced within contexts such as: (1) Virtual Reality (VR), where learners wear VR gear and go into an immersive computer-generated world with the illusion of "being there" or having a sense of presence, with immediate adjustments of the environment according to the learner's head or body movements; (2) Multi-User Virtual Environment (MUVE), where learners can enter a 3D virtual world with their digital avatars and virtually interact with other people (Hew & Cheung, 2010); and (3) Mixed Reality (MR) or Augmented Reality (AR), that combines digital information (e.g., images, videos, 3D objects, and audio layers) with real-world settings, and allows users to interact in real-time within a rich immersive experience (Barfield, 2015). Well-designed digital games can provide immersive experiences in any of these three types of environment.

Interactions within an immersive environment produce a suspension of disbelief for learners (i.e., sacrificing realism and logic for the sake of enjoyment) that can be further enhanced when the immersive environment incorporates design strategies that emphasize actional, symbolic, and sensory elements (Dede, 2009). One clear benefit of immersive environments is that they allow participants to safely engage in actions that might be considered too risky or difficult in natural environments (actional immersion). For example, training medical students on triage processes is difficult due to the constraints in which activities undertaken during training reflect the natural world conditions where triage is needed, such as a natural disaster or a plane crash. Replicating the realism and extent of injuries along with patient deterioration using natural world training is both expensive and incompatible for an individual learning experience. Given the natural world restrictions of triage training, researchers designed, built, and tested an immersive game to support learning about how to conduct a triage sieve, as taught in a Major Incident Medical Management and Support Course (MIMMS) in the United Kingdom.

The game, *Triage Trainer* (Knight et al., 2010), was evaluated relative to its effectiveness, compared to traditional learning methods (i.e., card sorting exercises). A total of 91 participants (i.e., 44 in the card-sorting group and 47 in the *Triage Trainer* group) were tested on their ability to correctly prioritize each casualty (tagging accuracy) as well as follow the procedure correctly (step accuracy). According to Knight et al. (2010), participants using *Triage Trainer* performed significantly better than the card-sorting group for tagging accuracy ($\chi^2(5) = 13.14$, $p < 0.05$) (i.e., 72% compared to 55%, respectively). In addition, the step accuracy results indicated four times as many participants in the *Triage Trainer* group (28%) correctly triaged all eight of the casualties compared to the card-sorting group (7%), and significantly more participants in the *Triage Trainer* group scored the maximum compared to the card-sorting group ($\chi^2(1) = 5.45$, $p < 0.05$).

In addition to cognitive effects, well-designed digital games that fully immerse learners in environments often elicit affective reactions (e.g., excitement, boredom, confusion, frustration) that differentially influence learning, such as the development of problem-solving skills and spatial abilities (e.g., Shute, Ventura, & Ke, 2015). Furthermore, there are several conditions of gameplay that one can experience in well-designed digital games (e.g., identity formation, persistent problem solving, practice, and interaction) that impact motivation, which in turn promotes engagement and meaningful learning (Clark, Tanner-Smith, & Killingsworth, 2014).

Consider the game *World of Warcraft* (WoW). This is a good example of a fully immersive digital game in which the learning takes place in a goal-driven problem space where players negotiate different contexts (i.e., levels, scenarios, interactions) solving assorted problems with their avatars (Gee, 2008). Playing WoW successfully requires various competencies (e.g., problem-solving skills and collaboration) as well as planning and executing strategies synchronously to accomplish goals. As players traverse each level in WoW, it is natural to reflect on and process gameplay choices, which helps to promote a more motivating gameplay/learning experience. Players additionally enjoy customizing different skills and abilities for their avatars because different combinations of abilities can lead to improved gameplay performance, which results in greater rewards earned.

An example customization by game players includes design modifications to the game that build models to be used for: (1) in-game performance improvement, and (2) addressing a naturally occurring and frustrating in-game problem—i.e., dealing with freeloaders. Thus, to improve in-game avatar performance, WoW players created an add-on modification called *Skada Damage Meter,* which displays how well each person in a group is performing based on feedback that is given to players as a percentage of damage or healing done per avatar. *Skada Damage Meter* displays a chart with various metrics such as overall damage done, damage per minute, overall healing done, and healing per minute. These metrics enable group leaders to identify which players are underperforming based on their avatar role (i.e., damage absorber, damage dealer, and healer). Developing this modification illustrates how players were sufficiently motivated to solve a WoW problem, which has a real-world parallel in workplace environments (i.e., individuals who attempt benefit from the success of others by trying to obscure their incompetent skills).

In addition to promoting problem-solving skills through gameplay, immersive games can serve as learning vehicles to support the development of knowledge and skills across various domains including: inquiry-based science learning with Quest Atlantis (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) and River City (Ketelhut, 2007), spatial skills with Portal 2 (Shute et al., 2015), and computational problem-solving (Liu, Cheng, & Huang, 2011) with TrainB&P (Train: Build and Program it). Immersion fosters learning by enabling multiple frames of reference and situated learning experiences (Dede, 2009). These multiple frames of reference provide different benefits for immersive learning. For instance, egocentric frames of reference support immersion and motivation through embodied learning, while exocentric frames of reference support abstract symbolic insights when one is further from the context of the environment.

Immersive environments also enhance a contextualized understanding of instructional content for learners in ways that are often decontextualized in formal learning settings. As mentioned earlier, these environments support meaningful learning experiences that are grounded by situated learning and constructivism learning theories. Situated learning is an active process that can generate excitement and curiosity in the learner to acquire knowledge by constructing meaning through specific problem-solving scenarios (Barab et al., 2005). Situated learning can also involve the adoption of multiple roles and perspectives and receiving guidance from expert modeling (Bransford, Brown, & Cocking, 2000). Through immersive interactions and gameplay, novice players can develop their skills by observing, communicating and interacting with other expert players; essentially emulating how junior scholars learn from their advisors in academic environments. In addition, players acquire in-game terminology through interactivity and communication with experts and novices, and language acquisition is an essential element to scaffolding.

Finally, immersive gameplay or other in situ interactions enable learners to traverse the zone of proximal development (ZPD; Vygotsky, 1978), which refers to the distance between what a learner can do with support by collaborating with peers or through guided instruction, and what they can do without support. The acquisition of knowledge begins with interaction, followed by the acquisition of language which provides meaning and intent so that behaviors can be better understood. Towards that end, well-designed IEs consist of rules, goals, feedback, skill mastery, and interactivity. To achieve quantifiable outcomes, players must acquire knowledge, skills, and other abilities. Immersive environments like digital games also promote play which is integral for human development, and is vital to assimilating and accommodating new information by interacting with a fluid environment (Shute et al., 2015). Well-designed IEs promote learning by requiring learners to apply various competencies (i.e., creativity, rule application, persistence) to solve novel problems thereby providing meaningful assessment environments for learners during gameplay. So how can these evolving competencies be accurately measured and thereby used as the basis for good evidence-based learning support?

5.3 Assessment in Immersive Environments

Assessment of student learning in IEs should not be measured using traditional summative tests (Shute, Leighton et al., 2016a). Such standardized tests provide a very narrow snapshot of student learning. Moreover, traditional assessments cannot provide immediate feedback to support learning or adapt the environment to learners' needs. Therefore, it is not surprising that the question of how to administer responsive, comprehensive, and balanced assessments within IEs is an emergent and complex question. Immersive environments provide novel learning opportunities that demand new assessment methodologies.

Shute (2011) used the term "stealth assessment" to refer to evidence-based, ongoing, and unobtrusive assessments, embedded within IEs (e.g., digital games, virtual reality, augmented reality). Stealth assessments capture, measure, and support the development of learners' targeted competencies in IEs which serve as vehicles for learning. Stealth assessment can be used to adapt the environment to accommodate learners' current levels/needs, as well as to provide appropriate feedback and other types of learning support (Shute, Ke, & Wang, 2017). According to Csikszentmihalyi (1990), such personalized support permits learners to maintain the state of flow (note: adaptivity is further discussed in the next section).

As a learner interacts with the IE (e.g., an augmented reality activity or video game), stealth assessment analyzes specific actions and interactions via data that are captured in the log file to estimate the learner's competency states in terms of evidence-based claims. Stealth assessment creates a student model and continuously updates it as the person interacts with the IE. Information from the student model, then, can be used as the basis for which to provide relevant feedback and/or adapt the IE to suit the learner's needs. In the process, this creates a personalized learning/playing experience.

Stealth assessment employs a principled assessment design framework called evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003). ECD involves the development of conceptual and computational models (e.g., the competency, evidence, and task models) that work together to accomplish valid assessment (see Fig. 5.1).

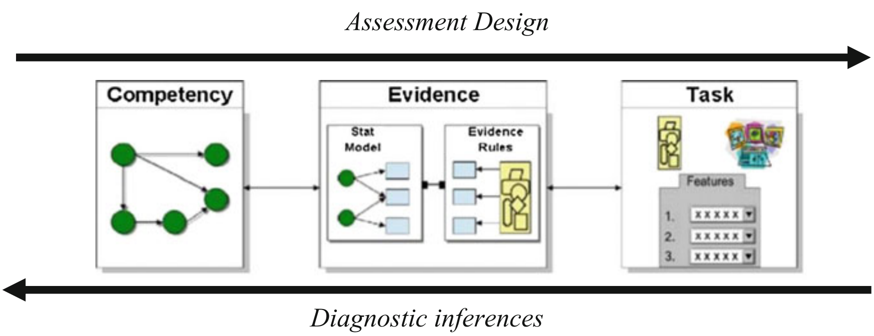

Fig. 5.1 Simplified ECD adapted from Mislevy et al. (2003)

The first model in ECD framework is the competency model, which explicitly specifies the knowledge, skills, and other attributes (collectively referred to as "competencies" in this chapter) to be measured by the assessment. This is intended to facilitate the operationalization of the construct with all of its associated facets and observable behaviors. The second model is the evidence model, which specifies the assignment of scores to the observable behaviors (i.e., the learner's performance), such as whether dichotomous (i.e., an item response or activity is assigned a value of 1 if correct, otherwise a 0) or polytomous (i.e., an item response or activity is assigned values other than just 0 or 1 to show increasing performance quality) scoring will be used, and how the scores will be accumulated. Finally, the third model is the task model, which outlines the types of tasks, including all features, requiring development to elicit the competencies of interest from the learner.

Stealth assessment's evidence-based models work together to accomplish ongoing analyses of all gameplay/interaction data. This provides more valid and reliable assessment results compared to traditional summative tests. Shute et al. (2017) delineate the steps for creating a stealth assessment in an IE:

1. Develop the competency model (CM) of targeted knowledge, skills, or other attributes based on comprehensive literature and expert reviews
2. Determine the IE (e.g., a game or other immersive media applications) into which the stealth assessment will be embedded
3. Create a full list of relevant actions/indicators that serve as evidence to inform the CM and its facets
4. Create new tasks in the IE, if necessary
5. Create a matrix to link actions/indicators to relevant facets of target competencies
6. Determine how to score indicators by classifying them into discrete categories for the "scoring rules" part of the evidence model (EM)
7. Establish statistical relationships between each indicator and associated levels of CM variables using, for example, Bayesian Networks (BNs) (EM)
8. Pilot test the BNs and modify parameters
9. Validate the stealth assessment with external measures
10. Use the assessment estimates to provide feedback and targeted learning supports in the IE.

We now examine a worked example of a stealth assessment of problem-solving skills that was developed and used within a modified version of a popular immersive 2-dimensional game based on the steps described above.

5.4 An Illustration of Stealth Assessment in a Game Environment

To make the process of creating a stealth assessment come alive, we present an example in which a problem-solving stealth assessment was developed and built into a game called "Use Your Brainz" (UYB; a modified version of the game Plants vs. Zombies 2; Shute, Wang, Greiff, Zhao, & Moore, 2016b). In the game, players position a variety of special plants on their lawn to prevent zombies from reaching their house. Each of the plants has different attributes. For example, some plants (offensive ones) attack zombies directly, while other plants (defensive ones) slow down zombies to give the player more time to attack the zombies. A few plants generate "sun," an in-game resource needed to utilize more plants. The challenge of the game comes from determining which plants to use and where to position them on the battlefield to defeat all the zombies in each level of the game.

To create a stealth assessment measuring problem-solving skills, Shute and colleagues first developed a competency model of problem solving based on an extensive literature review (step 1). The operationalized problem-solving CM included four main facets: (a) analyze givens and constraints, (b) plan a solution pathway, (c) use tools effectively/efficiently when solving the problem, and (d) monitor and evaluate progress. In parallel with developing the problem-solving CM, Shute and her team selected an appropriate IE (the UYB game) in which to embed the stealth assessment (step 2). They selected this game for several reasons. First, UYB requires ongoing problem-solving skills (like chess). Second, although it is a 2D game, it can provide an immersive experience in that its engaging environment requires players to continuously apply the various in-game rules to solve challenging problems. Third, this work was part of a joint project with GlassLab (see https://www.glasslabgames.org/), and Glasslab had access to the game's source code which allowed the researchers to modify the data to be captured in the log files and embed the stealth assessment models directly into the game.

After finalizing the problem-solving competency model, Shute and her team identified dozens of observable in-game indicators (after repeatedly playing the game and watching expert solutions on YouTube). The indicators are used as evidence to update the problem-solving CM (step 3; in this example step 4 was not needed). For example, the research team determined that planting three or more sun-producing plants (which provide the currency to use other plants) before the first wave of zombies arrive is an indicator of the "analyze givens and constraints" facet and shows that the player understands time and resource constraints. Table 5.1 includes some examples of problem-solving indicators in UYB.

Table 5.1 Example indicators for problem solving (from Shute, Wang, et al., 2016b)

Facets	Example indicators
Analyze givens and constraints	• Plants >3 Sunflowers before the second wave of zombies arrives • Selects plants off the conveyor belt before it becomes full
Plan a solution pathway	• Places sun producers in the back, offensive plants in the middle, and defensive plants up front/right • Plants Twin Sunflowers or uses plant food on (Twin) Sunflowers in levels that require the production of X amount of sun
Use tools and resources effectively/efficiently	• Uses plant food when there are >5 zombies in the yard or zombies are getting close to the house (within 2 squares) • Damages >3 zombies when firing a Coconut Cannon
Monitor and evaluate progress	• Shovels Sunflowers in the back and replaces them with offensive plants when the ratio of zombies to plants exceeds 2:1

The next task in the UYB project was to create a Q-matrix (Almond, 2010) with the four problem-solving facets in columns and all of the relevant indicators listed in rows (step 5; where the crossed cells contain the value of "1" if the indicator is related to the facet and a "0" if they're unrelated). Afterwards, they determined the scoring rules (step 6). This entails deciding about how to score the indicators by classifying them into discrete categories (e.g., yes/no, high/medium/low relative to the quality of the actions). For example, if a player planted six sunflowers before the second wave of zombies, the action will be automatically recorded as "yes" providing positive evidence of the first facet "analyze givens and constraints."

After categorizing all indicators, Shute and her team connected each indicator to the related CM variable(s) and established a statistical relationship between them (step 7). They used Bayesian Networks to create the statistical relationships, accumulate the incoming gameplay data, and update the beliefs in the competency model (note: they created one BN for each level, 43 BNs in total). Why were BNs used over other techniques? De Klerk, Veldkamp, and Eggen (2015) conducted a systematic literature review on various analytical approaches used in simulation-based and game-based assessments to analyze performance data (i.e., the data generated by learners' interaction with the IE). The most prevalent examples of such analytic tools include Bayesian Networks (BNs), Exploratory and Confirmatory Factor Analysis, Item Response Theory, Multidimensional Item Response Theory, Cluster Analysis, Artificial Neural Networks, and Educational Data Mining. Overall, BNs were the most used analytical and data modeling framework to analyze learners' performance data in game-based and simulation-based assessment. Moreover, there are several advantages to using BNs as a data modeling framework in IEs such as: (1) BNs provide an easy-to-view graphical representation of the competency model (direct and indirect relationships among variables) for clear operationalization; (2) BNs can "learn" from data as they're probability models (thus make probabilistic predictions)—the degree to which observed data meet expectations of the model can help improve the original model as more data become available; (3) Updating BNs is immediate (as performance data come from the IE) compared to other analytical approaches (like IRT), so they provide real-time diagnosis—overall and at sub-score levels; and (4) Enhancements to BN software permit large and flexible networks with as many variables as wanted (Almond et al., 2015). Moreover, by using only discrete variables, BNs can be scored very quickly, making them suited for embedded scoring engines.

Consider indicator #37 in Fig. 5.2 (use of iceberg lettuce in UYB). This indicator is connected to the "tool use" facet, and a player has just performed some action in the game which was judged as "poor" (e.g., placed an iceberg lettuce proximal to a fire-breathing plant, thus cancelling out the "freezing" effect of the lettuce). The real-time estimate that the learner is low on the "tool use" facet is $p = 0.61$ (for more details see Shute, Wang, et al., 2016b).

When establishing the BNs for UYB, the game experts and psychometricians in the team initially set the probabilities of the various states, per competency model variable (i.e., the prior probabilities in BNs). However, after pilot testing the BNs,

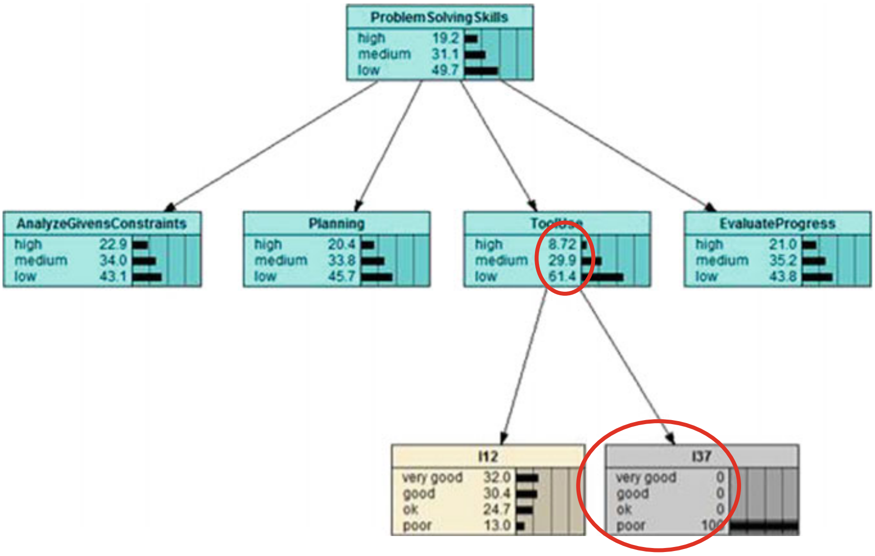

Fig. 5.2 An example of a BN with data for indicator #37 entered (poor use of iceberg lettuce)

data were used to modify the BN parameters (difficulty and discrimination) accordingly (step 8). Finally, to validate the stealth assessment using external measures (step 9), Shute and colleagues used two external measures of problem solving skill: (a) Raven's Progressive Matrices (Raven, 1941) which examines inductive ability (i.e., rule identification) based on given information; and (b) MicroDYN (Wustenberg, Greiff, & Funke, 2012), a simulation which measures problem solving skills based on acquiring and applying existing information (i.e., rule application).

Validation study participants were 7th grade students (n = 55) from a middle school in suburban Illinois. They students played the game for 3 h (1 h per day across three consecutive days). The results showed that the students' scores from the two external tests significantly correlated with the in-game stealth assessment estimates [Raven's ($r = 0.40$, $p < 0.01$) and MicroDYN ($r = 0.41$, $p < 0.01$)]. Therefore, the stealth assessment embedded in the UYB game appears to be valid. Other studies have been conducted using stealth assessment to measure various competencies, e.g., physics understanding (Shute et al., 2013) and persistence (Ventura & Shute, 2013). The overall findings from these studies also show significant correlations between external and the in-game estimates. Finally, it's important to note that this assessment approach, while illustrated in a 2D environment, can also be used in 3D games and environments (e.g., Portal 2 research, see Shute et al., 2015). We now discuss the next logical steps to take—making IEs adaptive based on assessment data.

5.5 Next Steps

After creating and embedding a stealth assessment into an IE and testing its psychometric properties (i.e., reliability, validity, and fairness), the next step is to provide adaptive or personalized learning supports (e.g., appropriate feedback and challenges) based on current estimates of competency states (Shute et al., 2017). This type of adaptation (i.e., *micro-adaptation*; see Kickmeier-Rust & Albert, 2010) keeps learners motivated to progress throughout the game/IE, engenders a state of flow, and aligns with their ZPD.

As mentioned earlier, Csikszentmihalyi (1990) asserted that when learners are fully engaged in tasks that are neither too difficult nor too easy, they enter the state of flow in which they learn best. Similarly, Vygotsky (1978) believed that the best learning experience happens when learners receive learning materials just beyond their current knowledge or skill level. Research has shown that adaptive learning activities generally yield better learning outcomes than non-adaptive activities (e.g., Kanar & Bell, 2013). We suspect that similar learning outcomes can be achieved via adaptive IEs. Moreover, learning/playing in an adaptive IE can facilitate learners' self-efficacy (Bandura, 1994) and self-determination (Ryan & Deci, 2000) because learners establish new beliefs about their personal capabilities when they progressively tackle challenges that are tailored to their current ability levels. In other words, the more learners overcome appropriately-challenging tasks, the more efficacious they feel in the IE in which they interact. The gratifying experience of efficacy makes the learners intrinsically motivated to continue facing new challenges (Klimmt, Hartmann, & Schramm, 2006).

To enhance learning—both processes and outcomes—learners' state of flow would be maintained by adjusting tasks/activities in the IE coupled with ongoing targeted feedback. In theory, this would motivate them to persist and enhance their self-efficacy (e.g., Van Oostendorp, van der Spek, & Linssen, 2013). To accomplish this goal, accurate, ongoing, and unobtrusive measurements of learners' current competency states (relative to cognitive, non-cognitive, and even affective variables) are needed to continuously adapt the IE to the learners' needs and capabilities in real-time. Research is needed on how to best prioritize the skill or affective state most in need of support.

One way to accomplish adaptation in an IE is via a task selection algorithm. For instance, Shute, Hansen, & Almond (2008) developed an adaptive algorithm that tends to select tasks for which the student has an approximately 50–50 chance of solving correctly. These tasks are likely to reside within the student's zone of proximal development (Vygotsky, 1978) and hence may be good candidates for promoting learning, particularly if accompanied by feedback. In contrast, non-adaptive (e.g., linear) IEs/games may present fixed sequences of activities or tasks, often arrayed from easy-to-difficult. This may lead to predictable and impersonal learning/gameplay experiences (Lopes & Bidarra, 2011) and perhaps boredom. Creating adaptive IEs empowered by stealth assessment is currently under development and we expect to see positive results on students' learning.

5.6 Conclusions

Immersive technologies are now available to use in formal education settings as they become more affordable. Historically, VR has been used in military training for many years, MUVEs have been around for more than fifteen years, and AR has been used in museums, factories, medical arenas, and the military since the early 1990s. Nonetheless, their use in public educational settings has not been feasible due to the cost and availability of the technologies, until recently. Currently, low-cost VR experiences are possible with products like Google Cardboard which only costs $15 and a smart phone (Brown & Green, 2016). Furthermore, according to a recent report (Adams Becker, Freeman, Giesinger Hall, Cummins, & Yuhnke, 2016), large investments are being made in the immersive media industry, and it is expected that the education sector will benefit from these investments within the next two to three years. In another report, Goldman Sachs predicted that the immersive media industry has the potential of being an $80-billion market by 2025 (Bellini et al., 2016).

Because of these trends, many companies (e.g., Facebook, Samsung, Google, and HTC) have entered the race for developing content with advanced technologies to make the immersive media experience possible for all (Brown & Green, 2016). Furthermore, industry leaders recognize the potential benefits of immersive well-designed games just as learning theorists posit that gameplay experiences in immersive environments can substantially improve learners' problem solving skills through multiple interactions with novel problem solving scenarios (e.g., Van Eck & Hung, 2010). However, there are still barriers to adopting IEs in formal education settings—mainly related to getting the assessment part right.

Our broad vision relating to assessment for learning involves the ongoing collection of data as students interact within various IEs during and after regular school hours. When these various data streams coalesce, the accumulated information can potentially provide increasingly reliable and valid evidence about what students know and can do across multiple contexts. To accomplish this goal, we need high-quality, ongoing, unobtrusive assessments embedded in various IEs that can be aggregated to inform a student's evolving competency levels (at various grain sizes) and aggregated across students to inform higher-level decisions (e.g., from student to class to school to district to state, to country).

The primary goal of this idea is to improve learning, particularly learning processes and outcomes necessary for students to succeed in the twenty first century, such as persistence, creativity, problem solving skill, critical thinking, and other constructs. Current approaches to assessment/testing are typically disconnected from learning processes. With innovative assessment technologies like stealth assessment, teachers do not need to disrupt the normal instructional process at various times during the year to administer external tests to students. Instead, assessment should be continuous and invisible to students, supporting real-time, just-in-time instruction and other types of learning support in all types of IEs.

For this vision of assessment—as ubiquitous, unobtrusive, engaging, and valid—to gain traction, there are a several hurdles to overcome. Several immediate concerns are presented here (for more details on challenges and future research, see Shute, Leighton, et al., 2016a).

1. *Ensuring the quality of assessments*. U.S. schools are under local control, thus students in a given state could engage in thousands of IEs during their educational tenure. Teachers, publishers, researchers, and others will be developing IEs, but with no standards in place, they will inevitably differ in curricular coverage, difficulty of the material, scenarios and formats used, and many other ways that will affect the adequacy of the IE, tasks, and inferences on knowledge and skill acquisition that can justifiably be made from successfully completing the IEs. More research is needed to figure out how to equate IEs or create common measurements from diverse environments. Towards that end, there must be common models employed across different activities, curricula, and contexts. Moreover, it is important to determine how to interpret evidence where the activities may be the same but the contexts in which students are working differ (e.g., working alone vs. working with another student).
2. *Making sense of different learning progressions*. IEs can provide a greater variety of learning situations than traditional face-to-face classroom settings, thus evidence for assessing and tracking learning progressions becomes more complex rather than general across individual students. As a result, we need to be able to model learning progressions in multiple aspects of student growth and experiences, which can be applied across different learning activities and contexts. Moreover, there is not just one correct order of progression as learning in IEs involves many interactions between individual students and situations, which may be too complex for most measurement theories that assume linearity and independence. So theories of learning progressions in IEs need to be actively researched and validated to realize their potential.
3. *Privacy/Security*. This issue relates to the accumulation of student data from disparate sources. However, information about individual students may be at risk of being shared far more broadly than is justifiable. And because of the often high-stakes consequences associated with tests, many parents and other stakeholders fear that the data collected could later be used against the students.

Despite these obstacles, constructing the envisioned ubiquitous and unobtrusive assessments within IEs across multiple learner dimensions, with data accessible by diverse stakeholders, could yield various educational benefits. First, the time spent administering tests, handling make-up exams, and going over test responses is not very conducive to learning. Given the importance of time on task as a predictor of learning, reallocating those test-preparation chores into meaningful pre-instructional activities that are more engaging for learners can benefit almost all students. Second, by having assessments that are continuous and ubiquitous, students are no longer able to "cram" for an exam. Although cramming can provide good short-term recall, it is a poor route to long-term retention and transfer of learning.

Standard assessment practices in school can lead to assessing students in a manner that conflicts with their long-term success. With a continuous assessment model in place, the best way for students to perform well is to engage with the content, interact with peers, and communicate ideas. The third direct benefit is that this shift in assessment mirrors the national shift toward evaluating students on acquired competencies. With increasing numbers of educators growing wary of traditional, high-stakes tests for students, ensuring students have acquired the "essential" skills needed to succeed in twenty first century workplace environments are consistent with the innovative type of assessment outlined in this chapter.

There is a need for innovative assessments given (a) the urgency for supporting new twenty first century skills, and (b) the increased availability of immersive technologies, both of which make it easy to capture the results of routine student work—in class, at home, or any place with available broadband access. One possibility is for twenty first century assessments to be so well integrated into students' day-to-day lives that they are unaware of its existence. This represents quite a contrast to our current testing contexts. However, while the benefits of using a seamless-and-ubiquitous model to run a business have been clear for more than four decades (e.g., using barcodes), applying this metaphor to education may require modifications given the desired outcome is knowledge rather than financial capital. For instance, there are certain risks to consider: students may come to feel like they are constantly being evaluated which could negatively affect their learning by causing unwanted stress. Another risk of a continuous assessment approach in education could result in teaching and learning turning into ways to "game the system" depending on how it is implemented and communicated. But the aforementioned hurdles and risks, being anticipated and researched in advance, can help to shape the vision for a richer, deeper, more authentic assessment (to support learning) of students in the future.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors, however we do acknowledge the intellectual support by Chris Dede and John Richards, as well as various reviewers that we received while writing this chapter.

References

Adams Becker, S., Freeman, A., Giesinger Hall, C., Cummins, M., & Yuhnke, B. (2016). *Horizon Report: 2016 K-12 Edition*. Austin, TX: New Media Consortium.

Almond, R. G. (2010). Using evidence centered design to think about assessments. In V. J. Shute & B. G. Becker (Eds.), *Innovative assessment for the 21st century: Supporting educational needs* (pp. 75–100). New York: Springer.

Almond, R. G., Mislevy, R. J., Steinberg, L., Yan, D., & Williamson, D. (2015). *Bayesian networks in educational assessment*. New York: Springer.

Bandura, A. (1994). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis: A game without guns. *Educational Technology Research and Development, 53*(1), 86–107.

Barfield, W. (2015). *Fundamentals of wearable computers and augmented reality* (2nd ed.). New York: CRC Press.

Bellini, H., Chen, W., Sugiyama, M., Shin, M., Alam, S., & Takayama, D. (2016). *Profiles in innovation: Virtual & augmented reality*. Goldman Sachs group, inc.

Bransford, J., Brown, A., & Cocking, R. (2000). *Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *TechTrends, 60*(5), 517–519. doi:10.1007/s11528-016-0102-z.

Clark, D., Tanner-Smith, E., & Killingsworth, S. (2014). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122. doi:10.3102/0034654315582065.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66–69. doi:10.1126/science.1167311.

De Klerk, S., Eggen, T. J. H. M., & Veldkamp, B. P. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education, 85,* 23–34.

Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (pp. 21–40). Cambridge, MA: The MIT Press. doi:10.1162/dmal.9780262693646.021.

Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology, 41*(1), 33–55.

Kanar, A. M., & Bell, B. S. (2013). Guiding learners through technology-based instruction: The effects of adaptive guidance design and individual differences on learning over time. *Journal of Educational Psychology, 105*(4), 1067.

Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology, 16*(1), 99–111.

Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games. *Journal of Computer Assisted learning, 26*(2), 95–105.

Klimmt, C., Hartmann, T., & Schramm, H. (2006). Effectance, self-efficacy, and the motivation to play video games. In J. Bryant & P. Vorderer (Eds.), *Psychology of entertainment* (pp. 291–313). Mahwah, NJ: Erlbaum.

Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., et al. (2010). Serious gaming technology in major incident triage training: A pragmatic controlled trial. *Resuscitation, 81*(9), 1175–1179.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: University of Cambridge Press.

Liu, C., Cheng, Y., & Huang, C. (2011). The effect of simulation games on the learning of computational problem solving. *Computers & Education, 57*(3), 1907–1918.

Lopes, R., & Bidarra, R. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games, 3*(2), 85–99.

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

Piaget, J. (1973). *To understand is to invent: The future of education*. New York: Grossman.

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology, 19*(1), 137–150.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.
Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction, 55*(2), 503–524.
Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing It–Or can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education, 18*(4), 289–316.
Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters & H. van Oostendorp (Eds.), *Techniques to facilitate learning and motivation of serious games* (pp. 59–78). New York, NY: Springer.
Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. (2016a). Advances in the science of assessment. *Educational Assessment, 21*(1), 1–27.
Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of portal 2 and lumosity on cognitive and noncognitive skills. *Computers & Education, 80,* 58–67.
Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in newton's playground. *The Journal of Educational Research, 106*(6), 423–430.
Shute, V., Ventura, M., Kim, Y., & Wang, L. (2014). Video games and learning. In W. G. Tierney, Z. B. Corwin, T. Fullerton, & G. Ragusa (Eds.), *Postsecondary play: The role of games and social media in higher education* (pp. 217–235). Baltimore, MD: John Hopkins Press.
Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016b). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior, 63,* 106–117.
Tennyson, R. D., & Jorczak, R. L. (2008). A conceptual framework for the empirical study of instructional games. In H. F. O'Neill & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp. 3–20). Boston: Elsevier.
Van Eck, R., & Hung, W. (2010). *A taxonomy and framework for designing educational games to promote problem solving*. In paper presentation at the videogame cultures & the future of interactive entertainment annual conference of the inter-disciplinary. *Net Group,* 227–263.
Van Oostendorp, H., Van der Spek, E., & Linssen, J. (2013). Adapting the complexity level of a serious game to the proficiency of players. *EAI Endorsed Transactions on Serious Games, 1*(2), 8–15.
Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior, 29*(6), 2568–2572.
Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence, 40*(1), 1–14.

Author Biographies

Valerie Shute is the Mack & Effie Campbell Tyner Endowed Professor in Education in the Department of Educational Psychology and Learning Systems at Florida State University. Before coming to FSU in 2007, she was a principal research scientist at Educational Testing Service where she was involved with basic and applied research projects related to assessment, cognitive diagnosis, and learning from advanced instructional systems. Her general research interests hover around the design, development, and evaluation of advanced systems to support learning–particularly related to twenty first century competencies. Her current research involves using games with stealth assessment to support learning—of cognitive and noncognitive knowledge, skills, and dispositions. Her research has resulted in numerous grants, journal articles, books, chapters in edited books, a patent, and a couple of recent books (e.g., Shute & Ventura, 2013, Measuring and supporting learning in games: Stealth assessment, The MIT Press; and Shute & Becker, 2010, Innovative assessment for the twenty first century: Supporting educational needs, Springer-Verlag). She is also the co-founder of www.empiricalgames.org.

Seyedahmad Rahimi is a Doctoral Candidate at Florida State University in the Instructional Systems & Learning Technologies program. He is also getting his second MS degree in Measurement & Statistics at Florida State University. He holds a BS degree in Computer Software Engineering from Islamic Azad University in Iran as well as an MS degree in e-learning technologies from Multimedia University in Malaysia. His Master's thesis was about the perception of college-level students on immersive learning environments like Second Life. He is currently working with Dr. Valerie Shute as one of her Research Assistants on an NSF grant examinining learning supports and adaptivity in games. His general research interests include game-based stealth assessment and adaptivity in learning games.

Benjamin Emihovich is a Doctoral Candidate in the Instructional Systems & Learning Technologies program at Florida State University. He holds a M.Ed. from the University of Florida and a B.A. in Psychology and Social Behavior from the University of California, Irvine. He is interested in exploring how well-designed video games can be used to improve a wide range of knowledge, skills, and abilities referred to as game-based learning (GBL). His dissertation research measures the impact of video gameplay on undergraduates' problem-solving skills. Video games have broad appeal with plenty of research opportunities available to meet the demands of a diverse learner population and those at-risk of failing.