



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



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.



106 The game, *Triage Trainer* (Knight et al., 2010), was evaluated relative to its
107 effectiveness, compared to traditional learning methods (i.e., card sorting exercises).
108 A total of 91 participants (i.e., 44 in the card-sorting group and 47 in the *Triage*
109 *Trainer* group) were tested on their ability to correctly prioritize each casualty
110 (tagging accuracy) as well as follow the procedure correctly (step accuracy).
111 According to Knight et al. (2010), participants using *Triage Trainer* performed
112 significantly better than the card-sorting group for tagging accuracy ($\chi^2(5) = 13.14$,
113 $p < 0.05$) (i.e., 72% compared to 55%, respectively). In addition, the step accuracy
114 results indicated four times as many participants in the *Triage Trainer* group (28%)
115 correctly triaged all eight of the casualties compared to the card-sorting group (7%),
116 and significantly more participants in the *Triage Trainer* group scored the maximum
117 compared to the card-sorting group ($\chi^2(1) = 5.45$, $p < 0.05$).

118 In addition to cognitive effects, well-designed digital games that fully immerse
119 learners in environments often elicit affective reactions (e.g., excitement, boredom,
120 confusion, frustration) that differentially influence learning, such as the develop-
121 ment of problem-solving skills and spatial abilities (e.g., Shute, Ventura, & Ke,
122 2015). Furthermore, there are several conditions of gameplay that one can experi-
123 ence in well-designed digital games (e.g., identity formation, persistent problem
124 solving, practice, and interaction) that impact motivation, which in turn promotes
125 engagement and meaningful learning (Clark, Tanner-Smith, & Killingsworth,
126 2014).

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

138 An example customization by game players includes design modifications to the
139 game that build models to be used for: (1) in-game performance improvement, and
140 (2) addressing a naturally occurring and frustrating in-game problem—i.e., dealing
141 with freeloaders. Thus, to improve in-game avatar performance, WoW players
142 created an add-on modification called *Skada Damage Meter*, which displays how
143 well each person in a group is performing based on feedback that is given to players
144 as a percentage of damage or healing done per avatar. *Skada Damage Meter* dis-
145 plays a chart with various metrics such as overall damage done, damage per minute,
146 overall healing done, and healing per minute. These metrics enable group leaders to
147 identify which players are underperforming based on their avatar role (i.e., damage
148 absorber, damage dealer, and healer). Developing this modification illustrates how
149 players were sufficiently motivated to solve a WoW problem, which has a

150 real-world parallel in workplace environments (i.e., individuals who attempt benefit
151 from the success of others by trying to obscure their incompetent skills).

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

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

178 Finally, immersive gameplay or other in situ interactions enable learners to
179 traverse the zone of proximal development (ZPD; Vygotsky, 1978), which refers to
180 the distance between what a learner can do with support by collaborating with peers
181 or through guided instruction, and what they can do without support. The acqui-
182 sition of knowledge begins with interaction, followed by the acquisition of lan-
183 guage which provides meaning and intent so that behaviors can be better
184 understood. Towards that end, well-designed IEs consist of rules, goals, feedback,
185 skill mastery, and interactivity. To achieve quantifiable outcomes, players must
186 acquire knowledge, skills, and other abilities. Immersive environments like digital
187 games also promote play which is integral for human development, and is vital to
188 assimilating and accommodating new information by interacting with a fluid
189 environment (Shute et al., 2015). Well-designed IEs promote learning by requiring
190 learners to apply various competencies (i.e., creativity, rule application, persistence)
191 to solve novel problems thereby providing meaningful assessment environments for
192 learners during gameplay. So how can these evolving competencies be accurately
193 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).

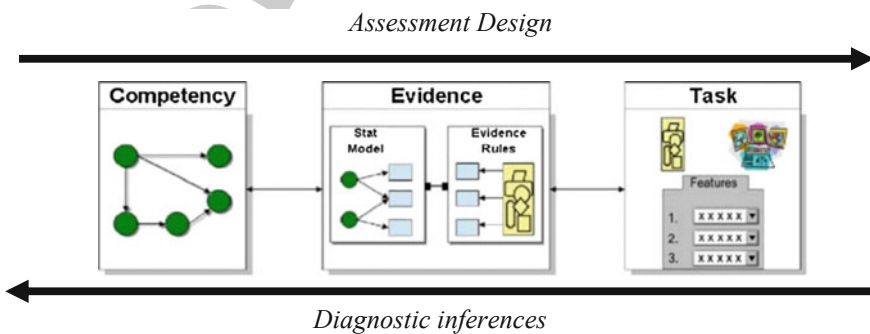


Fig. 5.1 Simplified ECD adapted from Mislevy et al. (2003)



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 which IE (e.g., a game or other immersive media applications) the stealth assessment will be embedded into
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

266 vs. *Zombies 2*; Shute, Wang, Greiff, Zhao, & Moore, 2016b). In the game, players
 267 position a variety of special plants on their lawn to prevent zombies from reaching
 268 their house. Each of the plants has different attributes. For example, some plants
 269 (offensive ones) attack zombies directly, while other plants (defensive ones) slow
 270 down zombies to give the player more time to attack the zombies. A few plants
 271 generate “sun,” an in-game resource needed to utilize more plants. The challenge of
 272 the game comes from determining which plants to use and where to position them
 273 on the battlefield to defeat all the zombies in each level of the game.

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

289 After finalizing the problem-solving competency model, Shute and her team
 290 identified dozens of observable in-game indicators (after repeatedly playing the game
 291 and watching expert solutions on YouTube). The indicators are used as evidence to
 292 update the problem-solving CM (step 3; in this example step 4 was not needed). For
 293 example, the research team determined that planting three or more sun-producing
 294 plants (which provide the currency to use other plants) before the first wave of zombies
 295 arrive is an indicator of the “analyze givens and constraints” facet and shows that the
 296 player understands time and resource constraints. Table 5.1 includes some examples
 297 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	<ul style="list-style-type: none"> Plants >3 Sunflowers before the second wave of zombies arrives Selects plants off the conveyor belt before it becomes full
Plan a solution pathway	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> Shovels Sunflowers in the back and replaces them with offensive plants when the ratio of zombies to plants exceeds 2:1

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

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

334 Consider indicator #37 in Fig. 5.2 (use of iceberg lettuce in UYB). This indi-
 335 cator is connected to the “tool use” facet, and a player has just performed some
 336 action in the game which was judged as “poor” (e.g., placed an iceberg lettuce
 337 proximal to a fire-breathing plant, thus cancelling out the “freezing” effect of the
 338 lettuce). The real-time estimate that the learner is low on the “tool use” facet is
 339 $p = 0.61$ (for more details see Shute, Wang, et al., 2016b).

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

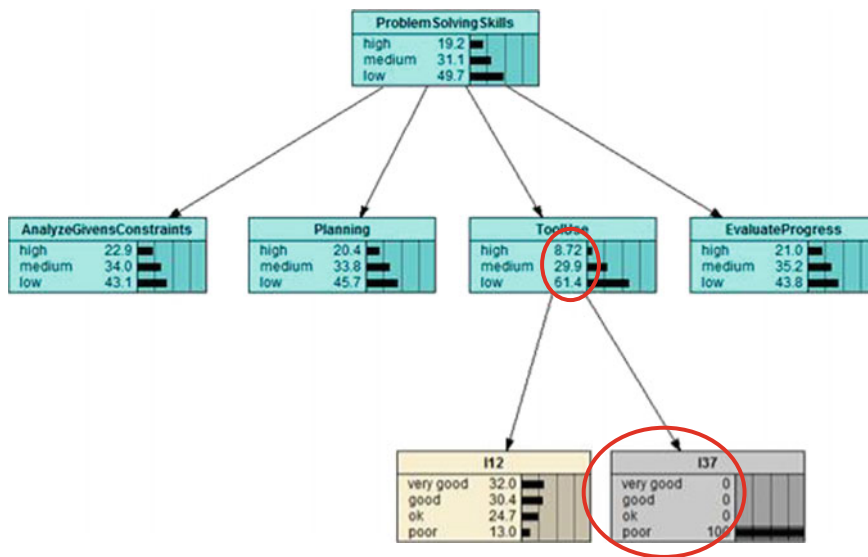


Fig. 5.2 An example of a BN with data for indicator #37 entered (poor use of iceberg lettuce)

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

351 Validation study participants were 7th grade students ($n = 55$) from a middle
 352 school in suburban Illinois. They students played the game for 3 h (1 h per day
 353 across three consecutive days). The results showed that the students’ scores from
 354 the two external tests significantly correlated with the in-game stealth assessment
 355 estimates [Raven’s ($r = 0.40$, $p < 0.01$) and MicroDYN ($r = 0.41$, $p < 0.01$)].
 356 Therefore, the stealth assessment embedded in the UYB game appears to be valid.
 357 Other studies have been conducted using stealth assessment to measure various
 358 competencies, e.g., physics understanding (Shute et al., 2013) and persistence
 359 (Ventura & Shute, 2013). The overall findings from these studies also show sig-
 360 nificant correlations between external and the in-game estimates. Finally, it’s
 361 important to note that this assessment approach, while illustrated in a 2D envi-
 362 ronment, can also be used in 3D games and environments (e.g., Portal 2 research,
 363 see Shute et al., 2015). We now discuss the next logical steps to take—making IEs
 364 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.



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

- 454 1. *Ensuring the quality of assessments.* U.S. schools are under local control, thus
455 students in a given state could engage in thousands of IEs during their educa-
456 tional tenure. Teachers, publishers, researchers, and others will be developing
457 IEs, but with no standards in place, they will inevitably differ in curricular
458 coverage, difficulty of the material, scenarios and formats used, and many other
459 ways that will affect the adequacy of the IE, tasks, and inferences on knowledge
460 and skill acquisition that can justifiably be made from successfully completing
461 the IEs. More research is needed to figure out how to equate IEs or create
462 common measurements from diverse environments. Towards that end, there
463 must be common models employed across different activities, curricula, and
464 contexts. Moreover, it is important to determine how to interpret evidence where
465 the activities may be the same but the contexts in which students are working
466 differ (e.g., working alone vs. working with another student).
- 467 2. *Making sense of different learning progressions.* IEs can provide a greater
468 variety of learning situations than traditional face-to-face classroom settings,
469 thus evidence for assessing and tracking learning progressions becomes more
470 complex rather than general across individual students. As a result, we need to
471 be able to model learning progressions in multiple aspects of student growth and
472 experiences, which can be applied across different learning activities and con-
473 texts. Moreover, there is not just one correct order of progression as learning in
474 IEs involves many interactions between individual students and situations,
475 which may be too complex for most measurement theories that assume linearity
476 and independence. So theories of learning progressions in IEs need to be
477 actively researched and validated to realize their potential.
- 478 3. *Privacy/Security.* This issue relates to the accumulation of student data from
479 disparate sources. However, information about individual students may be at
480 risk of being shared far more broadly than is justifiable. And because of the
481 often high-stakes consequences associated with tests, many parents and other
482 stakeholders fear that the data collected could later be used against the students.
483

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



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

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

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

525 References

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

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

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

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

- 535 Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest
536 Atlantis: A game without guns. *Educational Technology Research and Development*, 53(1),
537 86–107.
- 538 Barfield, W. (2015). *Fundamentals of wearable computers and augmented reality* (2nd ed.). New
539 York: CRC Press.
- 540 Bellini, H., Chen, W., Sugiyama, M., Shin, M., Alam, S., & Takayama, D. (2016). *Profiles in
541 innovation: Virtual & augmented reality*. Goldman Sachs group, inc.
- 542 Bransford, J., Brown, A., & Cocking, R. (2000). *Brain, mind, experience, and school*.
543 Washington, DC: National Academies Press.
- 544 Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom.
545 *TechTrends*, 60(5), 517–519. doi:10.1007/s11528-016-0102-z.
- 546 Clark, D., Tanner-Smith, E., & Killingsworth, S. (2014). Digital games, design, and learning: A
547 systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. doi:10.
548 3102/0034654315582065.
- 549 Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and
550 Row.
- 551 Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66–69.
552 doi:10.1126/science.1167311.
- 553 De Klerk, S., Eggen, T. J. H. M., & Veldkamp, B. P. (2015). Psychometric analysis of the
554 performance data of simulation-based assessment: A systematic review and a Bayesian
555 network example. *Computers & Education*, 85, 23–34.
- 556 Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting
557 youth, games, and learning* (pp. 21–40). Cambridge, MA: The MIT Press. doi:10.1162/dmal.
558 9780262693646.021.
- 559 Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in
560 K-12 and higher education settings: A review of the research. *British Journal of Educational
561 Technology*, 41(1), 33–55.
- 562 Kanar, A. M., & Bell, B. S. (2013). Guiding learners through technology-based instruction: The
563 effects of adaptive guidance design and individual differences on learning over time. *Journal of
564 Educational Psychology*, 105(4), 1067.
- 565 Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An
566 exploratory investigation in river city, a multi-user virtual environment. *Journal of Science
567 Education and Technology*, 16(1), 99–111.
- 568 Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: Protecting immersion in
569 didactically adaptive digital educational games. *Journal of Computer Assisted learning*, 26
570 (2), 95–105.
- 571 Klimmt, C., Hartmann, T., & Schramm, H. (2006). Effectance, self-efficacy, and the motivation to
572 play video games. In J. Bryant & P. Vorderer (Eds.), *Psychology of entertainment* (pp. 291–
573 313). Mahwah, NJ: Erlbaum.
- 574 Knight, J. F., Carley, S., Tregunna, B., Jarvis, S., Smithies, R., de Freitas, S., et al. (2010). Serious
575 gaming technology in major incident triage training: A pragmatic controlled trial.
576 *Resuscitation*, 81(9), 1175–1179.
- 577 Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge,
578 MA: University of Cambridge Press.
- 579 Liu, C., Cheng, Y., & Huang, C. (2011). The effect of simulation games on the learning of
580 computational problem solving. *Computers & Education*, 57(3), 1907–1918.
- 581 Lopes, R., & Bidarra, R. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE
582 Transactions on Computational Intelligence and AI in Games*, 3(2), 85–99.
- 583 Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus article: On the structure of
584 educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1),
585 3–62.
- 586 Piaget, J. (1973). *To understand is to invent: The future of education*. New York: Grossman.
- 587 Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical
588 Psychology*, 19(1), 137–150.

- 589 Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic
590 motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78.
- 591 Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer*
592 *Games and Instruction*, 55(2), 503–524.
- 593 Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing It–Or
594 can you? Evaluating an assessment for learning system called ACED. *International Journal of*
595 *Artificial Intelligence in Education*, 18(4), 289–316.
- 596 Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters &
597 H. van Oostendorp (Eds.), *Techniques to facilitate learning and motivation of serious games*
598 (pp. 59–78). New York, NY: Springer.
- 599 Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. (2016a). Advances in the science of
600 assessment. *Educational Assessment*, 21(1), 1–27.
- 601 Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of portal 2 and lumosity
602 on cognitive and noncognitive skills. *Computers & Education*, 80, 58–67.
- 603 Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in
604 newton's playground. *The Journal of Educational Research*, 106(6), 423–430.
- 605 Shute, V., Ventura, M., Kim, Y., & Wang, L. (2014). Video games and learning. In W. G. Tierney,
606 Z. B. Corwin, T. Fullerton, & G. Ragusa (Eds.), *Postsecondary play: The role of games and*
607 *social media in higher education* (pp. 217–235). Baltimore, MD: John Hopkins Press.
- 608 Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016b). Measuring problem solving
609 skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, 63,
610 106–117.
- 611 Tennyson, R. D., & Jorczak, R. L. (2008). A conceptual framework for the empirical study of
612 instructional games. In H. F. O'Neill & R. S. Perez (Eds.), *Computer games and team and*
613 *individual learning* (pp. 3–20). Boston: Elsevier.
- 614 Van Eck, R., & Hung, W. (2010). *A taxonomy and framework for designing educational games to*
615 *promote problem solving*. In paper presentation at the videogame cultures & the future of
616 interactive entertainment annual conference of the inter-disciplinary. *Net Group*, 227–263.
- 617 Van Oostendorp, H., Van der Spek, E., & Linssen, J. (2013). Adapting the complexity level of a
618 serious game to the proficiency of players. *EAI Endorsed Transactions on Serious Games*,
619 1(2), 8–15.
- 620 Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence.
621 *Computers in Human Behavior*, 29(6), 2568–2572.
- 622 Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*.
623 Cambridge, MA: Harvard University Press.
- 624 Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning?
625 *Intelligence*, 40(1), 1–14.

626 Author Biographies

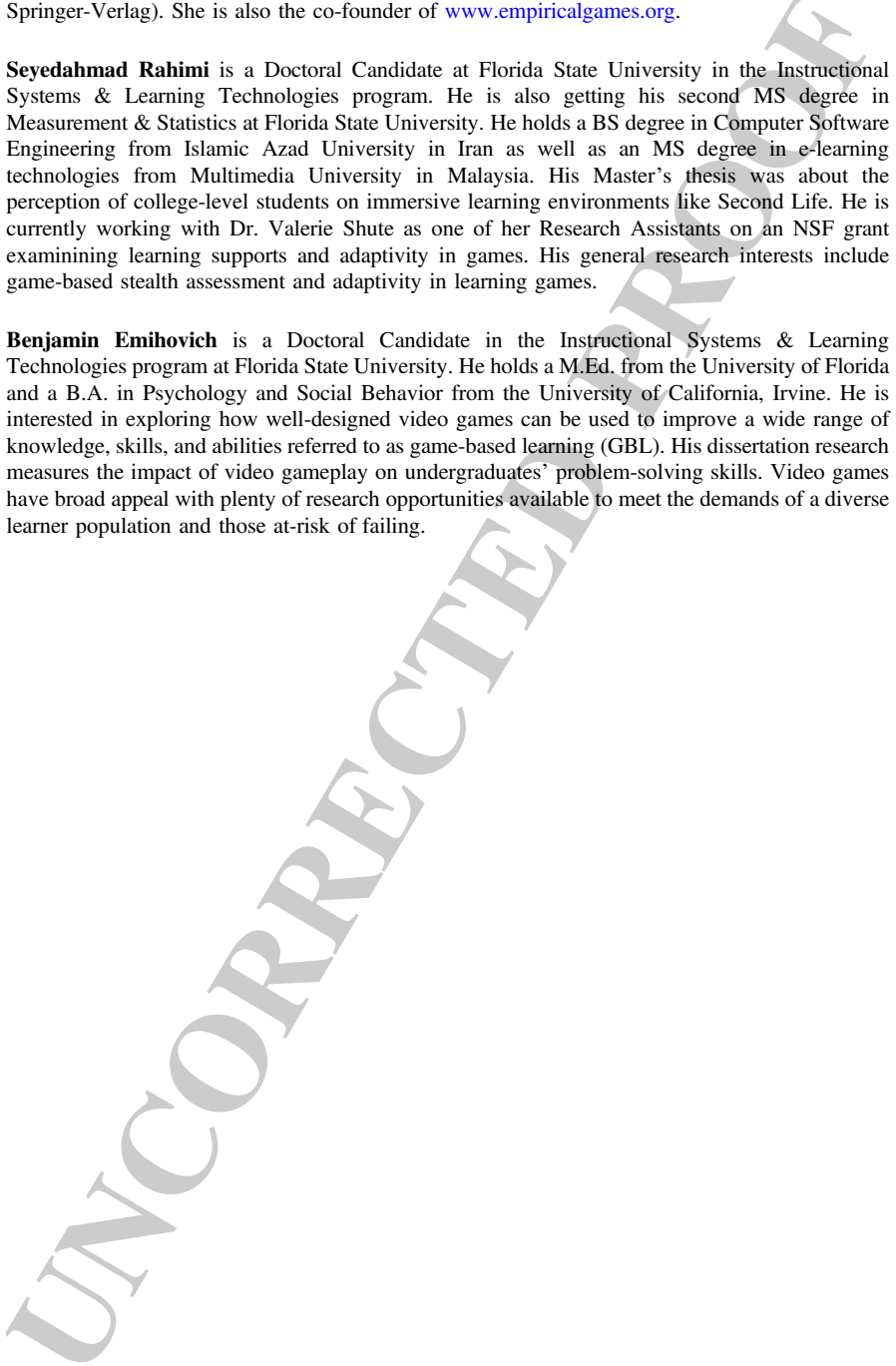
627 **Val Shute** is the Mack & Effie Campbell Tyner Endowed Professor in Education in the
628 Department of Educational Psychology and Learning Systems at Florida State University. Before
629 coming to FSU in 2007, she was a principal research scientist at Educational Testing Service
630 where she was involved with basic and applied research projects related to assessment, cognitive
631 diagnosis, and learning from advanced instructional systems. Her general research interests hover
632 around the design, development, and evaluation of advanced systems to support learning—
633 particularly related to twenty first century competencies. Her current research involves using
634 games with stealth assessment to support learning—of cognitive and noncognitive knowledge,
635 skills, and dispositions. Her research has resulted in numerous grants, journal articles, books,
636 chapters in edited books, a patent, and a couple of recent books (e.g., Shute & Ventura, 2013,



637 Measuring and supporting learning in games: Stealth assessment, The MIT Press; and Shute &
638 Becker, 2010, Innovative assessment for the twenty first century: Supporting educational needs,
639 Springer-Verlag). She is also the co-founder of www.empiricalgames.org.

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

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



658

Author Query Form

659

Book ID : 439860_1_En

661

662

Chapter No : 5

660

**Springer**

the language of science

663

Please ensure you fill out your response to the queries raised below and return this form along with your corrections.

664

665

Dear Author,

666

During the process of typesetting your chapter, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

667

668

669

674

673

676

678

679

682

683

Query Refs.	Details Required	Author's Response
AQ1	Kindly note that there is mismatch between 'author biography' and 'affiliation' provided, we have followed biography. Please check and amend in necessary.	
AQ2	The sentence 'Determine which ...' seems to be incomplete. Please check for missing words/phrases and complete the sentence.	

UNCORRECTED PROOF

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

<i>Instruction to printer</i>	<i>Textual mark</i>	<i>Marginal mark</i>
Leave unchanged	... under matter to remain	Ⓟ
Insert in text the matter indicated in the margin	∧	New matter followed by ∧ or ∧ [Ⓢ]
Delete	/ through single character, rule or underline or ┌───┐ through all characters to be deleted	Ⓞ or Ⓞ [Ⓢ]
Substitute character or substitute part of one or more word(s)	/ through letter or ┌───┐ through characters	new character / or new characters /
Change to italics	— under matter to be changed	↵
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	≡ under matter to be changed	≡
Change to bold type	~ under matter to be changed	~
Change to bold italic	≈ under matter to be changed	≈
Change to lower case	Encircle matter to be changed	≡
Change italic to upright type	(As above)	⊕
Change bold to non-bold type	(As above)	⊖
Insert 'superior' character	/ through character or ∧ where required	Υ or Υ under character e.g. Υ or Υ
Insert 'inferior' character	(As above)	∧ over character e.g. ∧
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	ʹ or ʸ and/or ʹ or ʸ
Insert double quotation marks	(As above)	“ or ” and/or ” or ”
Insert hyphen	(As above)	⊞
Start new paragraph	┌	┌
No new paragraph	┐	┐
Transpose	└┘	└┘
Close up	linking ○ characters	○
Insert or substitute space between characters or words	/ through character or ∧ where required	Υ
Reduce space between characters or words		↑