The potentials of game-based environments for integrated, immersive learning data

Jennifer S. Groff

MIT Media Lab & MIT Education Arcade, Cambridge, Massachusetts

Correspondence
Jennifer S. Groff, MIT Media Lab & MIT Education Arcade, 77 Massachusetts Avenue, NE48-328, Cambridge, MA 02139.
Email: jgroff@media.mit.edu

Abstract
In the last two decades, game-based learning environments have evolved into powerful learning tools. With the growth and evolution in complexity of these innovations, complementary work in game-based assessment design has also begun to take shape and has generated considerable interest among a variety of education stakeholders. This is in part due to their ability to offer real-time, adaptive and integrated formative feedback. As the use of digital devices and digital learning tools continues to expand in schools, the potential for harnessing learning data becomes increasingly possible and powerful. Yet there are inherent tensions and challenges in the nature of collecting assessment data inside a game-based learning environment which confront our foundational premises of what makes for a good assessment. In this article, we explore the evolution and opportunities of the emerging field of game-based assessment, the challenges and tensions these innovations present and how we may be able to collectively advance this work to benefit everyday classrooms.

KEYWORDS
assessment, digital games, educational technology, game-based assessments, learning games

1 | INTRODUCTION

The role of digital games in learning and education has evolved remarkably in the last two decades. Once a genre of learning tools that had to overcome the 'games' stigma, learning games are now a common learning technology both in the classroom and at home and the depth and breadth of their formats go well beyond The Oregon Trail, Math Blaster and the thousands of literacy and maths applications available for download (Richards, Stebbins, & Moellering, 2013). Many of the original learning games were criticised for their poor design and were dubbed 'edutainment' because they only attempted to leverage a behaviourist model of learning: complete the task and receive a reward. However, considerable advances have been made in learning theory and in our understanding of how we learn that have helped to transform the design of learning games into more robust, constructivist environments (Klopfer, Osterweil, & Salen, 2009). This has brought us into more modern learning theories, including constructivism, social constructivism, and constructionism, all placing the emphasis on the nature of learners constructing their understanding through their own
direct and social experience (Klopfer et al., 2009). This picture of learning can map quite well onto games which, when well-designed, also display more nuanced and complex dynamics for the user to engage with. This is why we now see such incredible engagement and excitement about the potential of learning games and such a broad range of powerful examples of how game-based environments can be used to facilitate deeper learning and assessment. This includes the development of large-scale, immersive game worlds which facilitate deep learning across a variety of domains, situated in rich storyline narratives where learners collaborate on a variety of problems and quests. Some embody many of the best parts of learning environment design—at times, even better than traditional, real world learning environments and classrooms where learning is situated in context, with individual and collaborative problem-solving and real-time data collection and where feedback guides the learners towards deeper, more meaningful learning (Shute & Ventura, 2013).

However, traditional assessment instruments and approaches that still largely serve as standardised test formats in schools have not followed this type of evolution. Many current approaches to standardised assessment, such as traditional, summative exams, require qualities of an assessment that still match with linearity and are associated with behaviourism—immediate feedback (or cause/effect), direct correlation, uniformity, and so on (Strickland & Strickland, 1998). Many of these assessments are often more narrow in scope and, as such, offer decontextualised, linear questions that are often incapable of assessing more complex skills and competencies. Since the 1990s, we have seen greater use of multiple-choice tests that are less likely to measure complex learning (National Research Council, 2010). In order to escape this mismatch and take advantage of the opportunities offered by game-based learning environments we must be willing to rethink the (very rigid) frames that have created our approach to assessment so far. It is in part for this reason that one of the most exciting recent advances in the field of learning games is how these environments can be used to collect critical learning data. Game-based Assessments (GBAs) offer immersive experiences in a situated context, collect a range of learning data and respond in real time for more authentic and meaningful feedback. In the past few years, there has been a range of examples of what these technologies can do—from understanding learners’ misconceptions in a key area of science to capturing their capacity for systems thinking, whilst guiding them adaptively on what they may need next. Although developing and implementing GBAs come with their own challenges as their development expands, we create the potential for a richly integrated learning environment system where learning data are collected and used in an ongoing fashion as assessment of (summative), for (formative) and as learning. Most importantly, such a rich data display could remove the need for any separate, large-scale summative tests and all the negatives that come with them (Phillips & Popović, 2012). In short, a new generation of educational games has the potential to liberate teachers from the need to administer standalone, standardised tests.

2 OVERVIEW OF THE FIELD

Digital learning games come in a wide range of styles and formats and address many topics. These include: (a) targeted games, which are individual apps that target one skill or aspect of a domain (such as phonemic awareness); (b) linear games, which include a storyline and a series of puzzles that explore the general mechanics of a concept (such as algebraic thinking); (c) open-ended or sandbox game, which offers players tools and context to construct items or outcomes in a game; and (d) virtual worlds, which engage players in problems and quests that often integrate a range of topics (such as problem-based quests that require the learner/player to use knowledge in biology, maths, communication, data analysis, etc.). It is also worth noting that there are many COTS (Commercial Off-The-Shelf) games which were built for entertainment but have been used as instructional tools in the classroom. Figure 1 provides a sampling of the types of learning games and examples from education and commercial spaces.

Each game genre comes with affordances and limitations. Targeted games can be integrated in the classroom because they more easily align with curricula and fit into class timescales. Linear games can be short or long, but often provide rich experiences for exploring a concept before further instruction is encountered in the classroom. Open-ended/sandbox games are gaining in popularity. For example, Civilization V is a commercial game that leads the player through the growth of a civilization and empire and there are many examples of its use in the classroom to target numerous learning goals, including trade routes and ethical thinking (Fitzgerald & Groff, 2010; McCall, 2011).
Virtual worlds that are on-going, such as popular multiplayer online games like *World of Warcraft* (WoW), are often the most challenging to integrate in the classroom, but can offer some of the richest learning environments (McCall, 2011).

In the last decade, a growing number of teachers has demonstrated outstanding and powerful learning experiences by leveraging digital games, demonstrating students’ deeper learning and engagement and that games can have a place in the classroom. In a recent survey conducted by the Joan Ganz Cooney Center, 32% of the United States pre-school to year 12 educators reported that they used digital games 2 to 4 times per week in their classrooms, with 18% using them every day (Millstone, 2012).

There has been considerable research in support of game-based learning and its effectiveness. For example, we have a cursory understanding of the valuable skill development that playing games can support. According to a review by McFarlane, Sparrowhawk, and Heald (2002), these include:

- Strategic thinking
- Planning
- Communication
- Application of numbers
- Negotiating skills

<table>
<thead>
<tr>
<th>Game Genre</th>
<th>EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Targeted Games</strong></td>
<td></td>
</tr>
<tr>
<td>individual apps that</td>
<td><em>Supercharged!</em></td>
</tr>
<tr>
<td>target one skill or</td>
<td><em>Angry Birds</em></td>
</tr>
<tr>
<td>aspect of a domain.</td>
<td></td>
</tr>
<tr>
<td>These environments</td>
<td></td>
</tr>
<tr>
<td>typically best</td>
<td></td>
</tr>
<tr>
<td>support the learning</td>
<td></td>
</tr>
<tr>
<td>and assessment of</td>
<td></td>
</tr>
<tr>
<td>discrete facts,</td>
<td></td>
</tr>
<tr>
<td>skills, and</td>
<td></td>
</tr>
<tr>
<td>knowledge, such as</td>
<td></td>
</tr>
<tr>
<td>“adding fractions”.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Linear Games</strong></td>
<td></td>
</tr>
<tr>
<td>A storyline and a</td>
<td>*Lure of the</td>
</tr>
<tr>
<td>series of puzzles</td>
<td><em>Ninja Garden</em></td>
</tr>
<tr>
<td>that explore the</td>
<td></td>
</tr>
<tr>
<td>general mechanics of</td>
<td></td>
</tr>
<tr>
<td>a concept.</td>
<td></td>
</tr>
<tr>
<td>These environments</td>
<td></td>
</tr>
<tr>
<td>typically best</td>
<td></td>
</tr>
<tr>
<td>support the learning</td>
<td></td>
</tr>
<tr>
<td>and assessment of</td>
<td></td>
</tr>
<tr>
<td>conceptual foundations</td>
<td></td>
</tr>
<tr>
<td>of skills, such as</td>
<td></td>
</tr>
<tr>
<td>“algebraic thinking”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Open-ended / Sandbox Games</strong></td>
<td></td>
</tr>
<tr>
<td>Players use tools and</td>
<td><em>Minecraft</em></td>
</tr>
<tr>
<td>context to construct</td>
<td><em>Civilization</em></td>
</tr>
<tr>
<td>items or outcomes.</td>
<td><em>SimCity</em></td>
</tr>
<tr>
<td>These environments</td>
<td></td>
</tr>
<tr>
<td>typically best</td>
<td></td>
</tr>
<tr>
<td>support a wide</td>
<td></td>
</tr>
<tr>
<td>variety of skills</td>
<td></td>
</tr>
<tr>
<td>and competencies,</td>
<td></td>
</tr>
<tr>
<td>depending on the</td>
<td></td>
</tr>
<tr>
<td>design and</td>
<td></td>
</tr>
<tr>
<td>application of the</td>
<td></td>
</tr>
<tr>
<td>game—including</td>
<td></td>
</tr>
<tr>
<td>“systems thinking”</td>
<td></td>
</tr>
<tr>
<td>and “collaboration”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Virtual Game Worlds</strong></td>
<td></td>
</tr>
<tr>
<td>Players engage in</td>
<td><em>Quest Atlantis</em></td>
</tr>
<tr>
<td>problems and quests</td>
<td>*World of</td>
</tr>
<tr>
<td>that often</td>
<td><em>Warcraft</em></td>
</tr>
<tr>
<td>integrate a range of</td>
<td></td>
</tr>
<tr>
<td>topics.</td>
<td></td>
</tr>
<tr>
<td>These environments</td>
<td></td>
</tr>
<tr>
<td>typically best</td>
<td></td>
</tr>
<tr>
<td>support a wide</td>
<td></td>
</tr>
<tr>
<td>variety of skills</td>
<td></td>
</tr>
<tr>
<td>and competencies,</td>
<td></td>
</tr>
<tr>
<td>depending on the</td>
<td></td>
</tr>
<tr>
<td>design and</td>
<td></td>
</tr>
<tr>
<td>application of the</td>
<td></td>
</tr>
<tr>
<td>game—including</td>
<td></td>
</tr>
<tr>
<td>“critical thinking”</td>
<td></td>
</tr>
<tr>
<td>and “collaborative</td>
<td></td>
</tr>
<tr>
<td>problem-solving”.</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 1** Taxonomy of game genres (adapted from Groff, McCall, Darvasi, & Gilbert, 2016; Squire, 2008). [Colour figure can be viewed at wileyonlinelibrary.com]
• Group decision-making
• Data-handling

A recent meta-analysis of over 77 studies that met review criteria found that digital games could enhance student learning, as measured by cognitive competencies that are relative to traditional instructional approaches (Clark, Tanner-Smith, & Killingsworth, 2013). These results support the sentiments of many educators. According to the Joan Ganz Cooney Center's 2012 Teacher Survey, more than 60% of teachers reported that these games helped them to personalise instruction, better assess knowledge, collect helpful data on lower-performing students and promote deeper engagement and collaboration among all students (Millstone, 2012).

2.1 | Game-based learning and game-based assessment

As the work in designing game-based learning environments has evolved, so has the need to understand what learners are doing in these environments. Of course, such questions are generally understood to be critical in designing a learning experience; but as game-based learning environments evolved in their scope and complexity, gathering such learning outcome data became imperative. Moreover, many saw that these environments presented a tremendous opportunity to collect critical data in a variety of dimensions—particularly elements that, traditionally, we have not been able to assess well, such as ‘collaborative problem-solving’ and ‘systems thinking’.

In the last few years, the interest and effort in using games as assessment engines have grown tremendously, with numerous examples and approaches emerging from a variety of groups, including traditional assessment organisations such as the U.S. CRESST (the National Center for Research on Evaluation, Standards, and Student Testing), ETS² (the Educational Testing Service) and GlassLab (The Games, Learning and Assessment Lab (http://glasslabgames.org)), an organisation specifically established for the study and development of GBAs.

Initial research and design efforts have produced a variety of examples, including the redesign of the commercial game Crayon Physics to become Newton's Playground, a game that is able to assess a learner's abilities in both physics and systems thinking; the retrofitting of assessment into the commercial game SimCity, creating SimCityEDU which assesses science, language arts, critical thinking and systems modelling; and the work of Groff et al. at the Learning Games Network that explores how large-scale standardised tests' elements could use game-based tasks to innovate on how these tests are administered. Each of these approaches has demonstrated varying levels of success and challenges, due in part to the inherent tensions between learning game design and summative assessment design. To be sure, game-based learning environments and game-based assessments are not necessarily the same thing. There are game-based learning tools that do not collect data and are not necessarily adaptive or responsive to the learner’s needs. There are also game-like environments that are designed to be assessments, with little to no learning occurring in the game. However, it is generally the sentiment of designers in this space that quality learning tools for ongoing classroom use are found in the overlap between these two, supporting assessment for learning rather than just assessment of learning (see Figure 2) (Groff, Clarke-Midura, Owen, Rosenheck, & Beall, 2015).

Game-based learning environments offer a number of opportunities and advantages in a variety of dimensions. First, digital games offer a rich medium that is now pervasive in many contexts. In the United States, over 72% of all teens play video games on a computer, game console or portable device like a cellphone (Lenhart, A. & Pew Research Center, 2015). Second, they allow us to situate learning in real-world contexts, helping learners to understand why and/or how a concept is relevant and applied in real life. Third, since games can offer rich dynamic environments, learners are engaged with and must apply a range of competencies that would otherwise be very difficult to approach in a traditional testing format.

2.2 | Designing game-based assessments

It is worth highlighting that one of the core advancements in assessment methodology that has emerged in the last 15 years was the foundation for many of the innovations in game-based assessment. Evidence-Centered Design (ECD) is
a principled framework for designing, producing and delivering educational assessments (Mislevy, Almond, & Lukas, 2003). Its heart is a construct-centred approach to designing assessments which begins by ‘asking what complex of knowledge, skills, or other attribute should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviours or performances should reveal those constructs and what tasks or situations should elicit those behaviours? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics’ (Messick, 1994). Although constructing a valid assessment is very complex, at a general level, ECD is the coherent integration of three main components called the Conceptual Assessment Framework, or CAF (Figure 3) (Mislevy et al., 2003, p. 6-):

- **Student Model—What Are We Measuring?** A Student Model defines one or more variables related to the knowledge, skills and abilities we wish to measure. A simple student model characterises students in terms of the proportion of a domain of tasks they are likely to answer correctly.

- **Evidence Model—How Do We Measure It?** The Evidence Model defines observable variables and indicators of learners’ performance to be interpreted for the evaluation of their level of proficiency or competency development.
• Task Model—Where Do We Measure It? The Task Model describes how to structure the kinds of situations that we need in order to obtain the kinds of evidence for the evidence models. It describes the presentation material that is to be provided to the examinee and the work products generated in response. It also contains variables that describe features of tasks, as well as how they are related to the presentation material and work products.

Learning game designers were already fairly versed in designing tasks (game mechanics) and to some extent in the Student Model that was the focus of the game. The complete ECD framework offered the rigour and coherence to help to deepen and enrich the design of learning games in order to be robust learning environments that more accurately modelled learning data. The ECD framework (and its derivations discussed below) has served as the foundation for the design of many of the innovations in GBAs.

2.3 | Stealth assessment

Shute and colleagues at the Florida State University have made considerable contributions to the field of GBA through their work on designing and researching game-based stealth assessments (Shute, 2011; Shute & Ventura, 2013). These assessments use the ECD framework while designing strategic data collection into the game mechanics, thereby overcoming one of the greatest challenges of GBAs—keeping the playful engagement of the game environment that can validly and reliably measure learning in the game without disrupting engagement and then leveraging that information to bolster learning (Shute, Levy, Baker, Zapata, & Beck, 2009). In this model, the stealth assessment serves as a complement by determining specific gameplay behaviours that can act as evidence of a claim (specified in the evidence model) and link them to the competency model (Shute & Ventura, 2013; Shute, Ventura, et al., 2009). The assessment tasks embedded in the game are clearly linked to claims about personal competencies and the estimates of competency levels can be used diagnostically and formatively to provide adaptively selected game levels, targeted feedback and other forms of learning support to students as they continue to engage in gameplay. Moreover, through the evidentiary models they use to build the stealth assessments, they demonstrate that these GBAs are statistically valid assessments (Shute & Ventura, 2013).

3 | OPPORTUNITIES AND ADVANTAGES OF GAME-BASED ASSESSMENT

When data collection is deeply integrated and used in the game design, we are able to garner additional benefits and advantages. To understand these opportunities, it is helpful to first look briefly at how current assessment infrastructures fall short.

Formative assessment—feedback on learning in a short time cycle in order to directly impact next steps for the learner—is critical to the learning cycle and yet is often not well integrated in everyday classroom practice (Sadler, 1998). Summative assessment—often delivered in long cycles (such as 1–2 times per year) and out of context of the learning (such as standardized tests)—is removed from context and often disrupts the classroom schedule. Multiple-choice and other fixed-response formats emphasise basic content knowledge and skills within subjects, resulting in a substantial narrowing of the curriculum and pedagogies from broader skills such as systems thinking and methods such as project-based learning (Kellaghan & Madaus, 1991; Shepard, 1991). Moreover, new 21st century skills and competencies such as persistence, creativity and collaborative problem-solving—which are included in the European Reference Framework of Key Competences for Lifelong Learning (OJEU, 2006)—have not only been identified as critical to thriving in our new digital society and knowledge economy, but can substantially impact student academic achievement (O’Connor & Paunonen, 2007; Poropat, 2009; Sternberg, 2006). Yet these are rarely, if ever, taught or assessed in learning environments. What is needed are new performance-based assessments that assess how students use knowledge and skills that are directly relevant for use in the real world—something that digital game-based learning environments are particularly able to support (Shute & Ventura, 2013). In short, assessment should be ongoing and invisible to students, supporting real-time or just-in-time instruction while adding that extra layer of engagement that digital games do so well (Shute, Levy, Baker, Zapata, & Beck, 2009).

Embedding assessments within games provides a way to monitor a player’s current level in valued competencies and then use that information as the basis for support, such as adjusting the difficulty level of challenges or providing...
timely feedback (Shute & Ventura, 2013). This type of assessment has been termed ‘stealth assessment’, where learner measurement is deeply embedded in games to unobtrusively, accurately and dynamically measure how players are progressing in relation to targeted competencies (Shute, 2011; Shute, Levy, et al., 2009).

In many ways, games and assessments are a natural marriage because it has been argued that all games are essentially assessments (Gee, 2003). In a typical digital game, as players interact with the environment, the values of different game-specific variables change; and solving major problems in games permits players to gain rank or ‘level up’ based on their performance (Shute & Ventura, 2013). Good learning game design is found in the alignment of good game mechanics that playfully support the exploration of the targeted concepts; similarly, in a good game-based assessment, the assessment goals and the game mechanics work in unison (Groff et al., 2015). Learning games and GBAs simply and specifically monitor educationally-relevant variables at different levels of granularity in games. By providing goals and challenges, games keep players at the edge of their capabilities, capitalising on deep principles of learning (Gee, 2003).

Well-designed games can serve as next-generation assessments that engage students in a seamless learning experience, assessing their learning without their even realising it, yet at the same time providing critical feedback to more effectively support their learning pathway. This helps teachers to personalise learning, instill conceptual understanding and knowledge transfer and motivate students to develop the persistence they need to achieve mastery (Phillips & Popović, 2012). Yet, perhaps most critically, digital games allow us to capture streams of data which can track both the learners’ process and their final outputs. This helps to better understand how they have arrived at an answer, rather than just scoring the answer as right or wrong. This is valuable information for the educator and useful feedback for the learner and, more generally, for learning scientists (DiCerbo, 2014).

Finally, digital games play a key role in the emerging ‘digital ocean’ of learning data—the intersection of many streams of data coming from the range of digital tools and technologies now available in the classroom and beyond. According to DiCerbo and Behrens (2014), when combined with appropriate analysis and standards for use, this digital ocean of data opens the door to new types of naturalistic observation and inference that could help to more deeply understand learning and improve the design of learning experiences and environments. The interest and potential for this are demonstrated in the recent upsurge in two related fields: learning analytics (LA) and educational data mining (EDM). LA is the use of data, analysis and predictive modelling to improve teaching and learning. Analytical tools, such as models and algorithms, process the data and seek to find reproducible, actionable findings that support at-risk students or programme recommendations that are better matched to a student’s aptitudes. EDM is a related, emerging discipline that is concerned with developing methods to explore the unique and increasingly large-scale data that come from educational settings and using these methods to better understand students and the settings in which they learn (http://educationaldatamining.org).

The main goal of both EDM and LA is to extract information from educational data to support education-related decision-making. From a general perspective, it can be argued that EDM focuses on techniques and methodologies, whilst LA deals more with applications (Lián & Pérez, 2015). Both fields are still in the very early stages and, although they are gradually becoming more sophisticated, it will take considerable research and development to improve the models that allow us to achieve these kinds of outcomes. As the potentiality of this work evolves, we can imagine a future where learners have seamless, integrated experiences across domains and their educational experience and where data are constantly collected, modelled and fed back to the learners in order to personalise their learning experience, not just in game-based environments, but across all their learning experiences, from science labs, to writing tasks, to maths practice, whilst never interrupting their experience to implement a formal test. Understanding how to assess and design formative experiences in game-based environments is a critical step in the pursuit of this reality.

3.1 | Examples of game-based assessments

Example 1: Newton’s Playground

Adapted from Shute and Ventura (2013, pp. 33–54)
Newton’s Playground is a stealth assessment that uses a digital game in simulating 2-D physics concepts, including gravity, mass, kinetic energy and transfer of momentum. It was developed with financial support from the Bill and Melinda Gates Foundation. Focusing on the development and assessment of three competencies—creativity, conscientiousness and conceptual physics understanding—, it is situated in a playful, game-based physics environment. After initially designing stealth assessments for these competencies in the game Crayon Physics Deluxe, developed by Petri Purho, Newton’s Playground was built from scratch in order to mitigate problems in the original game environment.

The objective of each problem in Newton’s Playground is to guide a green ball from a predetermined starting point to a red balloon (or balloons) which pops on contact and gives the student a trophy for the successfully completed level (and multiple trophies for multiple solutions). Everything obeys the basic rules of physics relating to gravity and Sir Isaac Newton’s three laws of motion. The player can nudge the ball to the left and right (if the surface is flat), but the primary way to move the ball is by drawing physical objects on the screen that ‘come to life’ once the object is drawn. The speed of (and importantly, the impulse delivered by) the swinging golf club is dependent on its size/mass distribution and the angle from which it was dropped to swing. The ball will then fly at a certain speed, length and trajectory. If drawn properly, it will hit the balloon. The various problems in Newton’s Playground require the player to create and use ramps, pendulums, levers, and so forth to move the ball. All solutions are drawn with coloured markers using the mouse. In a number of cases, the ball must go over a pit. If it falls into the pit, the player must start the problem over again.

Based on foundational conceptual physics (Feynman, 1964; Feynman, Leighton, & Sands, 1964; Hewitt, 2009), we interpret competency to involve the conceptual understanding of: (a) Newton’s three laws of motion; (b) potential and kinetic energy; and (c) conservation of angular momentum or torque. With all these competencies, Newton’s Playground task problems require the player to use one or more agents of force and motion in the solution. Successful solutions therefore inform one or more of the competencies to be developed in the student. Figure 4 shows the ballistic pendulum problem. An embedded code within the game uses relevant gameplay data to automatically identify agents and generate evidence indicators.

Example 2: The Radix Endeavor
Adapted from Groff et al. (2015)

The Radix Endeavor (Figure 5) is a massively multiplayer online game (MMOG) developed by The Education Arcade at the Massachusetts Institute of Technology to improve learning and interest in STEM in high school students. It focuses on statistics, algebra, geometry, ecology, evolution, genetics and human body systems. Players take on the role of mathematicians and scientists and embark on quests that encourage them to explore and interact with the virtual world through maths and science. Embedded in a narrative in the world, players encounter a villain who does not believe in the practices of science. By taking on quests in the game, they must reason about science issues that are applicable to the game characters’ everyday lives, refute the unscientific claims of the villain and make choices based on what they consider to be valid evidence. There are seven quest lines in biology and six in mathematics, each of which contains a series of sub-quests around a specific content area (i.e., human body systems, statistics, geometry, etc.). The quests become progressively

**FIGURE 4** Newton’s Playground, Ballistic pendulum problem (left) and solution (right)
harder, building off what was learned in the previous quest. The goal is to have a culminating quest activity at the end of each quest line where students apply what they have learned about a problem or new situation.

As with many other game-based assessments, Radix uses a modified version of the ECD framework called Experiment Centered Design or XCD (Conrad, Clarke-Midura, & Klopfer, 2014) (Figure 6). It serves as a guide for quest design to fit the project in a way that ensures how evidence is gathered and interpreted in game play and that is consistent with the learning objectives the game is attempting to teach. Each quest is designed around one or more experiments that students complete in a series of steps. For example, they may be asked to grow a particular type of flower or build a scale map of a city. These experiments have inputs (the flower or object) with particular properties. The outputs of the experiments (e.g., the flower or object that is bred) also have properties. These data are captured and stored on the back end of the platform and can be evaluated in real time to see what kind of experiments students are conducting—are they breeding the right objects? Are they producing the right proportion of offspring? Thus, in the quests, the experiments allow the Radix team to assess students’ understanding of the content. This information can then be used formatively to provide feedback to both students and teachers as a way to ensure that students learn the content. It can also be used to assign ‘side quests’ in particular content areas where students are struggling. In the culminating quests, which are essentially a performance assessment, the team uses the data to summatively assess what a student knows after working through the quest chain. Some of the variables being measured in these quests will also be evaluated with external measures. This modified XCD framework allowed the team to account for the use of experiments in game play and, perhaps more critically, make the process accessible to the range of experts involved (i.e., designers, content experts, programmers).

4  TENSIONS AND CHALLENGES WITH GAME-BASED ASSESSMENT

GBAs have a great deal to offer but, together with these incredible advantages and opportunities, there are both design and implementation challenges in realising their potential. Foremost, there are the tensions in design. Despite the aforementioned overlaps and synergies with the design of game environments and assessment instruments, paradoxically, there are also aspects that are diametrically opposed. Good learning games are fluid, dynamic systems that take time to learn and are subsequently able to focus on a few core competencies. This makes them effective for
learning but not for assessment (DiCerbo, 2016). In other words, they support deeper learning, but it takes some time and game play before understanding a student’s ability in just a couple of constructs. This is in stark contrast to formal, summative assessments which are administered in constrained time periods and seek to capture student performance in a broad range of constructs. Perhaps even more problematically, the psychometric principles used to develop assessments, such as reliability and validity, demand rigid constraints in design decisions for the GBA. These constraints are used in high-stakes summative assessments in an effort to ensure fairness and to be sure that you are accurately evaluating what you claim to be evaluating. Assessment expert and co-developer of the ECD framework Robert Mislevy explains (ETS, 2016):

> We do know that variations in kids’ performances depend in part on their skills in exploring, doing math and figuring out strategies. But they can also depend on other things that shouldn’t play a big role for the results, like their familiarity with the game setting, the language and the representations that are used, and the familiarity with the cultural aspects such as the marketplace setting. The results also depend on how kids react to the game mechanics, the open-endedness, the storyline and additional challenges of the game itself. The same features that make the activity a more valid assessment for some kids as it engages them can make it worse for another kid who hates it.

The tension that Mislevy is highlighting is that we know that learners vary tremendously in motivation/interests, background, modalities of learning, etc., but we do not yet have a way to easily collect and model data for each learner around these dimensions and offer appropriate assessments that align with all these variables. So where one game might meet a learner’s interests and prior knowledge/experiences, it may be completely irrelevant for another. Traditional assessment methodologies purposefully try to mitigate any variances in an effort to standardise tests. This tension was highlighted in a 2013 project at the Learning Games Network (LGN), a non-profit spin-off from the MIT Education Arcade research group in partnership with a national testing agency which oversees a national summative assessment that monitors student performance in a range of domains. The aim of the project was to support expanded
research and development in the assessment space and explore the boundaries of what a national, summative exam like this could evolve into. LGN was charged with developing 12 sample assessment task scenarios that embodied game-based principles. The 12 samples displayed a range of game genres and mechanics, including role-playing games (RPGs), puzzles and turn-based strategy games. Yet despite this variety, the general consensus from the testing agency was that all these formats and scenarios were too far a departure from the design principles mandated for summative assessments. Although we are starting to see a variety of digital tasks in large-scale assessments, this example embodies the difficult tension between traditional assessment design and the dynamic nature of game-based environments. This tension does not represent the end-of-the-line for game-based assessments, but rather the call for further research on how to develop these complex environments and the modelling of learning data within them. The good news is that research is underway on modifications of existing conceptualisations and methods that help to broaden these design constraints (DiCerbo, 2016). At the same time, the evolution of the digital ocean of data and ubiquitous learning technology may reduce the need for such summative assessments.

There are additional challenges in the implementation and scale of GBAs. First, they represent and demand a team of specialised designers. Although learning games are no longer a niche genre, they are still a distinct domain across the spectrum of instructional technologies. But add to that the nuanced nature of assessment design and the specialised expertise this demands and we see that building high-quality GBAs requires teams with unique skills. Individual designers and the teams of experts who build these tools need to be supported in understanding how to do this work and continue to forge ahead in innovation in this space.

Second, although we have a growing field of GBAs, considerably more research and development are needed to understand this space, not only to address some of the aforementioned design tensions and challenges, but also to better understand how these environments affect learning in the long-term and how to truly assess more complex skills that are not currently easily assessed in any other way. If the field of education values the opportunities presented by these unique tools, then more resources are needed to invest in the research and development required to move the field forward towards large-scale implementation. Cost is often raised as a concern with these technologies. It should be noted that, although there is a significant cost to the development of these types of game-based assessment environments, more traditional large-scale standardised assessments also require significant budgets.

Finally, the gap from research to practice is a long-standing challenge in education (Groff & Mouza, 2008). It can often take years for an innovation to make its way into classrooms, let alone at scale, if it ever makes the transition at all. Many schools still do not have the basic technical infrastructure to support game-based environments, yet hopefully this will continue to improve over time. Smaller, more focused games and apps often fit into classroom practice more easily because they often align with the focus and pace of many standards-based classrooms and demand less curricular reconfiguring than a large-scale, virtual world GBA (Groff & Mouza, 2008). GBA approaches such as the Playful Assessments project which offers small, focused ‘game-like’ assessments around a specific construct in order to capture the current level of competency and conceptual level of a learner may offer one way in which GBAs can more easily integrate in classrooms in the near term (Groff, J. (n.d.) Playful assessments: Capturing competency level through focused playful task environments, unpublished manuscript).

As with most learning technologies, teachers are often the key piece of this puzzle, but many report the lack of sufficient training and support as one of the biggest barriers to implementing learning technologies (Phillips & Popović, 2012). Games still face an acute challenge in this regard: in a 2014 survey by the Joan Ganz Cooney Center, although nearly three-quarters (74%) of primary teachers surveyed reported using digital games for instruction, only 12% said they provided data for assessment (Takeuchi & Vaala, 2014). This indicates that understanding the evolving capacity of the learning tools and how to effectively use them in the classroom still need considerable support.

5 | CONCLUSIONS AND RECOMMENDATIONS

GBAs and game-based learning environments that use learner data to provide a more personalised, adaptive experiences whilst simultaneously collecting assessment and competency data offer powerful tools that can impact learners’
overall educational experiences, but also the potential to redefine the educational landscape at large. Realising this will require a number of significant investments and strategic steps to make it a reality. These include:

- **More investment in critical research to develop GBAs**—This includes how to more effectively capture and model a variety of learning constructs in these environments, how to best model learner data in the game and report back to the teacher and learner and how to move beyond some design limitations of current summative assessment methodologies.

- **Investing in the expanded development of GBAs**—Getting to scale with GBAs will require more effective support of the professionals needed to collaboratively build these robust tools. That means helping to support more assessment experts to understand how to engage in the design of these non-traditional assessments and how to bring more game designers into the fold of assessment design and data. Additionally, it will mean greater support to help the field more broadly to build the menu of GBAs available across the many learning constructs that we value in the curriculum.

- **Greater support for educators to use GBAs**—As is the case with many learning technologies, teachers need more professional development and support to understand how to effectively use these in the classroom. This is especially true for GBAs, where the targeted learning goals are not always as blatantly apparent as they could be with a digital quiz.

As we can see, there is much opportunity, yet much to do. The current state of the field is inspiring and promising, yet there are real limitations on how we can interpret and apply the data from these learning technologies until further research and development is completed. For these reasons, game-based environments offer a rich, but not complete answer to better assessment and learning data identification. Instead, GBAs must be part of a larger kaleidoscope of data collection and learning analytics to provide a more diverse and robust picture of learning and learners. Although the field has come far and although we realise we have much further to go, GBAs offer a potential reality where learning environments are immersive, ubiquitous and robustly adaptive, whilst simultaneously removing the negative impacts of large-scale testing—an inspiring light indeed.

**ENDNOTES**

1. *The Oregon Trail* was one of the first popular digital learning games, designed to teach school children about the realities of 19th-century pioneer life on the Oregon Trail. The player assumes the role of a wagon leader guiding a party of settlers from Missouri to Oregon’s Willamette Valley in a covered wagon in 1848. The game was released in the 1970s and became popular in U.S. elementary schools in the 1980s. [https://en.wikipedia.org/wiki/The_Oregon_Trail_(video_game)](https://en.wikipedia.org/wiki/The_Oregon_Trail_(video_game))


**REFERENCES**


How to cite this article: Groff JS. The potentials of game-based environments for integrated, immersive learning data. *Eur J Educ*. 2018;53:188–201. [https://doi.org/10.1111/ejed.12270](https://doi.org/10.1111/ejed.12270)

**APPENDIX: INDEX OF GAMES REFERENCED**

*After the Storm*
https://www.classroominc.org/programs/after-the-storm

*Deep Sea Crisis*
https://www.glasslabgames.org/games/GEM

*Impulse*
https://edgeatterc.com/edge/games/impulse

*Lure of the Labyrinth*
labyrinth.thinkport.org/www/

*Mars Generation One: Argubot Academy*
https://www.glasslabgames.org/games/AA-1

*MathBlaster*
https://www.mathblaster.com

*Newton’s Playground*
https://www.educade.org/teaching_tools/newtons-playground

*The Oregon Trail*

*The Radix Endeavor*
https://www.radixendeavor.org

*SimCityEDU: Pollution Challenge!*
https://www.glasslabgames.org/games/SC

*Use Your Brainz*
https://www.glasslabgames.org/games/PVZ