K-12 Center at ETS

Invitational Research Symposium on
Technology Enhanced Assessments

Interweaving Assessments Into Immersive Authentic Simulations: Design Strategies for Diagnostic and Instructional Insights

Chris Dede

May 7–8, 2012

K-12 Center at ETS
Invitational Research Symposium on
Technology Enhanced Assessments

Interweaving Assessments Into Immersive Authentic Simulations: Design Strategies for Diagnostic and Instructional Insights

Chris Dede

Harvard Graduate School of Education

Executive Summary

Internships in 21st century workplace settings offer potential benefits for student motivation, academic learning, and mastery of skills for the global, knowledge-based, innovation-centered economy. Secondary schools that build their learning experiences around real-world internships or apprenticeships (such as High Tech High in San Diego or The Met in Providence) are popular and effective (Wagner, 2008). However, providing extended, mentored real-world activities outside classrooms is difficult, particularly for younger students. Moreover, internship/apprenticeship models are unscalable because the number of workplace sites willing to accept mentoring responsibilities for students is limited, and teachers accustomed to conventional classrooms often struggle to adapt to this form of education. Fortunately, virtual worlds and augmented realities now offer the opportunity for all students to experience simulated, authentic internships without leaving their classrooms.

Immersion is “being there,” the subjective impression that one is in a real place (Slator, 2009). For example, a well-designed movie draws viewers into the world portrayed on the screen, and they feel caught up in that virtual environment. Immersive interfaces, such as virtual worlds and augmented realities, now can enable scalable, authentic learning both in real-world settings and in the relatively barren context of classrooms (Dede, 2009):

- Multiuser virtual environment (MUVE) interfaces offer students an engaging, immersive “Alice in Wonderland” experience in which their digital avatars in a graphical, virtual context actively participate in experiences with the avatars of other participants and with computerized agents. MUVEs provide rich environments in which participants interact with digital objects and tools, such as historical photographs or virtual microscopes (Ketelhut, Nelson, Clarke, & Dede, 2010).
- Augmented reality (AR) interfaces enable *ubiquitous computing* models. Students carrying mobile wireless devices through real-world contexts interact with virtual information,

visualizations, and simulations superimposed on physical landscapes (such as a tree describing its botanical characteristics, a historic photograph offering a contrast with the present scene, or a cloaked alien spaceship visible only through the mobile device). This type of immersion infuses digital resources throughout the real world, augmenting students' experiences and interactions (Klopfer, 2008).

As a pedagogical approach, immersive interfaces align well with situated and constructivist learning theory, since immersion positions the learner within a virtual or real-world physical and social context while guiding, scaffolding, and facilitating participatory and metacognitive learning processes such as authentic inquiry, active observation, peer coaching, reciprocal teaching, and legitimate peripheral participation with multiple modes of representation (Palincsar, 1998; Squire, 2010).

The body of this paper describes in detail two curricula based on immersion and designed to give students apprenticeship experiences in science: EcoMUVE and EcoMOBILE. Our research team's EcoMUVE middle grades curriculum (http://ecomuve.gse.harvard.edu) focuses on teaching ecosystems science concepts, scientific inquiry (collaborative and individual), and complex causality (Metcalf, Kamarainen, Tutwiler, Grotzer, & Dede, 2012). The curriculum is inquiry-based; students investigate research questions by exploring Pond and Forest immersive digital ecosystems in teams of four, with each team member having a role based on a different area of expertise (e.g., botanist, microscopic specialist). The team works collaboratively to analyze their combined data and understand the ecosystem interrelationships (Figures 1–7). The module culminates with each team creating an evidence-based concept map representing their understanding of the causal relationships in the ecosystem and presenting it to the class.

As a complement to EcoMUVE, the EcoMOBILE project (http://ecomobile.gse.harvard.edu) is exploring the unique affordances of augmented reality, as well as the capabilities of data collection probeware, to support setting-enhanced learning in environmental science education (Kamarainen, Metcalf, Grotzer, et al.,2012). As an example, *hotspots* are placed on a map of the physical setting, such as a pond, and these hotspots become accessible to students at the real location in the field. At a hotspot, students can experience augmented reality visualizations overlaid on the real environment, as well as interactive media including text, images, audio, video, 3D models, and multiple-choice or open-ended questions (Figures 8–12)**.** In a field trip to a real pond, one EcoMOBILE activity focuses on understanding the relationship between living and nonliving factors, data collection and interpretation skills, and the functional roles (producer, consumer, decomposer) of organisms in an ecosystem.

Immersive interfaces offer unique potential for interwoven assessment because of three Es: engagement, evocation, and evidence. In our research, we find that immersive authentic simulations are very engaging for students, can simulate almost any aspect of an internship in a real research group, and can collect an impressive array of evidence documenting the degree of a student's motivation and learning. However, understanding what students do and do not know as a result of open-ended learning activities requires new types of assessment designs and analytic methods.

Design approaches that seem promising for interwoven, unobtrusive, real-time diagnostic assessments include the following:

- *Capturing exploratory paths*. The paths that a student takes in exploring a virtual world to determine the contextual situation, identify anomalies, and collect data related to a hypothesis for the causes of an anomaly are an important predictor of the student's understandings about scientific inquiry.

- *Analyzing usage of guidance systems*. Gathering data on when students first choose to use an interwoven individualized guidance system, which messages they viewed, where they were in the immersive simulation when they viewed them, and what actions they took subsequent to viewing a given guidance message provides diagnostic insights that can aid instruction.
- *Interacting with animated pedagogical agents (APAs)*. APAs are "lifelike autonomous characters [that] co-habit learning environments with students to create rich, face-to-face learning interactions" (Johnson, Rickel, & Lester, 2000, p. 47). The trajectory over time of questions students ask of an APA is diagnostic—typically learners will ask for information they do not know but see as having value. This can help us comprehend a student's thought processes and methods of knowledge acquisition. Also, APAs scattered through an immersive authentic simulation can collect diagnostic information in various ways, such as the APA requesting a student to summarize what he or she has found so far.
- *Documenting progress and transfer in similar settings*. Shifting a student to a similar, but not identical environment in which he or she must identify a problem (earlier in the curriculum) or resolve a problem (later in the curriculum) can provide insights into a student's progress and aid transfer. Further, centering these benchmarking assessments on learners' common misconceptions then immediately conveying the results to students, can prompt "aha" moments that help to synthesize new levels of understanding.
- *Attaining "powers" through accomplishments*. Like leveling up in games, students can attain new powers through reaching a threshold of experiences and accomplishments. These new capabilities document team achievements, promote engagement, facilitate learning, and offer additional opportunities for interwoven assessment.

All of these types of assessment are based not on proxies (e.g., test items, essays) for real world performance, but instead on authentic actions in rich simulated contexts.

The assessment strategy overarching these design approaches is similar to what a workplace mentor would use in measuring an intern's motivation and learning. For a middle grades student in a research team internship, part of the goal would be to reach some threshold of scientific understandings and performances, but beyond that threshold the apprenticeship's primary objectives would focus on intrinsic motivation, self-efficacy, and love of learning (Tai, Liu, Maltese, & Fan, 2006). These important outcomes are difficult to measure. Immersive authentic simulations are analogous to such an internship; they are more than is needed for learning basic concepts and formulaic skills but are potentially a powerful method of scaffolding intellectual, emotional, and social capacity and commitment for future involvement (Dieterle, 2009), as well as for assessing students' progress toward these outcomes. The emphasis is not on achieving higher performance on standardized measures but on preparing learners for life. An immersive simulation with all the methods of interwoven diagnostic assessment described above would be inefficient if the metric is coverage of curricular standards—but then that is not how the value of internships and other real-world apprenticeship experiences are judged.

In an instructional model oriented to presentational teaching with the goal of high achievement on tests and essays, immersive simulations with open-ended learning-by-doing may seem a needless extravagance. But if our ultimate goal is personal and collective success in the 21st century global, knowledge-based, innovation-centered economy, then instruction, curricula, and assessment should center on real and simulated authentic learning-by-doing experiences, with just-in-time support based on interwoven diagnostic assessments, to help students surmount challenges as these are encountered (Dede, 2011).

A major next step for the EcoMUVE and the EcoMOBILE projects is to incorporate interwoven diagnostic assessments in their curricula. My colleagues on these projects and I see this opportunity as very important, now that we have developed these environments and are studying their motivation and learning. The remainder of this paper describes ways in which immersive simulations offer unique opportunities for interwoven diagnostic assessments, using EcoMUVE and EcoMOBILE as examples of how these visions could be implemented and delineating the substantial challenges in design and analysis that must be overcome to realize this potential.

The Importance of Providing and Assessing Authentic Learning Experiences for Students

Research shows the benefits of authentic experiences, such as internships in real-world contexts, for student motivation and mastery of academic knowledge. In 21st century workplaces, students also can learn digital literacies and other sophisticated skills for the global, knowledge-based, innovation-centered economy (Wagner, 2008). Historically, access to these authentic experiences has been limited, because providing students with extended, mentored learning activities outside of the classroom is difficult. A few secondary schools (e.g., High Tech High in San Diego, The Met in Providence) have built their learning experiences around real-world apprenticeships. However, this model is unscalable due to shortfalls both in number of sites willing to accept educational responsibility for the internships and in the number of teachers with the capacity to implement and sustain this form of instruction, which centers on real-world understandings and performances rather than test-based knowledge and skills.

As internships and similar real-world learning experiences illustrate, research in cognitive science has established that knowledge and skills are richly intertwined, rather than knowledge as content on which skills act as a process (Scardamalia & Bereiter, 2006). For example, applying mathematics as a perspective on understanding the world requires recognizing situations in which mathematical models might apply. This is neither a process-free content nor a content-devoid process, but a complex knowledge-skill mixture, an understanding. Categorizing what students do and do not know as understandings based on interwoven content knowledge and process skills is a more accurate depiction of how the mind works than the separation between these that current frameworks typically impose, and how students actualize those understandings in practice are performances.

In internships and apprenticeships, mentors are constantly creating stretch opportunities to help students attain new levels of learning (Hamilton & Hamilton, 2006). In this process, interweaving learning experiences with diagnostic assessments that measure learners' performances to determine their current understandings is central. Feedback from these assessments can provide mentors, students, teachers, and parents with insights that are formative for next steps in understanding and performing. But internship mentors don't give tests—so how is interwoven assessment accomplished? And how can we scale up students' access to authentic learning experiences with interwoven mentoring and assessment?

In the remainder of this paper, I first delineate ways in which immersive interfaces, such as virtual worlds and augmented realities, now offer powerful affordances enabling scalable, authentic learning both in real-world settings and in the relatively barren context of classrooms. As specific cases that illustrate these capabilities, I describe two research projects my colleagues and I are conducting at Harvard to design and study immersive middle-school ecosystems science curricula: EcoMUVE and EcoMOBILE. Next, I delineate how virtual worlds and augmented realities offer unique opportunities for interwoven diagnostic assessments, in part because immersive simulations motivate student performance, evoke authentic behaviors, and generate rich datasets about each student's detailed actions. As illustrations, I discuss ways of evolving the current design of EcoMUVE and EcoMOBILE to realize this potential. As part of this vision, I discuss substantial challenges in such an evolution. For example, incorporating interwoven assessments that do not disrupt the flow and authenticity of immersive experiences requires sophisticated design that seamlessly interweaves learning and measurement. Also, interpreting the diagnostic and instructional implications of detailed performance records in open-ended learning environments is very difficult. I conclude the paper with recommendations for the field and next steps for research.

New Capabilities for Immersive Learning via Virtual Worlds and Augmented Realities

Immersion is "being there," the subjective impression that one is in a real place (Slator, 2009, p. 3549). For example, a well-designed movie draws viewers into the world portrayed on the screen, and they feel caught up in that virtual environment. Technologies can induce immersion via the sensory stimuli, participants' abilities to influence what happens in the environment, and the use of narrative and symbolism (Dawley & Dede, in press). Two types of immersive interfaces underlie a growing number of formal and informal learning experiences (Dede, 2009):

- Multiuser virtual environment (MUVE) interfaces offer students an engaging "Alice in Wonderland" experience in which their digital avatars in a graphical, virtual context actively participate in experiences with the avatars of other participants and with computerized agents. MUVEs provide rich environments in which participants interact with digital objects and tools, such as historical photographs or virtual microscopes (Ketelhut et al., 2010).
- Augmented reality (AR) interfaces enable *ubiquitous computing* models. Students carrying mobile wireless devices through real-world contexts interact with virtual information, visualizations, and simulations superimposed on physical landscapes (such as a tree describing its botanical characteristics, a historic photograph offering a contrast with the present scene, or a cloaked alien spaceship visible only through the mobile device). This type of immersion infuses digital resources throughout the real world, augmenting students' experiences and interactions (Klopfer, 2008).

The military and the entertainment industry have expended substantial resources in developing these immersive media, which have many applications in precollege and higher education, as well as for training and professional development.

For example, both of these immersive interfaces enable developing learning experiences in which students encounter richly detailed, simulated real-world situations with challenges that can be resolved through applying academic knowledge and skills (e.g., the biology-based protocols and practices used by a diagnostician examining a virtual patient to determine what illness is present). Sometimes these learning experiences are developed as games that foster problem solving skills related to a particular role or situation (Barab, Gresalfi, & Ingram-Goble, 2010; Shaffer & Gee, 2007). The EcoMUVE and EcoMOBILE curricular case studies in this paper are instead gamelike, authentic simulations that draw on theory-based and empirical insights about engagement, motivation, flow, and self-efficacy. However, they deliberately do not include some types of engagement common in games (e.g., extrinsic rewards, scoring systems, competition among participants) that may undercut intrinsic motivation and learning for some students, using instead the importance and relevance of the real-world challenge as the primary motivator.

How are immersive media different from other types of learning experiences, such as tutoring systems or nonimmersive games and simulations, which more tightly constrain learner interactions in an instructional trajectory? (An example of the latter would be a two-dimensional game in which one fires a cannon at a distant object, changing the angle of the cannon's barrel to determine where its shell will impact.) Instruction in immersive settings can foster constructivist/interpretivist learning by providing rich, loosely structured experiences and guidance (such as apprenticeships, coaching, and mentoring) that encourage meaning-making without imposing a fixed set of knowledge and skills or a narrow path to success (Dunleavy & Dede, in press). Constructivist/interpretivist theories of learning assume that meaning is imposed by the individual rather than existing in the world independently (Dede, 2008). People construct new knowledge and understandings based on what they already know and believe, which is shaped by their developmental level, their prior experiences, and their sociocultural background and context.

Constructivist learning theory outlines five conditions most likely to enhance learning: (a) embed learning within relevant environments, (b) make social negotiation integral to the learning experience, (c) provide multiple perspectives and multiple modes of representation, (d) provide self-directed and active learning opportunities, and (e) support and facilitate metacognitive strategies within the experience (Driscoll, 2000). Knowledge is interwoven in the setting in which it is used; learning involves mastering authentic tasks in meaningful, realistic situations—physical or virtual. Learners build personal interpretations of reality based on experiences and interactions with others, creating novel and situation-specific understandings.

Immersive interfaces also enable applying situated learning theory, in which learning is defined as embedded within and inseparable from participating in a system of activity deeply determined by a particular physical and cultural setting (Chaiklin & Lave, 1993). The unit of analysis is neither the individual nor the setting, but the relationship between the two, as indicated by the student's level of participation (Greeno, 1998). From this perspective, learning and cognition are understood both as progress along trajectories of participation in communities of practice and as the ongoing transformation of identity (Wenger, 1998).

As a pedagogical approach, immersive interfaces align well with both constructivist and situated learning theories. Immersion positions the learner within a virtual or real-world physical and social context while guiding, scaffolding, and facilitating participatory and metacognitive learning processes such as authentic inquiry, active observation, peer coaching, reciprocal teaching, and legitimate peripheral participation with multiple modes of representation (Palincsar, 1998; Squire, 2010).

Immersive interfaces also can aid with the important educational challenge of transfer. Transfer is defined as the application of knowledge learned in one situation to another situation and is demonstrated if instruction on a learning task leads to improved performance on a transfer task, ideally a skilled performance in a real-world setting (Mestre, 2002). One of the major criticisms of instruction today is the low rate of transfer generated by presentational instruction; even students who excel in educational settings often are unable to apply what they have learned to similar real-world contexts (Dede, 2008). Teaching in a barren setting, such as a classroom, requires far-transfer: applying knowledge learned in a situation to a quite different context whose underlying semantics are associated but distinct.

When transfer is defined as learning that empowers future learning (Schwartz, Sears, & Bransford, 2005), researchers measure transfer by focusing on extended performances where students learn how to learn in a rich environment and then solve related problems in real-world contexts. The potential advantage of immersive interfaces and situated learning is that their simulation of real-world problems and contexts means that students must achieve only near-transfer: applying the knowledge learned in a situation to a similar context with somewhat different surface features. Flight and surgical simulators demonstrate near-transfer of psychomotor skills from digital simulations to real-world settings; research on the extent to which less elaborate types of immersive interfaces can foster transfer of cognitive and psychosocial skills is an important frontier for the field.

Overall, immersive interfaces are best used as a medium for learning and assessing sophisticated knowledge and skills that are contextually dependent and are not based on a narrow formulaic method. An example used in the EcoMUVE and EcoMOBILE curricular cases is scientific inquiry. This is a well-studied cognitive and psychosocial understanding/performance based in part on subject matter knowledge appropriate for a particular context (e.g., physics in a situation involving magnetic fields, such as high voltage power lines) and in part on skills such as using appropriate tools and techniques to gather, analyze, and interpret data; developing prescriptions, explanations, predictions, and models using evidence; thinking critically and logically to make the relationships between evidence and explanations; and recognizing and analyzing alternative explanations and predictions (National Committee on Science Education Standards and Assessment, National Research Council, 1996). Scientific inquiry (in its deepest sense) cannot effectively be learned through a tutoring system or some form of direct instruction; inquiry skills are best motivated and taught either in a real-world apprenticeship as part of a research team or in an immersive simulation of this experience.

Capabilities of virtual worlds for motivation and learning. In a virtual world, each participant uses an avatar (a digital representation of oneself) to interact with digital agents, artifacts, and contexts. Theories about motivation from social psychology describe various reasons why participants might become highly engaged in a virtual world and might be motivated to frequently seek out this experience (Dawley & Dede, in press). Aspects of a virtual experience that promote intrinsic motivation include intrapersonal factors such as challenge, control, fantasy, and curiosity, as well as interpersonal factors such as competition, cooperation, and recognition (Bartle, 2003). The challenge dimension of engagement is heightened when a participant achieves a state of flow through facing challenges that are difficult but surmountable at his or her current level of skill (Csikszentmihalyi, 1991). Other generic, intrinsic factors that heighten motivation include the perceived instrumental value of an activity (Brophy, 1999), perceived personal competence in accomplishing the goals of an activity (Dweck, 2002; Schunk & Parajes, 2005), and perceived autonomy in making choices within an activity (Ryan & Deci, 2000).

Przybylski, Rigby, and Ryan (2010) summarized these dimensions of motivation as applied to videogames, many of which involve immersive interfaces. Lepper and Henderlong (2000) described various ways that extrinsic incentives used to promote participating in an activity, but unrelated to the intrinsic nature of the experience, can undercut learning and intrinsic motivation, if overdone. Both they and Habgood and Ainsworth (2011) suggested strategies for ensuring that educational experiences that begin with extrinsic motivators culminate in participants having strong intrinsic motivation.

The experience of situated embodiment lies at the heart of immersive experiences in which one feels psychologically present in a context that is not where the person is physically located (Winn, 2003). In virtual worlds (VWs) and immersive simulations, situated embodiment is based on the willing suspension of disbelief (Dawley & Dede, in press). Situated embodiment in virtual environments and immersive simulations offers the potential for identity exploration, in which a participant plays a role different than that portrayed by that person in everyday life (Turkle, 1997, 2005). Laurel (1993) and Murray (1997) described design strategies that can enhance participants' identity exploration, such as providing options to modify the avatar's appearance, gender, or clothing; creating role-play opportunities in historical or fantasy-based settings; and experiencing learning opportunities to be someone other than themselves and reflect on the shift.

Social network knowledge construction (SNKC) is a pedagogical model for virtual world learning (Dawley, 2009). SNKC takes advantage of the various social network communication mechanisms that are available to participants in virtual worlds, leading learners through a five-stage process: identify, lurk, contribute, create, and lead. Dawley (2009) listed over 15 in-world and out-of-world communication mechanisms potentially available in virtual worlds. When leveraged effectively, these communication options can support increased engagement and motivation, group action, individual transformation, and shared meaning-making opportunities (Dawley & Dede, in press). Community presence to induce a sense of belonging and group purpose is another affordance supported in virtual worlds through communication mechanisms such as groups, guilds, and clans (Warburton, 2009).

Educational virtual worlds are practical in school settings when the cost of development is amortized over a large number of students. Immersive media can be so motivating that students voluntarily spend substantial extra time in them outside of the classroom (National Research Council, 2011). Virtual worlds designed for informal learning are not so engaging that they will displace entertainment games, but some sites such as Whyville have garnered a large set of followers who spend substantial time there (Kafai, Quintero, & Felton, 2010). Sustainability, distribution, and customer acceptance are the three major hurdles preventing a rapid expansion of virtual worlds into educational settings (National Research Council, 2011). These barriers are unfortunate, because the 2010 National Education Technology Plan recommendations reference the potential power of this medium for learning and assessment (U.S. Department of Education, 2010):

> 2.3 Conduct research and development that explores how interwoven assessment technologies, such as simulations, collaboration environments, virtual worlds, games, and cognitive tutors, can be used to engage and motivate learners while assessing complex skills. (p. xvii)

This underscores the importance of research on embedding diagnostic assessment formative for instruction into immersive learning environments.

A case example of virtual worlds for learning: the EcoMUVE curriculum. Complex causality underlies many problems in the 21st century, such as global climate change and the world financial crisis, in part because many people use only simple causal models in their reasoning. For example, students typically use simple linear causal forms in their science learning —reasoning that one thing directly makes another thing happen. They also tend to focus on obvious variables, ones they can perceive directly (Grotzer, 2004). This hampers learning about complex causal relationships in science. As an illustration, ecosystems are complex systems that are impacted by nonobvious as well as obvious causes, distributed causality, and effects at a distance and over long periods of time. An understanding of complex causality is necessary to understand the dynamics involved in concepts such as energy transfer, matter recycling, decomposition, and interaction between biotic (living) and abiotic (nonliving) factors (Grotzer, Dede, Metcalf, & Clarke, 2009). Even after instruction, students often retain inaccurate interpretations about ecosystems' structural patterns and systemic causality (Grotzer & Basca, 2003).

In response to these challenges for learning, our EcoMUVE middle grades curriculum (http://ecomuve.gse.harvard.edu) focuses on the potential of immersive authentic simulations for teaching ecosystems science concepts, scientific inquiry (collaborative and individual), and complex causality (Metcalf , Kamarainen, Grotzer, & Dede, 2012). This four-year project is funded by the Institute of Education Sciences, U.S. Department of Education. The *EcoMUVE* curriculum consists of two MUVE-based modules, which center on pond and forest virtual ecosystems. Each module consists of ten 45-minute lessons and represents an ecological scenario involving complex causality. The curriculum is inquiry-based; students investigate research questions by exploring the virtual ecosystem and collecting data from a variety of sources over time, assuming roles as ecosystems scientists in this immersive authentic simulation (Figure 1).

Figure 1. A student uses an avatar to explore a virtual ecosystem.

An illustrative activity in a virtual pond ecosystem. The description in this paper focuses on the pond module (Metcalf, Kamarainen, Tutwiler, Grotzer, & Dede, 2011). The curriculum uses a jigsaw pedagogy; students work in teams of four (Figure 2) and are given roles: areas of expertise (e.g., botanist, microscopic specialist). Students use interactive learning quests to learn more about the content specific to their module, (e.g., what is pH). Each student then performs data collection specific to his or her role in the virtual pond ecosystem, sharing these data with teammates within the immersive interface via tables and graphs. Each team works collaboratively to analyze the combined data and understand the ecosystem interrelationships. The module culminates in each team creating an evidence-based concept map representing their understanding of the causal relationships in ecosystem and presenting it to the class.

Naturalist	Microscopic Specialist	Water Chemist	Private Investigator
Find out how the populations of pond organisms: largemouth bass, bluegill, minnows, and great blue herons change over time.	Find out how the populations of microscopic bacteria, blue green algae, and green algae change over time.	Use the atom tracker to find out what happens to the carbon atom on different days.	Gather clues from the landscaper, the golf course manager, the utility worker, the park ranger, the birdwatcher, other people near the pond.
Use the field guide to learn about the different fish species.	Measure the dissolved oxygen in the water on different days.	Measure the dissolved oxygen in the water on different days.	Observe the weather on different days; collect measurements of temperature, cloud cover, and wind speed.
Use the atom tracker to find out what happens to the carbon atom on different days.	Use the atom tracker to find out what happens to the oxygen atom on different days.	Use the atom tracker to find out what happens to the phosphorus atom on different days.	Measure chlorophyll a in the water on different days.
Measure the turbidity in the water (and use your eyes) to see changes over time.	Measure the temperature in the water on different days.	Measure the pH in the water on different days.	Measure the temperature in the water on different days.
Measure the dissolved oxygen in the water on different days.	Measure chlorophyll a in the water on different days.	Measure the nutrients (phosphates and nitrates) in the water on different days.	Measure the nutrients (phosphates and nitrates) in the water on different days.
Work together to create a concept map that represents the causal relationships of the pond ecosystem based on whole team's observations.			

Figure 2. Students work in teams of four with complementary data collection roles.

Students can explore the virtual pond and the surrounding area, even under the water; see realistic organisms in their natural habitats; and collect water, weather, and population data (Figures 3–5). Students visit the pond over a number of virtual "days" and eventually make the surprising discovery that, on a day in late summer, many fish in the pond have died. Students are challenged to figure out what happened—they travel backward and forward in time to gather information to solve the mystery and understand the complex causality of the pond ecosystem (Figure 6).

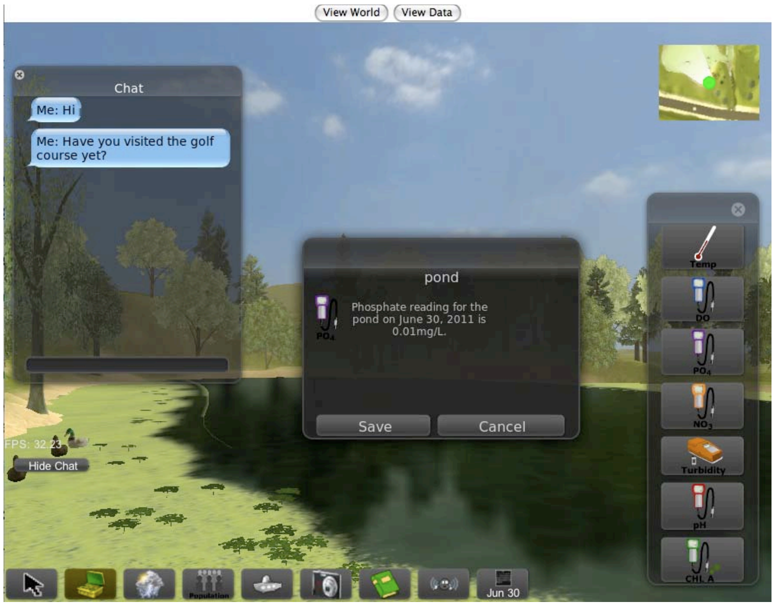

Figure 3. Students can collect water, weather, and population data at the digital pond.

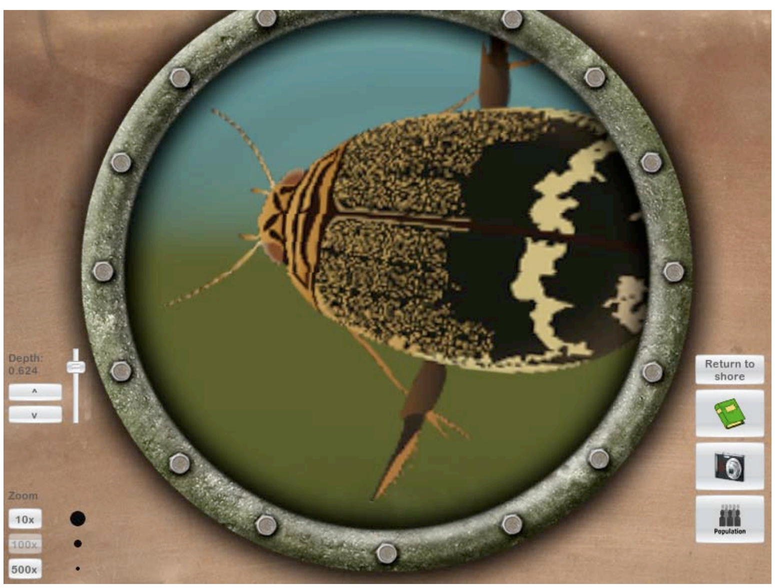

Figure 4. The submarine tool allows students to see and identify microscopic organisms.

Figure 5. Talking to Manny and observing the bags of fertilizer.

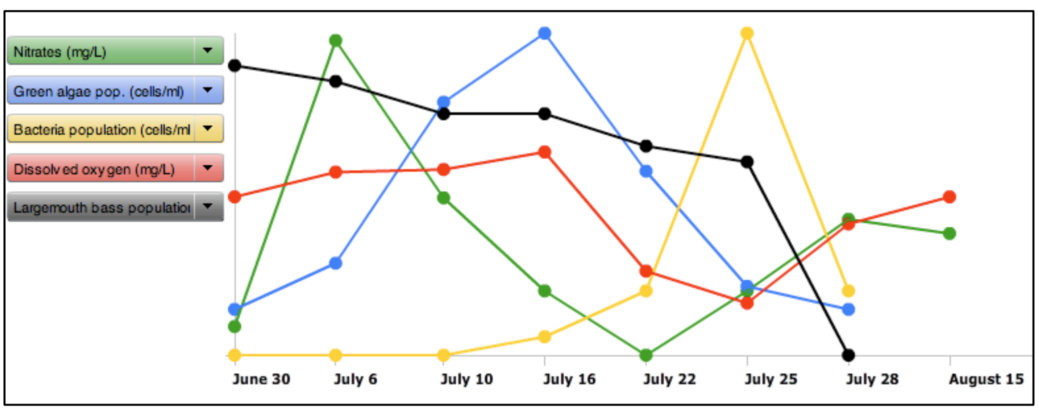

Figure 6. Summarizing and interpreting data gathered over time.

We are studying the extent to which student teams can develop concept maps that explain the complicated process by which the fish died (Figure 7). Our research measures also include student gains in motivation to do science and feelings of self-efficacy about their abilities as scientists, student gains in ecosystems science content, and students' abilities to perform individual and collaborative scientific inquiry. We implemented the EcoMUVE curriculum in 121 classrooms in 2011 and are analyzing the many types of data we collected. Early findings (Grotzer, Kamarainen, Tutwiler, Metcalf, & Dede, 2012; Kamarainen, Metcalf, Grotzer, et al., 2012; Metcalf et al., 2012) are very encouraging in terms of gains on all these dimensions.

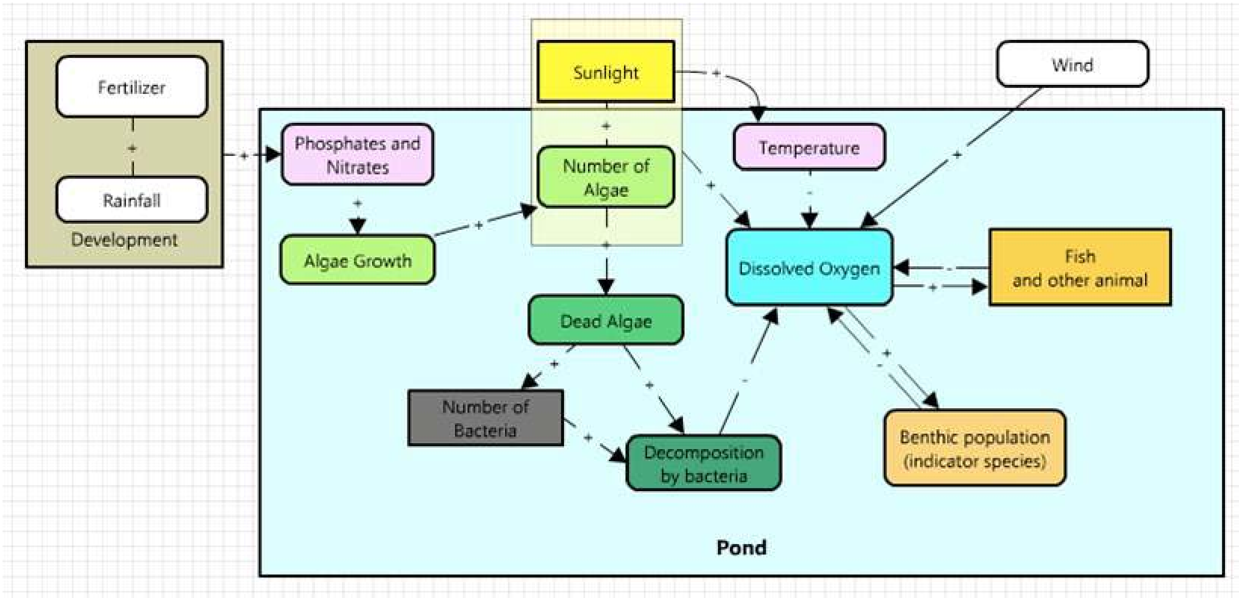

Figure 7. Causal process underlying fish kill.

Capabilities of augmented realities for motivation and learning. As a complementary type of immersive interface to virtual worlds, AR utilizes mobile, context-aware technologies (e.g., smartphones, tablets) and software that enables participants to interact with digital information interwoven within the physical environment (Dunleavy & Dede, in press). Two forms of AR are currently available to educators: location-aware and vision-based AR. Location-aware AR presents digital media to learners as they move through a physical area with a GPS-enabled smartphone or similar mobile device. The media (i.e., text, graphics, audio, video, 3D models) augment the physical environment with narrative, navigation, and/or academic information relevant to the location. In contrast, vision-based AR presents digital media to learners after they point the camera in their mobile device at an object (e.g., QR code, 2D target).

As an interface that fosters a blend of digital immersion and the real world, AR shares with virtual worlds many of the affordances for learning already discussed. The unique power of AR as a learning tool is its ability "to enable students to see the world around them in new ways and engage with realistic issues in a context with which the students are already connected" (Klopfer & Sheldon, 2010, p. 86). This includes the ability to present to a group of learners multiple incomplete yet complementary perspectives on a problem situated within a physical space, enabling collaborative pedagogical techniques and design approaches, such as jigsaw and differentiated role-play, which lend themselves well to inquiry-based activities requiring argumentation (Dunleavy, Dede, & Mitchell, 2009). By embedding these multiple perspectives within the environment and contextualizing them within a problem-based narrative, AR also affords educators the ability to leverage physical space as an additional layer of content for students to observe, manipulate and analyze (Squire et al., 2007).

Numerous studies document high engagement by students and teachers as a result of using the handhelds: adopting roles; negotiating meaning within active, inquiry-based compelling narratives; solving authentic problems; and physically exercising (Dunleavy & Dede, in press). The ability to access outside resources (i.e., Internet) and additional software on the devices to more effectively resolve the learning environment's challenges is another unique affordance of AR. Overall, both virtual worlds and AR as immersive interfaces are exciting opportunities for research about how diagnostic assessment formative for instruction can aid these powerful new types of learning environments.

A case example of a reality augmented for learning: The EcoMOBILE experience. As a curricular case study, the EcoMOBILE project (http://ecomobile.gse.harvard.edu) is exploring the unique affordances of AR, as well as the capabilities of data collection probeware, to support setting-enhanced learning in environmental science education (Kamarainen et al., 2012). The EcoMOBILE curriculum is a blend of the EcoMUVE learning experiences with these new, field-based activities. This research is funded by the National Science Foundation and by Qualcomm, Inc. and is supported with resources from Texas Instruments, Inc.

The ability to understand ecosystems is enriched by experiences in real environments. Field trips, both real and virtual, lead to increases in science knowledge (Garner & Gallo, 2005; Harrington, 2011), and outdoor experiences can positively influence student attitudes about nature (Ballantyne & Packer, 2002). Yet, the real world can be a difficult learning environment: Students may be distracted by the novelty of the social and physical context of the experience, making a focus on appropriate learning tasks difficult (Orion & Hofstein, 1994). Also, students may be overwhelmed by too much information and may struggle to discern where to direct their attention. As a result of these and logistical factors, field trips tend to be one-time experiences with limited connections to what students experience in the classroom curriculum or in their everyday lives.

As one means of improving this situation, research has shown that using data collection probes in science can enhance various aspects of teaching and learning in the classroom and in the field. Using probes in a lab setting coupled with computer-mediated presentation of the results scaffolds critical evaluation of graphs and data (Nicolaou et al., 2007), supports student learning of science concepts (Metcalf & Tinker, 2004), and empowers inquiry-based science learning (Rogers & Price, 2008). Real-time probeware helps students to see data and related concepts as concretely related to particular phenomena (Vonderwell, Sparrow, & Zachariah, 2005).

A combination of both AR and environmental probes may further enhance the field trip experience in ways that neither technology could accomplish on its own. The EcoMOBILE team is studying whether using probes and/or mobile devices equipped with AR experiences can enhance learning by situating data collection activities in a larger, meaningful context that connects to students' activities at the real-world setting.

In our pilot studies, the technology components have included an AR experience running on wireless-enabled mobile devices, along with either water measurement test kits or graphing calculators with environmental probes. The AR experiences were created using the FreshAiR AR development platform (http://playfreshair.com) designed by MoGo Mobile, Inc. The FreshAiR platform allows an author to create AR games and experiences that can be accessed anywhere from an Android mobile device with wireless connectivity and GPS capabilities (Figure 8). Hotspots are placed on a map of the physical setting, and these hotspots become accessible to students at the real location in the field

(Figure 9). At a hotspot, students can experience AR visualizations overlaid on the real environment, as well as interactive media including text, images, audio, video, 3D models, and multiple-choice or open-ended questions (Figure 10).

Figure 8. Students work in teams to explore the pond ecosystem.

Figure 9. A green hotspot showing the direction and distance to the next location.

Figure 10. Using a visual target to see a 3D image on the mobile device.

An illustrative pilot activity in an ecosystem augmented for learning. To help us better understand how to design AR-based curricula, in fall 2011 we developed three distinct environmental science AR experiences as pilot implementations for middle grades students. We provided a variety of content types and process structures to assess which elements are most engaging and powerful. As an example, during a field trip to a pond the Scientific Discoveries pilot focused on understanding of the relationship between biotic and abiotic factors, data collection and interpretation skills, and the functional roles (producer, consumer, decomposer) of organisms in an ecosystem. This learning experience included one class period before the field trip, the field trip itself, and one class period after the field trip (Kamarainen, Metcalf, Grotzer, et al., 2012). Prior to the field trip, the students had access to *learning quests*, which are online modules providing a 5–10 minute activity that introduces the students to the ideas behind dissolved oxygen, turbidity, and pH.

In the Scientific Discoveries pilot, students collected water measurements using Texas Instruments (TI) NSpire devices with Vernier environmental probes. The TI NSpire provides graphing calculator capabilities along with a Data Quest data collection mode that allows display of multiple probe readings on a single interface. Probes were provided to measure four variables: dissolved oxygen concentration, turbidity (Figure 11), pH, and water temperature.

Figure 11. Collecting water quality data on turbidity using digital probes.

The field trip lasted approximately 3.5 hours. The activities during the field trip included the following AR experiences, along with more conventional field trip activities:

Upon arriving at a hotspot near the pond, students working in pairs were prompted to make observations about the organisms around the pond and classify (producer, consumer, decomposer) an organism they observed. Students answered questions about their observations, and received constructive feedback based on their answers.

- At the next hotspot, students were prompted to collect water measurements using the TI NSpire and environmental probes. The AR delivered additional information that helped them make sense of the measurements they had collected. Student recorded their data on a worksheet.
- Students were then prompted to collect water measurements at a second location that they could choose. Students once again recorded their data and were prompted to compare the two measurements.
- At a later hotspot, students were prompted to sketch on paper an organism they had observed near the pond.
- Two more hotspots provided visual overlays, videos, and additional information related to consumers and decomposers, as well as posed questions related to the role of these organisms in the ecosystem.
- As the final activity in the field, students met with another pair of students who had collected the other two water quality variables, and the two pairs compared their measurements before returning to the classroom.

The AR supported students' use of the probes by helping them navigate to a location to collect a sample, providing step-by-step instructions for use of the probes, entering the reading in response to a multiple-choice question, and delivering immediate feedback related to the student-collected measurement (Figure 12).

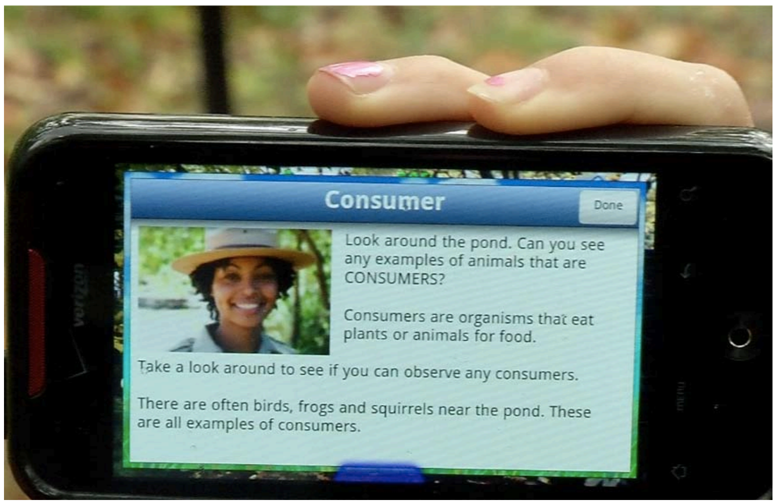

Figure 12. Handheld device delivering background information about their research assignment.

On the next school day after the field trip, back in the classroom, students compiled all of the measurements of temperature, dissolved oxygen, pH, and turbidity that had been taken during the field trip. They looked at the range, mean, and variations in the measurements and discussed the implications for whether the pond was healthy for fish and other organisms. They talked about potential reasons why variation may have occurred, how these measurements may have been affected by environmental conditions, and how to explain outliers in the data.

As discussed in Kamarainen, Metcalf, Grotzer, et al. (2012), our findings from the Scientific Discoveries pilot showed that students were highly engaged with the technology and also with science. Teachers were able to use pedagogical approaches that may otherwise be difficult in an outdoor learning environment. Student learning gains on the content survey were significant both from a statistical perspective and from the viewpoint of the teachers, who compared these gains to memories of prior field trips without technological support. These results suggest that combining AR with use of probes inside and outside the classroom holds potential for helping students to draw connections between what they are learning and new situations. Overall, our findings indicate that there are multiple benefits to using this suite of technology for teaching and learning.

A major next step for the EcoMUVE and the EcoMOBILE projects is to incorporate interwoven diagnostic assessments in their curricula. We see this opportunity as very important, now that we have developed these environments and are studying their motivation and learning. The remainder of this paper describes ways in which immersive simulations offer unique opportunities for interwoven diagnostic assessments using EcoMUVE and EcoMOBILE as examples of how these visions could be implemented and delineating the substantial challenges in design and analysis that must be overcome to realize this potential.

Embedding Diagnostic Assessments in Immersive Simulations

Some types of technology-based learning-by-doing environments are tightly structured in several respects. The phenomenon to be studied is clearly defined, without extraneous or unrelated information clouding the situation. Also, at each decision point, the range of possible actions is limited to a few alternatives, all directly related to the phenomenon rather than possibly off-task. Often, the understandings and performances to be learned are relatively simple and can be encapsulated in factual statements and procedural recipes. These tightly structured characteristics typically apply to nonimmersive simulations (Quellmalz & Haertel, 2004) and to tutoring systems (Koedinger & Corbett, 2006; VanLehn, 2006). Many computer-based games also fit this description.

In contrast, internships in workplace contexts, immersive simulations, field trips to real-world settings, and some computer-based games (e.g., role-playing adventure games set in virtual environments, such as Quest Atlantis [Barab et al., 2010]) are more open-ended learning experiences. Students encounter many phenomena, only some of which are related to what they seek to learn; and participants' range of possible actions at any given point is quite broad. The understandings to be developed are complex (e.g., problem finding in unstructured situations) and nonformulaic.

In tightly structured simulations, well-understood methods are available for collecting data about student behaviors, analyzing these performances to determine what learners understand and do not understand at a particular stage of instruction, and providing individualized feedback to students and teachers that is formative for next steps in learning and instruction (Brown, Hinze, & Pellegrino, 2008). Tutoring systems and related assistance environments also have mature methods for collecting data about learner actions, interpreting this information, and providing diagnostic feedback (Heffernan, Heffernan, Decoteau, & Militello, 2012). Other papers from the technology enhanced assessments symposium (Behrens, DiCerbo, & Ferrara, 2012; Heffernan & Koedinger, 2012; Levy, 2012) discussed ways in which embedding diagnostic assessments in these tightly structured learning environments is complicated and challenging. However, well-understood and mature strategies are available for formatively improving learning and instruction for the types of educational content and outcomes that these systems can support. In contrast, this paper focuses on less well understood, immature methods for the more problematic challenge of embedding diagnostic assessments in open-ended learning experiences such as immersive simulations.

Immersive interfaces offer unique potential for interwoven assessment because of three Es: engagement, evocation, and evidence. In our research, we find that immersive authentic simulations are very engaging for students, even though we deliberately avoid some types of motivation common in games (Dede, 2009). Students try hard to succeed, and the learning experiences promote their self-efficacy as a scientist. This means that, unlike many types of testing, all students are putting forth their best efforts.

Further, immersive interfaces can evoke a wide spectrum of performances. As EcoMUVE and EcoMOBILE illustrate, within the structure of a narrative about problems in an ecosystem, we can simulate almost any performance that students might demonstrate within an internship in a real research group. This means that a very broad palette of learning experiences—and assessment situations by which a mentor would assess progress towards mastery—can create opportunities for students to reveal their degrees of engagement, self-efficacy, understandings, and performances.

Drawing on a broad spectrum of performances is important in determining the true extent of what a student knows and does not know. For a decade, we developed and studied the River City curriculum (http://muve.gse.harvard.edu/rivercityproject/); middle grades students learned epidemiology and biological principles, as well as collaborative scientific inquiry, by traveling back about 130 years to an immersive virtual town plagued by diseases (Ketelhut et al., 2010). In our detailed analysis of students' ongoing activities and interactions, we found evidence of learning that was not captured by pre/post-tests or by a scientific-conference presentation student teams gave at the end of the unit (Ketelhut, Dede, Clarke, Nelson, & Bowman, 2007). The evidentiary trail of learning trajectories afforded by interwoven diagnostic assessments is richer and often more valid than a snapshot summative measure, even a rich artifact like a synthesis presentation.

Finally, immersive interfaces can collect an impressive array of evidence about what a learner knows (and does not know), what he or she can do (and cannot do), and whether he or she knows when and how to apply disciplinary frames and prior knowledge to a novel problem. Immersive environments—because of their situated nature and because they generate log files—make it easy to design for eliciting performances, to collect continuous data, and to interpret structures of evidence. In a virtual world, the server documents and timestamps actions by each student: movements, interactions, utterances, saved data, and so on. In an AR, the mobile device can save moderately detailed information about movements and actions, and using the device to record learners' voices as their team interacts could provide another resource for analysis. Given engagement, evocation, and evidence, immersive learning interfaces potentially are the most powerful and valid assessment medium available—but can we realize this potential?

Difficulties of interwoven diagnostic assessment in immersive simulations. Quellmalz, Timms, and Schneider (2009) examined issues of embedding assessments into games and simulations in science education. Their analysis included both tightly-structured and open-ended learning experiences. After studying several immersive games and simulations related to learning science, including River City, they noted that the complex tasks in simulations and games cannot be adequately modeled using only classical test theory and item response theory. This shortfall arises because these complex tasks have four characteristics (Williamson, Bejar, & Mislevy, 2006). First, completion of the task requires the student to undergo multiple, nontrivial, domain-relevant steps and/or cognitive processes. Second, multiple elements, or features, of each task performance are captured and considered in the determination of summaries of ability and/or diagnostic feedback. Third, the data vectors for each task have a high degree of potential variability, reflecting relatively unconstrained work product production. Fourth and finally, evaluation of the adequacy of task solutions requires the task features to be considered as an interdependent set, for which assumptions of conditional independence typically do not hold.

Quellmalz et al. (2009) concluded that, given the challenges of complex tasks, more appropriate measurement models for simulations and games—particularly those that are open-ended—include Bayes nets, artificial neural networks, and model tracing. They added that new psychometric methods beyond these will likely be needed. Beal and Stevens (2007) used various types of probabilistic models in studying students' performance in simulations of scientific problem solving. Bennett, Persky, Weiss, and Jenkins (2010) described both progress in applying probabilistic models and the very difficult challenges involved. Behrens, Frezzo, Mislevy, Kroopnick, and Wise (2007) described ways of embedding assessments into structured simulations; and Shute, Ventura, Bauer, and Zapata-Rivera (2009) delineated a framework for incorporating stealth assessments into games.

Not surprisingly given these analytic difficulties and the open-ended nature of immersive learning environments compared to structured simulations and games, we encountered many challenges in attempting to understand students' progress during the course of the River City curriculum (Ketelhut et al., 2007). River City was deliberately designed as a very open-ended learning environment in which many paths to success were available. Further, resolving the town's issues with illness required students first to infer that three different diseases were simultaneously present and then to shift their activities to studying one of these diseases. While scientifically valid and important as a learning experience in complex inquiry, this complicated situation necessitated relatively unfocused data gathering by students until the realization of superimposed illnesses was reached.

As a result of all these factors in River City, interpreting students' detailed actions to understand their intent—and what level of ongoing performance they actually achieved—was very difficult, although Clarke (2009) developed analytic methods to accomplish this. For each student, by combining records of movements, interactions with the world, chat-logs with team members, and artifacts produced, we were able to formulate case studies documenting individual learning trajectories. However, these case studies required intensive human effort and expertise, and the methods used are impractical and unscalable for real-time diagnosis formative for instruction. SAVE Science (http://www.savescience.net/) is a research project currently studying the issues of data-mining records of student actions in an immersive virtual environment for learning science inquiry (Ketelhut et al., 2012).

The virtual performance assessment project. To better understand the assessment challenges involved with the MUVE interface, in 2008 my colleagues and I began our Virtual Performance Assessment (VPA) project, funded by the Institute for Education Sciences of the U.S. Department of Education, as well as the Gates Foundation. We are developing and studying the feasibility of immersive virtual performance assessments to assess the scientific inquiry skills of middle grades students as a standardized component of an accountability program (http://vpa.gse.harvard.edu). The goal is to provide states with reliable and valid technology-based performance assessments linked to state and national science education standards for inquiry processes (Clarke-Midura, Dede, & Norton, 2011).

Applying the evidence-centered design (ECD) approach (Mislevy & Haertel, 2006; Mislevy & Rahman, 2009) allowed us to articulate every aspect of the VPAs, from the knowledge, skills, and abilities (KSAs) measured to the types of evidence that allow making claims about what students know and do not know. Using the principled assessment designs for inquiry (PADI) system has enabled us to create multiple forms of the same assessment, for a generalizability study, and to reframe science inquiry constructs (theorizing, questioning and hypothesizing, investigating, analyzing and synthesizing) into specific KSAs aligned with current national standards (Clarke-Midura, Mayrath, & Dede, in press). Clarke-Midura is currently studying the extent to which the attribute hierarchy method (Wang & Gierl, 2011) and Bayesian network models (Mislevy, Steinberg, & Almond, 2003) provide measurement models suitable for the types of data generated from virtual performance assessments.

The VPA work has clarified how to design ECD-based summative assessments of sophisticated cognitive performances in virtual worlds. We have established that this type of assessment is practical and affordable at scale, as well as more valid for sophisticated performances like scientific inquiry than paper-and-pencil, item-based assessments (Clarke & Dede, 2010). While we are still determining the psychometric properties of our virtual performance assessments, we believe this medium has great potential to complement more standard forms of testing. However, our research has also revealed that the design of a virtual world as a summative assessment results in an environment necessarily too narrow in its structure to allow many of the powerful forms of open-ended learning described above.

All these issues lead to the conclusion that the difficulties of embedding diagnostic assessments in immersive simulations, as well as in other open-ended learning environments with complex tasks, cannot be fully resolved—at least in the near-term—through more sophisticated analytic techniques. New types of design strategies are also needed to create immersive authentic simulations that have interwoven aspects amenable to diagnostic measurement and real-time formative intervention.

Design Strategies for Ameliorating Analytic Difficulties in Interwoven Assessments

This section describes design strategies that can preserve the open-ended learning that immersive interfaces empower, while at the same time enabling tractable real-time analysis of data diagnostic of students' understandings. An important constraint in these designs is that, as much as possible, the assessment activities and data collection must be unobtrusive. Otherwise, the assessment dimension of the experience disrupts immersion and engagement through undercutting flow and authenticity, which in turn can undermine learning (National Research Council, 2011). These design strategies are illustrated with examples from River City and with potential applications to EcoMUVE and EcoMOBILE.

Paths and heat maps. The paths that a student takes in exploring a virtual world to determine the contextual situation, identify anomalies, and collect data related to a hypothesis for the causes of an anomaly are an important predictor of the student's understanding of scientific inquiry. In River City, we used log file data to generate event paths (Figure 13) for both individual students and their three person teams. Students and teachers found this a useful source of diagnostic feedback on the relative exploratory skills—and degree of team collaboration—that these performances exhibited.

Dukas (2009) extended this research by developing an avatar log visualizer (ALV), which generates a series of slides depicting the relative frequency events of one or more subpopulations of students, aggregated by user-specified location and time bins. Figure 14 displays an ALV visualization that contrasts the search strategies of the high-performing and low-performing students in a class, displaying the top 10 scores on the content post-test (in green) and the lowest 10 scores (in pink).

Figure 13. Event paths in River City Session 3 for a three-person team.

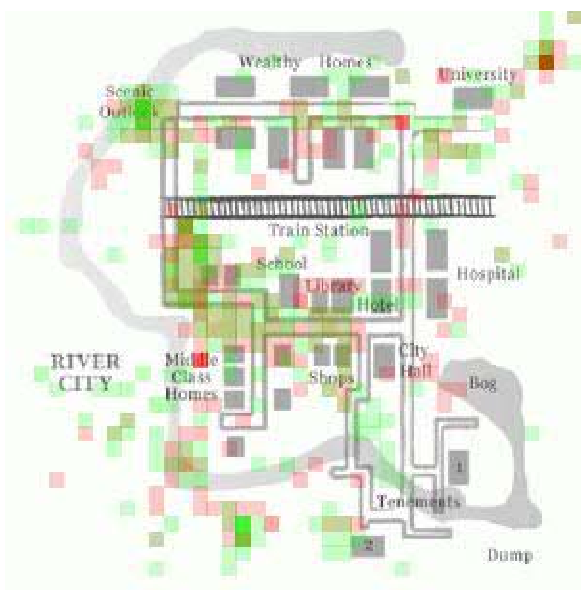

Figure 14. An avatar log visualizater (AVL) heat map showing high-performing and low-performing students in River City Session 4.

The high performing students' preferred locations provide an expert model usable in diagnostic feedback, formative about their search strategies, to students in subsequent classes. The low performing students' locations may offer insights into what types of understanding they lack.

Path analysis is a potentially powerful form of unobtrusive assessment, although choosing the best way to display student paths through a learning environment is a complex type of visualization not well understood at present (Dukas, 2009). The utility of this diagnostic approach also depends on the degree to which exploration in the virtual world is an important component of learning. In EcoMUVE, for example, while the forest module involves substantial exploration, only one role out of four team roles in the pond module involves searching the region. However, students can rotate this role from session to session so that all get this experience and benefit from diagnostic feedback. In EcoMOBILE, creating a heat map of which hotspots students prefer may be diagnostic about what they are not learning. We already have designed learning experiences so that students who have a misconception go to a different hotspot than those who are correct, thus generating a different path (and enabling teachers to see at a glance which students are confused). Also, we plan to have students creating their own ARs in a year or two, and assessing what they are highlighting—and missing—in their designs could provide useful feedback.

Accessing an individualized guidance system. As another example of the rich analytic power possible through the use of log files, Nelson (2007) developed a version of River City that contained an interwoven individualized guidance system (IGS). The guidance system utilized personalized interaction histories collected on each student's activities to generate real-time, customized support. The IGS offered reflective prompts about each student's learning in the world, with the content of the messages based on in-world events and basic event histories of that individual (Figure 15).

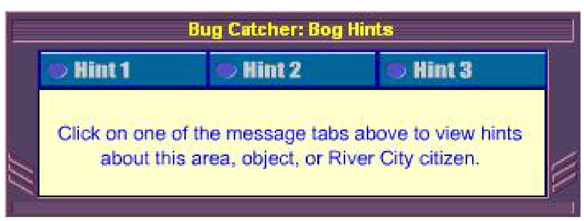

Figure 15. Individualized guidance system (IGS) interface.

As an example, if a student were to click on the admissions chart in the River City hospital, a predefined rule stated that if the student had previously visited the tenement district and talked to a resident there, then a customized guidance message would be shown reminding the student that he or she had previously visited the tenement district, and asking the student how many patients listed on the chart came from that part of town.

Multilevel multiple regression analysis findings showed that use of this guidance system with our MUVE-based curriculum had a statistically significant, positive impact ($p < .05$) on student learning (Nelson, 2007). In addition to using the log files to personalize the guidance provided to each student, we conducted analyses of guidance use. We knew when and if students first chose to use the guidance system, which messages they viewed, where they were in the virtual world when they viewed them, and what actions they took subsequent to viewing a given guidance message. This potentially provides diagnostic information that could guide instruction.

We could embed such an IGS system throughout the EcoMUVE; this would provide diagnostic insights comparable to those provided by any hint system. For EcoMOBILE, we could develop a similar system, but care is needed to ensure that the primary focus of students' attention does not shift from interacting with the real world to instead utilizing resources on the handheld device.

Asking questions of an "expert" agent. The APAs are "lifelike autonomous characters [that] co-habit learning environments with students to create rich, face-to-face learning interactions" (Johnson, Rickel, & Lester, 2000, p. 47). Bowman (2011) created Dr. C (Figure 16), an APA that served as a mentor for middle school students studying space science. Interwoven in a web browser, Dr. C was a computerized simulation of the primary scientist involved with NASA's Mars Student Imaging Project. One version of Dr. C provided both career (content-focused) and psychosocial (interpersonal-focused) mentoring; the latter meant Dr. C would answer some questions about the scientist's personal and professional life. The back end of the system was a large set of short responses to frequently asked questions, with a relatively simplistic pattern recognition system scanning the text of the student's question for words it recognized. A study revealed that the Dr. C template for expert mentoring was flexible, reliable, and engaging, distributing substantial amounts of content knowledge on an as-needed basis (Bowman, 2011).

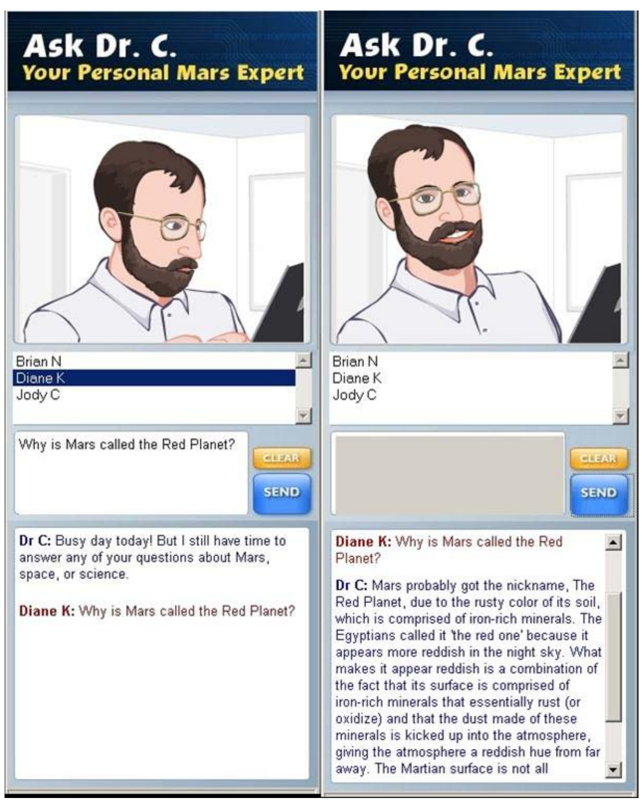

Figure 16. Dr. C interacting with a student.

Research suggests that APAs can fill various roles of mentorship, including expert, motivator, collaborator, and learning companion (Chou, Chan, & Lin, 2003). For example, Baylor and Kim (2005) created three versions of an APA: the Expert, designed as older than the participants, formal in appearance and language, and providing domain-specific information; the Motivator, casual in appearance and language, providing encouragement; and the Mentor, less formal than the Expert yet older than the Motivator, providing a mix of information and encouragement. The results from their study confirmed that the agent (APA) roles were not only perceived by the students to reflect their intended purposes, but also led to significant changes in learning and motivation as intended by their design. Specifically, the Expert agent (APA) led to increased information acquisition, the Motivator led to increased self-efficacy, and the Mentor led to overall improved learning and motivation (Baylor & Kim, 2005).

One can imagine tailoring a wide range of APAs to various student needs and embedding these in immersive learning environments (Dede, 2012). Beyond engaging students and providing a limited form of mentoring, APAs have advantages for interwoven diagnostic assessment in immersive authentic simulations in two respects: First, the questions students ask of an APA are themselves diagnostic—typically learners will ask for information they do not know, but see as having value. Sometimes a single question asked by a student of an APA may reveal as much about what that learner does and does not know than a series of answers the student provides to a teacher's diagnostic questions. Both EcoMUVE and EcoMOBILE could embed APAs of various types for eliciting a trajectory over time of questions that reveal aspects of students' understanding and motivation, as well as aiding learning and engagement by the APA's responses.

Second, APAs scattered through an immersive authentic simulation can draw out student performances in various ways. In EcoMUVE and EcoMOBILE, a student could meet an APA who requests the student's name and role. Even a simple pattern recognition system could determine if the student made a response indicating self-efficacy and motivation ("ecosystems scientist" or some variant) versus a response indicating lack of confidence or engagement ("sixth grader" or some other out-of-character reply). As another example, an APA could request a student to summarize what the student has found so far, and some form of latent semantic analysis could scan the response for key phrases indicating understanding of terminology and relevant concepts. The important design considerations of this method for evoking performances are that (a) the interaction is consistent with the overall narrative, so not too disruptive of flow, (b) the measurement is relatively unobtrusive, and (c) the interactions themselves deepen immersion.

Structured benchmark assessments that measure learning progress and scaffold transfer. Periodically, designers could structure brief benchmarking episodes into the immersive learning experience. In a virtual world, these might be the equivalent of the VPA assessments: shifting a student to a semantically identical but syntactically dissimilar environment in which the student must identify a problem (earlier in the curriculum) or resolve a problem (later in the curriculum) by performances based on current understandings. In EcoMUVE, this might involve teleporting to a pond with a superficially different problem that had similar underlying causality. In EcoMOBILE, a student might go into an area of the ecosystem without AR supports for learning and, absent that guidance, demonstrate the degree to which he or she can conduct various practices as an ecosystem scientist.

The ECD of this structured assessment would enable benchmarking to document a student's progress, without undercutting the open-ended nature of the overall learning experience. As another benefit, these benchmarking assessments could help to scaffold transfer. Further, by centering on learners' common misconceptions about that domain (e.g., in the case of ecosystems, action across distance and time, invisible causes, the flows of matter and of energy), immediately conveying the results of these benchmarks to students could prompt "aha" moments that help to synthesize new levels of understanding.

Reporting mechanisms consistent with the immersive simulation's narrative and flow. Game designers have developed a variety of reporting mechanisms that provide diagnostic feedback to participants, formative to improved player strategies, without disrupting flow or narrative (Gee, 2008). These include *dashboards* of various types that provide ongoing progress levels for key variables (e.g., health and stamina in an adventure game). In the case of EcoMUVE and EcoMOBILE, finding various species, classifying them along various dimensions (e.g., producers, consumers, decomposers), and adding those species to a food web that models energy flow all offer opportunities for dashboard variables important for learning the domain. Beyond providing ongoing feedback, this design strategy increases immersion and engagement—and learning, if the variables tracked correspond closely to instructional goals.

In River City, we went beyond dashboards to give students an opportunity parallel to leveling up in games: On each visit to River City, teams of students could attain new powers through reaching a threshold of experiences and accomplishments (Nelson et al., 2007). We centered our powers narrative on a specific location inside the world, a spooky building initially closed to all students. When a team of students achieved powers for a given session, they gained access to a new room in the house where they found a number of magical tools, such as a special interactive map that allowed students to check on the health of all residents of the city. The powers design approach provides a way of motivating a team's collective success on a variety of interwoven assessments, then uses that achievement to promote further learning with added opportunities for diagnostic inferences. EcoMUVE's and EcoMOBILE's designs could include domain-appropriate "powers" (e.g., in the field of view, highlighting interesting areas for special study) that use interwoven assessments to promote engagement and learning.

An analogy for the design of immersive diagnostic assessments. The list of design approaches above is illustrative rather than inclusive, but gives a sense what is possible in immersive interfaces. All of these types of assessment are based not on proxies (e.g., test items, essays) for real world performance, but instead authentic actions in rich simulated contexts. One way to understand the assessment strategy overarching these design approaches is to note that this paper begins by discussing the value of learning in authentic real-world contexts, such as an internship in a workplace setting. Suppose that internship was on an ecosystems science research team. One can imagine the mentor for the internship structuring that experience to elicit understandings and performances that provide diagnostic information: Is the learner:

- gaining in confidence and developing an identity as a scientist?
- developing intrinsic motivation about understanding and stewarding natural environments?
- accomplishing progressively less easy performances related to science inquiry, as well as a greater proportion of difficult tasks performed together by student and mentor?

- increasingly using disciplinary terminology and representations?
- progressively more able to communicate her insights, both informally to the general public and formally to the scientific community?

The mentor would continuously structure experiences to evoke these skills, but would do so with care to avoid disrupting the engagement and flow of the internship experience. Certainly, the mentor would not give periodic tests, nor would the intern be asked to write essay papers (although he or she might contribute to the team's research paper).

For a middle grades student in such an internship, the goal would be to reach some threshold of scientific understandings and performances, but beyond that threshold to focus the internship on intrinsic motivation, self-efficacy, and love of learning (Tai et al., 2006). These important outcomes are difficult to measure. Immersive authentic simulations are analogous to the internship; they are more than is needed for learning basic concepts and formulaic skills, but are a powerful method of scaffolding intellectual, emotional, and social capacity and commitment for future involvement (Dieterle, 2009), as well as for assessing students' progress towards these outcomes. The emphasis is not on achieving higher performance on standardized measures, but on preparing learners for life. As such, an immersive simulation interwoven with all the methods of diagnostic assessment described above would be inefficient if the metric is coverage of curricular standards—but then that is not how we judge the value of internships and similar real-world apprenticeship experiences.

In an instructional model oriented to presentational instruction with the goal of high achievement on tests and essays, immersive simulations may seem a needless extravagance. But if our ultimate goal is personal and collective success in the 21st century global, knowledge-based, innovation-centered economy, then instruction, curricula, and assessment should center on real and simulated authentic learning-by-doing experiences with just-in-time teaching-by-telling available to help learners surmount challenges as these are encountered through interwoven diagnostic assessments (Dede, 2011).

Next Steps in Attaining Immersive Assessments: Recommendations for the Field

The National Research Council (2011) report on learning science through computer games and simulations made several recommendations on a research agenda (pp. 126–127):

- Applications of the ECD approach to the development of assessments of learning through simulations and games. Developers and testing experts should collaborate to clearly identify desired learning goals and the kinds of evidence needed to show learner progress toward these goals; they should use these specifications to design tasks and test items in ways that will provide the needed evidence. Modeling of the motivation and thinking of the learner will need to evolve simultaneously with the "physical" modeling of the game or simulation.

  In addition to ECD, I would include design for ameliorating the challenges of interwoven assessment in open-ended learning environments.

- The development and use of flexible statistical models and machine learning to make meaning from the large amounts of data provided by simulations and games. These measurement methods are well suited to application in simulations and games because they can handle uncertainty about the current state of the learner, provide immediate feedback during tasks, and model complex patterns of student behavior and multiple forms of evidence. Continued research on these methods will help to improve assessment in simulations and games.

In addition to new types of mathematical models, I would include embedded measurement models analogous to those utilized by mentors in internship or apprenticeship settings.

- Researchers should continue to advance the design and use of techniques that (a) rapidly measure and adapt to students' progress in a specific learning progression, (b) dynamically respond to an individual student's performance, and (c) allow for the summative evaluation of how well students are learning.

A culminating achievement of this type of research would be to alter, in real time, the context and activities of the immersive simulation to make salient what the student needs to understand next in their learning trajectory.

Collectively, these research achievements could help to scaffold two Grand Challenges in the 2010 National Educational Technology Plan (U.S. Department of Education, 2010, p. 78):

> 1.0: Design and validate an integrated system that provides real-time access to learning experiences tuned to the levels of difficulty and assistance that optimize learning for all learners and that incorporates self-improving features that enable it to become increasingly effective through interaction with learners.
>
> 2.0: Design and validate an integrated system for designing and implementing valid, reliable, and cost-effective assessments of complex aspects of 21st-century expertise and competencies across academic disciplines.

These in turn could empower powerful digital teaching platforms that enable customizing classroom learning for each student (Dede & Richards, 2012).

Acknowledgments

The EcoMUVE and EcoMOBILE projects are supported by research grants from the Institute of Education Sciences (IES) of the U.S. Department of Education; the Qualcomm, Inc. Wireless Reach initiative; and the National Science Foundation. I also thank Texas Instruments and MoGo, Mobile, Inc. for resources and support. I have greatly benefited from the insights of my colleagues Dr. Tina Grotzer, Dr. Amy Kamarainen, Dr. Shari Metcalf, and Shane Tutwiler from the EcoMUVE/EcoMOBILE projects, as well as from the insights of Dr. Jody Clarke-Midura, Director of the Virtual Performance Assessment Project funded by IES and the Gates Foundation. In addition, I appreciate the feedback of Drs. Cassie Bowman, Ed Dieterle, Brian Nelson, and Diane Jass Ketelhut, all members of the River City project team. The viewpoints expressed in this article are mine and are not necessarily those of my collaborators or the funders.

References

Ballantyne, R., & Packer, J. (2002). Nature-based excursions: School students' perceptions of learning in natural environments. *International Research in Geographical and Environmental Education, 11*(3), 218–230.

Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher, 39*(7), 525–536.

Bartle, R. (2003). *Designing virtual worlds*. Berkeley, CA: Peachpit/New Riders Publishing.

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, *15*, 95–115.

Beal, C. R., & Stevens, R. H. (2007). Student motivation and performance in scientific problem solving. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 539–541). Amsterdam, Netherlands: IOS Press.

Behrens, J. T., Frezzo, D., Mislevy, R., Kroopnick, M., & Wise, D. (2007). Structural, functional, and semiotic symmetries in simulation-based games and assessments. In E. Baker, J. Dickieson, W. Wulfeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 59–80). Mahwah, NJ: Lawrence Erlbaum Associates.

Behrens, J. T., DiCerbo, K. E., & Ferrara, S. (2012, May). *Intended and unintended deceptions in the use of simulations.* Paper presented at the K-12 Center at ETS invitational research symposium on technology enhanced assessments. Retrieved from http://www.k12center.org/rsc/pdf/session2-pearson-behrens-dicerbo-ferrara-paper-tea2012.pdf

Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010).Measuring problem solving with technology: A demonstration study for NAEP. *Journal of Technology, Learning, and Assessment, 8*(8), 1–45.

Bowman, C. D. D. (2011). Student use of animated pedagogical agents in a middle school science inquiry program. *British Journal of Educational Technology, 43*(3). Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8535.2011.01198.x/pdf

Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist*, *34*, 75–85.

Brown, J., Hinze, S., & Pellegrino, J.W. (2008). Technology and formative assessment. In T. Good (Ed.), *21st century education*. Thousand Oaks, CA: Sage.

Chaiklin, S., & Lave, J. (1993). *Understanding practice: Perspectives on activity and context*. New York, NY: Cambridge University Press.

Chou, C., Chan, T., & Lin, C. (2003). Redefining the learning companion: The past, present, and future of educational agents. *Computers and Education*, *40*, 255–269.

Clarke, J. (2009). *Exploring the complexity of inquiry learning in an open-ended problem space* (Unpublished doctoral dissertation). Harvard Graduate School of Education, Cambridge, MA.

Clarke, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education*, *42*(3), 309–328.

Clarke-Midura, J., Dede, C., & Norton, J. (2011). Next generation assessments for measuring complex learning in science. In *The Road Ahead for State Assessments* (pp. 27–40). Cambridge MA: Rennie Center for Education and Public Policy. Retrieved from http://renniecenter.issuelab.org/research

Clarke-Midura, J, Mayrath, M., & Dede, C. (in press). Thinking outside the bubble: Virtual performance assessments for measuring complex learning. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. Charlotte, NC: Information Age Publishing.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (1st ed.) New York, NY: Harper Perennial.

Dawley, L. (2009). Social network knowledge construction: Emerging virtual world pedagogy. *On The Horizon*, *17*(2), 109–121.

Dawley, L., & Dede, C. (in press). Situated learning in virtual worlds and immersive simulations. In M. J. Bishop & J. Elen (Eds.), *Handbook of research on educational communications and technology* (Vol. 2., 4th ed.). New York, NY: Macmillan.

Dede, C. (2005). Why design-based research is both important and difficult. *Educational Technology*, *45*(1), 5–8.

Dede, C. (2008). Theoretical perspectives influencing the use of information technology in teaching and learning. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 43–62). New York, NY: Springer.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66–69.

Dede, C. (2011a). Reconceptualizing technology integration to meet the challenges of educational transformation. *Journal of Curriculum and Instruction, 5*(1), 4–16.

Dede, C. (2012). Customization in immersive learning environments: Implications for digital teaching platforms. In C. Dede & J. Richards (Eds.). *Digital teaching platforms: Customizing classroom learning for each student* (pp. 119–133). New York, NY: Teacher's College Press.

Dede, C., & Richards, J. (Eds.). (2012). *Digital teaching platforms: Customizing classroom learning for each student*. New York, NY: Teacher's College Press.

Dieterle, E. (2009). Neomillennial learning styles and River City. *Children, youth, and environments, 19*(1), 245–278.

Driscoll, M. (2000). *Psychology of learning for instruction*. Needham Heights, MA: Allyn & Bacon.

Dukas, G. (2009) *Characterizing student navigation in educational multiuser virtual environments: A case study using data from the River City project* (Unpublished doctoral dissertation). Harvard Graduate School of Education, Cambridge, MA.

Dunleavy, M., & Dede, C. (in press). Augmented reality teaching and learning. In M. J. Bishop & J. Elen (Eds.), *Handbook of research on educational communications and technology* (Vol. 2, 4th ed.). New York, NY: Macmillan.

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*(1), 7–22.

Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways). In J. Aronson (Ed.), *Improving academic achievement* (pp. 37–60). New York, NY: Academic Press.

Garner, L., & Gallo, M. (2005). Field trips and their effects on student achievement and attitudes: A comparison of physical versus virtual field trips to the Indian River lagoon. *Journal of College Science Teaching, 34*(5), 14–17.

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

Greeno, J. (1998). The situativity of knowing, learning, and research. *American Psychologist*, *53*, 5–26.

Grotzer, T. A. (2004, October). Putting everyday science within reach: Addressing patterns of thinking that limit science learning. *Principal Leadership*, 16–21.

Grotzer, T. A., & Basca, B. B. (2003). Helping students to grasp the underlying causal structures when learning about ecosystems: How does it impact understanding? *Journal of Biological Education*, *38*(1), 16–29.

Grotzer, T. A. Dede, C., Metcalf, S., & Clarke, J. (2009, April). *Addressing the challenges in understanding ecosystems: Why getting kids outside may not be enough.* Paper presented at the National Association of Research in Science Teaching (NARST) Conference, Orange Grove, CA.

Grotzer, T. A., Kamarainen, A., Tutwiler, M. S., Metcalf, S. J., & Dede, C. (2012, April). *Learning for focus on processes and steady states in ecosystems dynamics using a virtual environment*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

Habgood, M. P. J., & Ainsworth, S. E. (2011): Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games, *Journal of the Learning Sciences*, *20*(2), 169–206.

Hamilton, S. F., & Hamilton, M. A. (2006). School, work, and emerging adulthood. In J. J. Arnett & J. L. Tanner (Eds.), *Emerging adults in America: Coming of age in the 21st century* (pp. 257–277). Washington, DC: American Psychological Association.

Harrington, M. C. R. (2011). Empirical evidence of priming, transfer, reinforcement, and learning in the real and virtual trillium trails. *Learning Technologies, IEEE Transactions on Issue*, *4*(2), 175–186.

Heffernan, N. T., Heffernan, C. L., Decoteau, M., & Militello, M. (2012). Effective and meaningful use of educational technology: Three cases from the classroom. In C. Dede & J. Richards (Eds.), *Digital teaching platforms: Customizing classroom learning for each student* (pp. 88–102). New York, NY: Teachers College Press.

Heffernan, N. T., & Koedinger, K. R. (2012, May). *Integrating assessment within instruction: A look forward.* Paper presented at the K-12 Center at ETS invitational research symposium on technology enhanced assessments. Retrieved from http://www.k12center.org/rsc/pdf/session4-koedinger-paper-tea2012.pdf

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*, 47–78.

Kafai, Y.M., Quintero, M., & Felton, D (2010) Investigating the 'why' in Whypox: Casual and systematic explorations of a virtual epidemic. *Games & Culture 5*(1), 116–135**.**

Kamarainen, A., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M.S., & Dede, C. (2012). *EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips.* Manuscript submitted for publication.

Kamarainen, A., Metcalf, S., Tutwiler, S., Grotzer, T., & Dede, C. (2012, April). *EcoMUVE: Shifts in affective beliefs and values about science through learning experiences in immersive virtual environments.* Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

Ketelhut, D. J., Avirup, S., Yates, A., Shelton, A., Natarajan, U., Nelson, B., Schifter, C., & Karakus, M. (2012). *Insights into science learning using immersive environments as assessments: Data mining SAVE science.* Manuscript submitted for publication.

Ketelhut, D., Dede, C., Clarke, J., Nelson, B., & Bowman, C. (2007). Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 37–58). Mahwah, NJ: Erlbaum.

Ketelhut, D. J., Nelson, B. C., Clarke, J. E., & Dede, C. (2010). A multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*, *41*, 56–68.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. Cambridge, MA: MIT Press.

Klopfer, E., & Sheldon, J. (2010, Winter). Augmenting your own reality: Student authoring of science-based augmented reality games. *New Directions for Youth Development*, *128*, 85–94.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). New York, NY: Cambridge University Press.

Laurel, B. (1993). *Computers as theatre*. New York, NY: Addison-Wesley.

Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257–307). San Diego, CA: Academic Press.

Levy, R. (2012, May). *Psychometric advances, opportunities, and challenges for simulation-based assessment.* Paper presented at the K-12 Center at ETS invitational research symposium on technology enhanced assessments. Retrieved from http://www.k12center.org/rsc/pdf/session2-levy-paper-tea2012.pdf

Mestre, J. (2002). *Transfer of learning: Issues and a research agenda*. Arlington, VA: National Science Foundation.

Metcalf, S., Kamarainen, A., Grotzer, T., & Dede, C. (2012, April). *Teacher perceptions of the practicality and effectiveness of immersive ecological simulations as classroom curricula*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, Canada.

Metcalf, S., Kamarainen, A., Tutwiler, M. S., Grotzer, T., & Dede, C. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations, 3*(1), 86–90.

Metcalf, S. J., & Tinker, R. F. (2004). Probeware and handhelds in elementary and middle school science. *Journal of Science Education and Technology*, *13*(1), 43–49.

Mislevy, R., & Haertel, G. (2006). *Implications of evidence-centered design for educational testing* (Draft PADI Technical Report 17). Menlo Park, CA: SRI International.

Mislevy, R., & Rahman, T. (2009). *Design pattern for assessing cause and effect reasoning in reading comprehension* (PADI Technical Report 20). Menlo Park, CA: SRI International.

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

Murray, J. H. (1997). *Hamlet on the holodeck: The future of narrative in cyberspace*. New York, NY: Free Press.

National Committee on Science Education Standards and Assessment, National Research Council. (1996). *National science education standards*. Washington, DC: National Academies Press.

National Research Council. (2011). *Learning science through computer games and simulations*. Washington, DC: National Academies Press.

Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology 16*(1), 83–97.

Nelson, B., Ketelhut, D. J., Clarke, J., Dieterle, E., Dede, C., & Erlandson, B. (2007). Robust design strategies for scaling educational innovations: The River City MUVE case study. In B. E. Shelton & D. A. Wiley, *The design and use of simulation computer games in education* (pp. 219–242). Rotterdam, Netherlands: Sense Press.

Nicolaou, C. T., Nicolaidou, I. A., Zacharia, Z. C., & Constantinou, C. P. (2007). Enhancing fourth graders' ability to interpret graphical representations through the use of microcomputer-based labs implemented within an inquiry-based activity sequence. *Journal of Computers in Mathematics and Science Teaching*, *26*(1), 75.

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of Research in Science Teaching, 31*(10), 1097–1119.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, *49*, 345–375.

Przybylski, A. K, Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154–166.

Quellmalz, E. S., & Haertel, G. (2004). *Technology supports for state science assessment systems*. Paper commissioned by the National Research Council Committee on Test Design for K-12 Science Achievement. Washington, DC: National Research Council.

Quellmalz, E. S., Timms, M. J., & Schneider, S. A. (2009). *Assessment of student learning in science simulations and games*. Paper prepared for the National Research Council Workshop on Gaming and Simulations. Washington, DC: National Research Council.

Rogers, Y., & Price, S. (2008). The role of mobile devices in facilitating collaborative inquiry in situ. *Research and Practice in Technology Enhanced Learning*, *3*(3), 209.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–118). New York, NY: Cambridge University Press.

Schunk, D. H., & Pajares, F. (2005). Competence beliefs in academic functioning. In A. J. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 85–104). New York, NY: Guilford Press.

Schwartz, D. L., Sears, D., & Bransford, J. D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 1–51). Greenwich, CT: Information Age

Shaffer, D. W., & Gee, J. P. (2007). Epistemic games as education for innovation. In J. D. M. Underwood & J. Dockrell (Eds.), *Learning through digital technologies* (pp. 71–82). Leicester, UK: British Journal of Educational Psychology.

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *The social science of serious games: Theories and applications* (pp. 295–321). Philadelphia, PA: Routledge.

Slator, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 3549–3557.

Squire, K. (2010). From information to experience: Place-based augmented reality games as a model for learning in a globally networked society. *Teachers College Record*, *112*(10), 2565–2602.

Squire, K. D., Jan, M., Matthews, J., Wagler, M., Martin, J., Devane, B., & Holden, C. (2007). Wherever you go, there you are: The design of local games for learning. In B. Sheldon & D. Wiley (Eds), *The design and use of simulation computer games in education* (pp. 265–296). Rotterdam, Netherlands: Sense Publishing.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science*, *312*(5777), 1143-1144.

Turkle, S. (1997). *Life on the screen: Identity in the age of the internet*. New York, NY: Simon & Schuster.

Turkle, S. (2005). *The second self: Computers and the human spirit* (20th anniversary ed.). Cambridge, MA: MIT Press.

U.S. Department of Education. (2010). *Transforming American education: Learning powered by technology* (National Educational Technology Plan 2010). Washington, DC: Office of Educational Technology, U.S. Department of Education. Retrieved from http://www.ed.gov/technology/netp-2010

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education 16*(3), 227–265.

Vonderwell, S., Sparrow, K., & Zachariah, S. (2005). Using handheld computers and probeware in inquiry-based science education. *Journal of the Research Center for Educational Technology*, *1*(2), 1–11.

Wagner, A. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need--and what we can do about it*. New York, NY: Basic Books.

Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement, 48*(2), 165–187.

Warburton, S. (2009). Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology, 40*(3), 414–426.

Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. New York. NY: Cambridge University Press.

Williamson, D. M., Bejar, I. I., & Mislevy, R. J. (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah, NJ: Erlbaum.

Winn, W. (2003). Learning in artificial environments: Embodiment, interwovenness, and dynamic adaptation. *Technology, Instruction, Cognition, and Learning*, *1*, 87–114.

K-12 Center at ETS

Invitational Research Symposium on

Technology Enhanced Assessments

The Center for K–12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place, and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state and local decision makers.