

Chapter 10

Assessing and Evaluating Virtual World Effectiveness

In November 2007, U.S. Secretary of Education, Margaret Spellings, suggested that U.S. colleges and universities implement a standardized test not unlike the K-12 *No Child Left Behind Act*. Measures need to be made to see which institutions are assessing and evaluating their innovation and courses in general (Tyre, 2007). As educational institutions become more data-driven and accountability is increasingly more important, assessment and evaluation of courses is crucial. As it pertains to online delivery of courses, especially those in 3D virtual worlds, many are shying away from assessment and evaluation primarily because it is not an easy venture. Since integrating 3D worlds in education is an emerging idea, something my former college football coach reiterated time again comes to mind. He said, “Prior Planning Prevents Poor Performance.” As we begin to develop or further develop these environments for education purposes, planning how the courses will be assessed and evaluated prior to going live can prevent potential missing data that could help improve courses and meet the requirements for institutional accountability.

The no significant difference phenomenon of the 1990s in studies comparing online courses to their face-to-face counterparts has driven many to design within group studies that lacked some of the rigor to really show the relative power of online education. One important reason for the incoherent findings in online environments is that methodological flaws in the study designs often do not allow a rejection of the null hypothesis of no difference. In 1999, the “Institute for Higher Education Policy” (Phipps & Merisotis, 1999) pointed out that the majority of all published studies comparing online distance education with classroom instruction had serious methodological flaws and poor study designs. Randolph (2007) re-examined these studies and came to a similar conclusion. Most of the studies were quantitative–descriptive, qualitative–descriptive, or correlative studies in which participants were not randomly selected, extraneous variables or feelings and attitudes of students (reactive effects) not controlled for, or the validity and reliability of the measures not reported. Bernard et al. (2004) found that methodological and experimental differences (including inadequacies and missing information) explained a large amount of the reported variation in the research literature. Dwyer, Millet, & Payne (2006) proposed a comprehensive national system for determining the nature and extent of college learning, focusing on four dimensions of student learning:

- Workplace readiness and general skills
- Domain-specific knowledge and skills
- Soft skills, such as teamwork, communication, and creativity
- Student engagement with learning

Since this book is grounded in game-based learning and 3D online worlds, it has never been a better time to infuse methodologies commonly used outside of education. This chapter will look at how we employed some of these methodologies and how we used the power of available technologies to harness our data collection processes. If the hallmark of games is their interactivity, their ability to grant players agency within the narrative fiction of the game world and its rules, the theoretical models need to account for players' action in creating the experience (Squire, 2006). Relatively few evaluation studies have been conducted on the use of computers in education and on the learning outcomes of the different modes of educational software (Presby, 2001).

In 1991, Brant, Hooper, and Sugrue examined the effectiveness of computer simulations based upon their placement within a larger sequence of instruction. Their design involved the stratified random sampling of 101 college students from an introductory animal science course. Participants were in one of three treatments: one experimental group of students ($n = 34$) that solved computer simulation problems *before* a classroom lecture on the topic; a second experimental group ($n = 32$) that worked on the simulation problems *after* a lecture; and a control group ($n = 35$) that was not exposed to the simulation, receiving only a lecture on the topic. Using a 17-item post-test that assessed students' ability to apply genetics principles to solve breeding problems as their dependent measure, they found that the effectiveness of a simulation is influenced by its placement in the instructional sequence. That is, the group that experienced the simulation prior to the lecture significantly outscored the untreated control group on the genetics test (effect size = 0.91) as did the group that engaged in the simulations after the lecture, but the magnitude of this difference was smaller (effect size = 0.36) (Brant, Hooper, & Sugrue, 1991). In another study, Carlsen and Andre (1992) introduced students ($n = 83$) to a simulation about electrical circuits that was combined with either a traditional text or a conceptual change/refutation text. The treatments were presentation of the simulation before the text, simultaneous with the text, or no simulation. The main cognitive measure was a post-test consisting of 26 items designed to assess participants' conceptualizations about series circuits. It was found that simulation groups' scores were not significantly different than those of the no simulation group but the authors assert that evidence existed that the mental models of the simulation group participants were more advanced.

Guided by the five-step process first elucidated by Heck, Steigelbauer, Hall, and Loucks (1981) we will (1) *identity innovation components* such as teacher behaviors, student activities, or ways innovation resources and materials are used; (2) *identify additional components and variations* that constitute variations of implementation that range from ideal use to unacceptable use for each component; (3) *refine the innovation components* as part of our research plan outlined above;

(4) *finalize the innovation components* as we construct a component checklist consisting of innovation components and a set of variations within each component that is field-tested with a small group of innovation users; (5) *collect innovation data* as we administer the checklist in written or interview format to our innovation users and analyze data in order to determine prevailing innovation configuration patterns.

Given the likelihood of a mixed early reaction to the general concept of postsecondary education assessments, an incremental approach to implementation may be appropriate for initial consideration. Here are several related issues for consideration:

- Regarding assessment development, the options range from having one organization develop and test the needed assessments to the clearly less desirable option (from the point of view of comparability and efficiency) of having each of the 4,071 institutions develop its own assessments.
- The outcomes associated with successful performance on the different dimensions of student learning could vary. For example, mastery of work-readiness skills could lead to a certificate, while performance on domain areas could be tied to a new valuation of the bachelor's degree.
- Performance indicators could be developed for individuals, institutions, or both.
- The number of students taking the assessment could range from all students in higher education to a sample from each institution.
- The number of times that students take the assessments could range from one to multiple times. Several key questions may guide the expert panel as it considers where on the different continua it wishes to place its marks:
- Should there be individual scores? Would this help future employers and graduate and professional schools know more about the inputs into their systems? How should this consideration be balanced with the cost savings of a sampling approach?
- Should there be institutional scores? Would an institutional score help both prospective students and their families have a more informed sense of what the educational experience will be like? What would an institutional score signal to employers and graduate and professional schools about their graduates?
- What should the rollout plan be for the new postsecondary education system? Should a demonstration program be conducted, while plans for a longer-term nationwide system are developed?
- What are the desired types of analyses – pre-/post-test, individual growth models, value-added analyses? Each of these analyses has important data thresholds that need to be met.

Since much of our work has been supported by the National Science Foundation in the United States, it is only fitting we share the Foundation's Division of Research on Learning's Cycle of Innovation and Learning as a framework from which we have operated (see Fig. 10.1).

These five steps to design, implementation, and evaluate an innovation clearly lead to synthesizing lines of work and study new ideas and questions posed by the

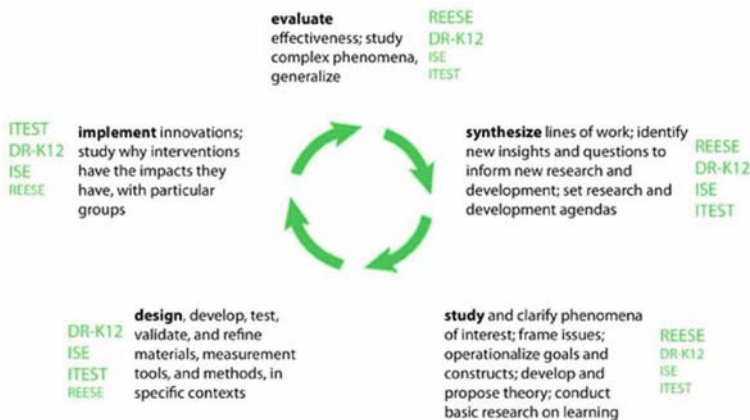


Fig. 10.1 DRL cycle of innovation and learning (Note: Programs whose primary emphases relate to particular components appear in larger type.)

implementation. The model is truly a cycle in that new ideas and research questions facilitate new designs, implementation, and evaluations. As it pertains to our work, we have gone through this cycle numerous times and thus new iterations of our software and courses have evolved tremendously.

Entitled “Being Fluent with Information Technology,” a National Research Council Committee (1999) acknowledged tendencies to focus on skills when approaching technology literacy. The report explained that literacy today requires a complement of knowledge and related abilities to be fluent in information technology (FIT). Much of this report aligns with what we suggested in Chapter 1 of this book on 21st Century Skills. According to the report, FITness is a long-term process of self-expression, reformulation, and synthesis of knowledge in three realms: “Contemporary skills, the ability to use today’s computer applications, enable people to apply information technology immediately . . . are an essential component of job readiness . . . [and] provide . . . practical experience on which to build new competence. Foundational concepts, the principles and ideas of computers, networks, and information, underpin the technology . . . explain the how and why of information technology . . . give insight into its limitations and opportunities . . . [and] are the raw material for understanding new information technology as it evolves. Intellectual capabilities, the ability to apply information technology in complex and sustained situations, encapsulate higher-level thinking in the context of information technology . . . empowers people to manipulate media to their advantage and to handle unintended and unexpected problems when they arise . . . [and] foster more abstract thinking about information and its manipulation.”

The report offers an intellectual framework that can help distinguish between achievements (those of a particular time) and learning outcomes (results over time) when assessing what competencies students need to have. The proposed framework might also help differentiate among research (of teaching and learning theories),

evaluation (of learning programs and processes), and assessment (of learning outcomes) as scholars and their audiences seek to show who and what measure up or make the grade. Although the specific skills for each area will change with the technology, the concepts are rooted in the basic information and abilities required to function in technology-enabled environments.

What follows is an example evaluation plan we designed in conjunction with colleagues at Information In Place, Inc. on a National Science Foundation-funded project where we are creating training simulations for prospective science teachers. *STIMULATE* (*Science Training Immersive Modules for University Learning Around Teacher Education*) seeks to use Serious Game technology to train prospective science teachers in laboratory safety and managing a safe classroom environment. These simulations are immersive and take a first-person perspective not unlike training simulations used by the military and medical fields.

First, each class was randomly assigned into one of the two treatment groups. Pre-tests were given to all participants one week before the intervention began. Once the intervention period began, treatment group #1 played three interactive *STIMULATE* game modules over a period of six weeks. They had access to the game during non-class time. At the same time, treatment group #2 received a written case study scenario that was the same as the ones used in *STIMULATE*, and their interactivity was through classroom analysis and discussion of the case-based reasoning approaches. At the end of each game session and in both treatments, the professor led the class in a whole group discussion of an after action review analysis focusing on decisions made, evidence supporting those decisions, and a discussion of specific domain-specific content addressed in each scenario. These after action reviews were videotaped so that individual classroom interactions could be analyzed in more detail.

One week after the intervention, both groups completed post-tests. One week after the post-test, individual semi-structured interviews were conducted with eight students and four professors to better understand student and classroom specific patterns and implementation issues.

Simulation and Game Design

We find that the model presented by Garris, Ahlers, and Driskell (2002) and (Fig. 10.2) helps us to articulate how the prior work done in the areas of computer-based instruction, inquiry-based science, and learner-centered design amalgamate to inform our work. This model involves the design of computer-based instructional program that incorporates instructional content and certain features games. They suggest that the six key dimensions that characterize games include the following: fantasy, rules/goals, sensory stimuli, challenge, mystery, and control. Next, they assert that the combination of instructional content and game characteristics initiates a *game cycle* that involves user judgments or reactions (such as enjoyment or interest), user behaviors (such as greater persistence or time on task), and system

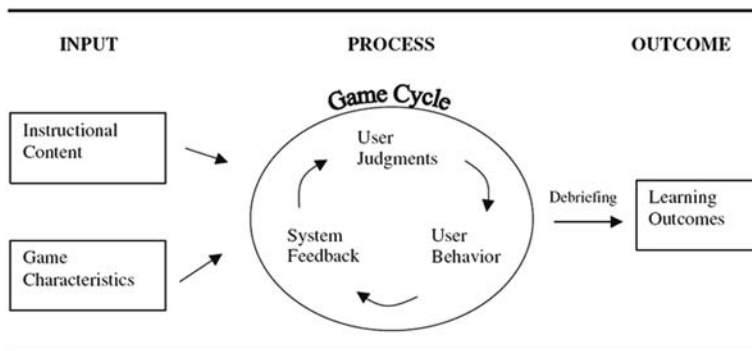


Fig. 10.2 Simulation design from Garris, Ahlers, and Driskell (2002)

feedback. During this cycle, users are actively constructing knowledge from their experiences within the virtual world in which they are immersed. This model also includes a *debriefing* phase that serves to provide a critical link between the game cycle and the achievement of the desired learning outcomes. This debriefing often includes the review and analysis of events that occurred in the game itself (Garris, Ahlers, & Driskell, 2002), what we called the *after action review*.

Dondi & Moretti (2007) nicely categorized learning objectives, required features, game typology, and possible number of players. Table 10.1 briefly illustrates this categorization.

Design-Based Research

Researchers working in these areas are helping to chart the way by identifying best practices in commercial and educational game design that are also consistent with both cognitive and constructivist learning theories. Many of our projects engage in “Design-Based Research” (Squire & Barab, 2004, Brown, 1992; Cobb, Confrey, deSessa, Lehrer, & Schauble, 2003; Design-Based Research Collective, 2003), the results of which are then integrated into the three dimensions of the “contextual model of learning” in free-choice environments as posited by Falk and Dierking (2000). We believe our work is beginning to establish an international dialogue among educators as to how game-based learning can most effectively reflect and inform the personal, physical, and socio-cultural contexts of free-choice learning. Although not fully embraced by the research community, particularly those who advocate for randomized controlled trials in education, we feel that this research paradigm is highly appropriate for this innovation.

We generally engage in two cycles of design, development, enactment, analysis, redesign, and refinement of our intervention in order to generate design knowledge and build theory. Our studies employ a *concurrent triangulation* research design. This mixed-methods strategy utilizes both quantitative and qualitative data in an

Table 10.1 Dondi's learning game categorization

| Learning objective(s) | Definition | Required features | Typology | Number of players |
|--|---|--|--|--|
| Applying concepts and rules | Translate knowledge into new contexts | Presents of a set of rules and instructions, possibility of applying the rules to different contexts, balance between reality and abstraction, chance as a factor, finite scenarios in game, documentation of the scenario, real-time monitoring, time considerations, open-endedness, session based | Sport games, action-adventure games, driving games, drill-and-practice games | Single player vs. the PC |
| Decision-making (strategy + problem solving) | Knowledge analysis based on problem solving, prediction, drawing conclusions, choice making, reasoned argumentation | Narrative based, chance as a factor, finite scenarios in game, documentation of the scenario, real-time monitoring, time considerations, considering background knowledge | Strategy games, adventure games, role play games, simulations | Single player in connection with other players/opponents |

Table 10.1 (continued)

| Learning objective(s) | Definition | Required features | Typology | Number of players |
|-----------------------------------|--|---|--|-------------------|
| Social interaction/value/cultures | Understanding the social environment of others | Chance is not a factor, involves deep negotiation or coalition skills, reflection as a factor, persistent state game, Narrative based, presence, time is not a factor, difficulty determined by the abilities of more than one player, relevant feedback, | Strategy games, role play games, simulations | Multiple players |
| Ability to learn/self-assessment | Evaluation | Available evaluation tools, engine provides different scenarios each time, relevant documentation, time is not a factor, presence, after-action review, reflections, chance is not a factor, balance between action and reflection | Role play games and simulations | Single player |

attempt to confirm, cross validate, or corroborate findings within a single study. We implement the quantitative and qualitative methods and measures during each of the “testing cycles” and with equal weight to obtain different but complementary data regarding our interventions (Creswell, 2003; Creswell, Plano Clark, Gutmann, & Hanson, 2003).

With a focus on linking processes to outcomes in particular settings, this iterative process requires the collection and coordination of a complex array of data sources including video and audiotapes, student work, classroom observations, responses to interviews, and formative test results (Cobb, Confrey, deSessa, Lehrer, & Schauble, 2003).

The multiple sources of qualitative data generally emerge from our studies that are analyzed according to standard procedures for qualitative analysis (e.g., Coffey & Atkinson, 1996; Erickson, 1992; LeCompte, Millroy, & Preissle, 1992), with each data source analyzed slightly differently based on the type of data it yields and the purposes of the data.

Testing the Intervention

Through systematic feasibility and usability studies of successive versions of our interventions, we collect data that are used to inform and guide the creation and refinements of our program prototypes. What follows is a comprehensive description of our data sources, potential measures, and how we use the information generated.

We propose that good design-based research exhibits the following five characteristics: First, the central goals of designing learning environments and developing theories or “prototheories” of learning are intertwined. Second, development and research take place through continuous cycles of design, enactment, analysis, and redesign (Cobb, 2001; Collins, 1992). Third, research on designs must lead to sharable theories that help communicate relevant implications to practitioners and other educational designers (Brophy, 1998). Fourth, research must account for how designs function in authentic settings. It must not only document success or failure, but also focus on interactions that refine our understanding of the learning issues involved. Fifth, the development of such accounts relies on methods that can document and connect processes of enactment to outcomes of interest.

To better understand the importance of integrating design-based research, it is important to clarify the distinction between existing methods for understanding learning and cognition, and those central to design-based research. Collins, Joseph, and Bielaczyc (2004) contrast several different methodologies with design-based research. They posit seven major differences between traditional psychological methods and the design-experiment methodology. Barab, MaKinster, & Scheckler (2004) summarized this notion by stating, “Central to this distinction is that design-based research focuses on understanding the entropy of real-world practice, with context being a core part of the story and not an extraneous variable to be trivialized. Further, design-based research involves flexible design revision, multiple

Table 10.2 Comparison of psychological experimentation vs. design-based research

| Category | Psychological experimentation | Design-based research |
|------------------------------|--|--|
| Location of research | Conducted in a laboratory | Occurs in real-life settings where most learning takes place |
| Complexity of variables | Involves a single or a few dependent variables | Involves multiple dependent variables including climate variables (e.g., collaboration among learners, available resources), outcome variables (e.g., learning of context and transfer), system variables (e.g., dissemination and sustainability) |
| Focuses of research | Involves identifying a few variables and holding them constant | Involves characterizing the situation in all its complexity |
| Unfolding of procedures | Uses fixed procedures | Involves a flexible design revision in which there is a tentative initial set that is revised depending on the success in practice |
| Amount of social interaction | Isolates the learner to control interaction | Involves complex social interactions with participants sharing ideas, distractions, etc. |
| Characterizing the findings | Focuses on testing the hypothesis | Involves looking at multiple aspects of the design and developing a profile that characterizes the design in practice |
| Role of the participants | Participants as subjects | Involves different participants in the design so to bring different expertise into producing and analyzing the design |

dependent variables, and capturing social interaction. In addition, participants are not ‘subjects’ assigned to treatments but instead are treated as co-participants in both the design and even the analysis. Last, given the focus on characterizing situations (as opposed to controlling variables), the focus of design-based research may be on developing a profile or theory that characterizes the design in practice (as opposed to simply testing hypotheses).” Table 10.2 shows Barab’s comparison of psychological experimentation versus design-based research.

Finally, we would like to include the characteristics of design-based research proposed by Wang & Hannafin (2005) as another way of illustrating the power of this paradigm (Table 10.3).

Assessment Techniques

In an attempt to depict how the features and components of a project developed by my colleague, Dr. James Minogue, are related to resources, activities, and outcomes, we used a logic model to guide our design (Fig. 10.3). Although most logic

Table 10.3 Wang’s design-based research characteristics

| Characteristics | Explanations |
|--------------------------------------|---|
| Pragmatic | Research refines both theory and practice, the value of theory is appraised by the extent to which principles inform and improve practice |
| Grounded | The design in theory driven and grounded in relevant research, theory and practice, the design is conducted in real-world settings and the design process is embedded in and studied through design-based research |
| Interactive, iterative, and flexible | Designers are involved in the design processes and work together with participants, processes are iterative cycles of analysis, design, implementation, and redesign, the initial plans is usually insufficiently detailed so that the designers can make deliberate changes when necessary |
| Integrative | Mixed research methods are used to maximize the credibility of ongoing research, methods vary during the different phases as new needs and issues emerge and the focus of the research evolves, rigor is purposely maintained and discipline applied appropriate to the developmental phase |
| Contextual | The research process, findings, and changes are well documented, results are connected with the design process and the setting, the content and depth of generated design principles vary, and the guidance for applying generated principles is needed |

models include short-, intermediate-, and long-term outcomes, given the focus on development, we felt that the identification of long-term outcomes would be a bit premature.

This example depicts how a project employs design-based research with the inclusion of experts. The logic model is a good way to illustrate the design process to easily organize one’s thoughts on the initial design phase. The model can be changed as the iterative design process unfolds.

Usability/Feasibility

In software development, usability and feasibility are two very important concepts by which the design process is informed. It is critical to be sure the software, or in this case the virtual learning environment and simulations, is understood by the end users and that it can be sustained as technology evolves.

We have attacked this issue by collecting data through remote (e.g., simulation back end, videoconference, telecommunications) access and face-to-face cognition interviews. Convenience sampling is generalized and used because it is increasingly difficult to stratify participants from a distance. These participants are asked pointed questions focusing on how the environment is used, how decisions are made in world, and how content is understood as it relates to real-world scenarios.

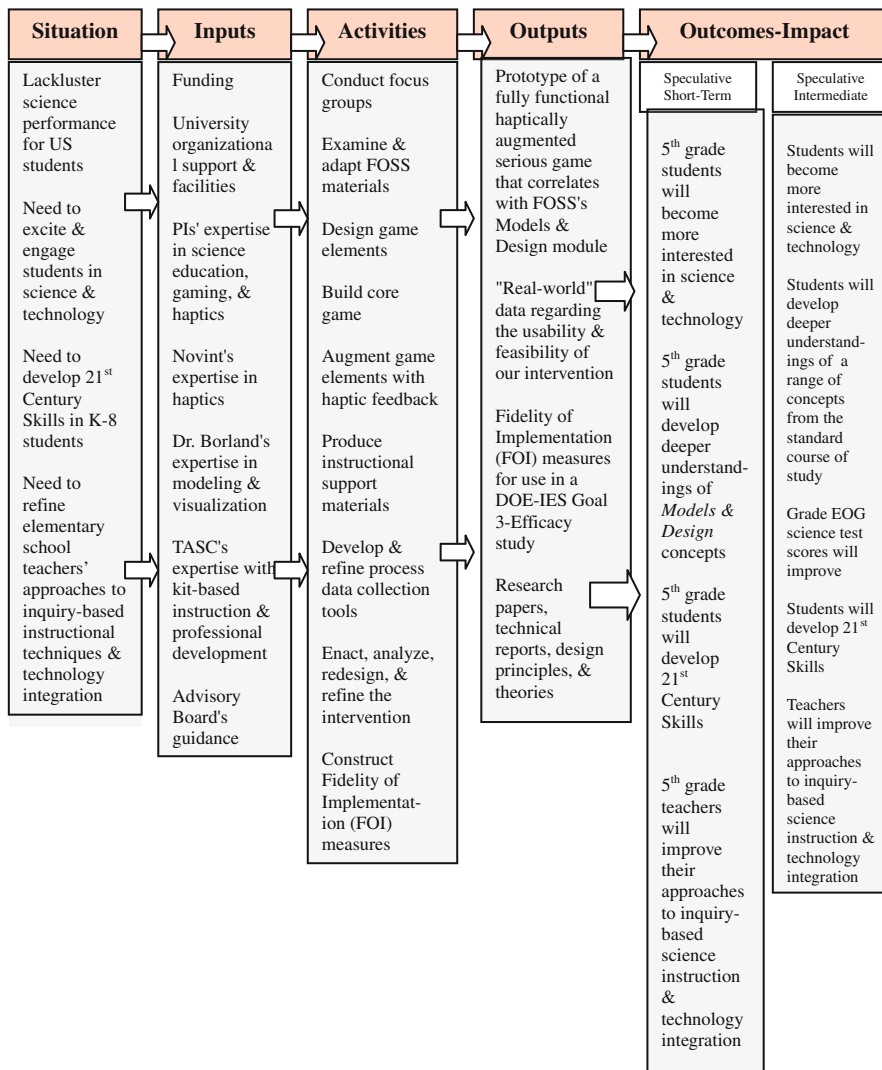


Fig. 10.3 A logic model for the ASPECT project

The multiple sources of qualitative data that emerge from this technique are analyzed according to standard procedures for qualitative analysis (e.g., Coffey & Atkinson, 1996; Erickson, 1992; LeCompte, Millroy, & Preissle, 1992), with each data source analyzed slightly differently based on the type of data it yields and the purposes of the data.

The purpose of this phase of the research is to determine the perceived effectiveness of the proposed design of environment or simulation scenarios as well as to use the outcomes to further improve the design of the product. In this first phase of the research, we use methods of qualitative naturalistic inquiry (Lincoln & Guba,

1985) with a focus on learner-centered design (Quintana et al., 2004) and participatory design (Schuler & Namioka, 1993) approaches. The team creates a preliminary design document, which provides written descriptions and storyboards of the key scenarios. We then use methods of rapid prototyping (Tripp & Bichelmeyer, 1990) which enable the team to “test” the ideas with potential users of the environment or simulation in order to obtain early feedback to improve designs as well as to inform the overall effort of what issues arise with regard to designing this type of environment for this type of audience.

To understand the perceived effectiveness of design outcomes, we provide surveys and conduct focus groups. Surveys are used to collect demographic data on participants as well as to respond to 5–10 questions related to the design ideas. Focus group discussions are then held to specifically examine how well the scenarios potentially impact usability and feasibility of the proposed audience.

Usability Data

We are equally interested in gaining insight into the usability of our intervention. Thus, another key component of our research plan involves the collection, analysis, and careful application of *usability data*. Following the design-based research model, we regularly collect and analyze multiple sources of data. These sources include the following: (a) classroom observation protocols; (b) videotapes of testing sessions; (c) student think alouds; (d) student questionnaires; (e) students and teacher interviews; and (f) formative knowledge assessments and attitudinal assessments.

Classroom observations. Adopting ethnographic techniques, we become part of the user community and make careful observations of our intervention in use. These focused observations of our test sessions require the development of classroom observation protocols and coding schemes. One such instrument we tend to use is the *Science Teacher Inquiry Rubric* (STIR) (Beerer & Bodzin, 2003). This instrument allows us to quantify and plot classroom activities along an inquiry continuum. Comparing pre-intervention and post-intervention activities helps us assess whether or not our intervention is actually promoting inquiry as we intend it to.

Videotapes of testing sessions. When usability issues exist, participants often hesitate, struggle, and/or become frustrated. Thus, content analysis and resulting codings of user’s speech and actions during the videotaped testing sessions likely yield critical information about a wide range of factors including workflow, navigation, and terminology.

Think alouds. A researcher from the team also works individually with one student during each of the testing sessions. This researcher has the student user engage in concurrent think aloud strategies in an attempt to gain insight into how students process information as they engage in our environments and simulation. Users are asked to verbalize their actions, perceptions, and expectations regarding the application’s interface and functionality (Dumas & Redish, 1999; Ericsson & Simon, 1993).

Questionnaires. Although limited somewhat by the relatively low number of students involved, it is still expected that this approach will generate valuable data regarding users' level of comprehension of our program's purpose and functionality, initial expectations of where features are located within a system's interface, and reactions to the visual design of an interface.

As part of the collection of usability data, we also develop and administer open-ended and Likert-scale questionnaires to all student users. Open-ended prompts may include the following: *What do you like best about the instructional program? What do you like least about the instructional program? What aspects of the instructional program would you like the designers to change? How?* Again, through our work with Dr. James Minogue, his instrument, *AIM* – a solo taxonomy, prompts participants to answer questions with regard to knowledge gain and transfer. Written responses to the open-ended items are coded and trends are identified. The quantitative results of the Likert-scale items are analyzed descriptively and both data sources are fed into the analysis, redesign, and refinement cycle.

Interviews. We also engage in retrospective probing (Wickens & Hollands, 2000) of the student users. Semi-structured interviews of a randomly chosen sub-sample of students are conducted immediately after they have completed a task or series of tasks with our intervention. Designed to reveal the users' memories of their experiences, responses highlight major usability concerns or issues that are prominent in the users' minds.

It is expected that this approach will generate valuable data regarding users' level of comprehension of our program's purpose and functionality, initial expectations of where features are located within a system's interface, and reactions to the visual design of an interface. These interviews are often audiotaped, transcribed, coded, and analyzed in an effort to further isolate areas of strength and weakness regarding the usability of our intervention.

Usability data are also garnered from the potential teacher participants. Given that they will be integral to the design, development, testing, and refinement process, it is equally critical to tap into their observations and feelings regarding the usability of each iteration of the interventions. Straight forward and important questions such as *Does the software program crash when students use it? Are the activities planned for a particular lesson do-able within the allotted time?* are asked. Again, the results of such sessions are recorded and its content analyzed to inform subsequent "design and test" efforts.

Feasibility Data

Early on in the project we conduct focus groups with diverse groups of participants from the targeted audience. The focus of these sessions is to document the viability of integrating our environment or simulations in authentic education delivery settings. Here we operationally define *diversity* as potential students with varying age, gender, race, years of online learning experience, and their reported use of

technology. Each of these focus groups is videotaped or screen recorded to allow for subsequent analysis.

Additional critical insights into the feasibility of our intervention in our particular setting are gained throughout the development and testing phases. Although much of this data collection is gathered informally, this information constitutes a key piece of the feasibility studies.

In addition to the above described focus group sessions and informal conversations with the participants, feasibility data are collected via survey instruments. Due to the 3D, game-like nature of our environments and simulation, we like to use the *Self-Efficacy in Technology and Science* (SETS) (Ketelhut, 2005). This instrument focuses on efficacy as it pertains to science as inquiry and common informal technology uses such as video games, online chat, etc. Analysis of this survey data is descriptive in nature and we look for relationships between specific items/topics and characteristics of respondents in order to better assess the technical, organizational, and cultural feasibility of our intervention.

Recognizing the importance of administrative support in the ultimate success of educational innovations, we also interview district level officials and school level administrators. Through our content analysis, we posit these sessions will highlight any potential barriers (be they logistical, financial, or philosophical) to the implementation of our program, as well as gauge the level of support for a larger scale implementation in the future.

In short, through these activities, we aim to accurately assess the pedagogical feasibility, management feasibility, economic feasibility, and client acceptability of our computer-based instructional program and this information will help inform the initial design of our intervention.

Server-Side Data Collection

During each testing session, we use *Just-In-Time* (JIT) analysis so that we can record technical problems, immediately generate a prioritized master list of problems, and fix as many as possible on the spot. If problems are not remedied immediately, we use affinity analysis in which each problem is written on a sticky note, notes are placed on a wall or board, notes (problems) are grouped into emergent categories, and assigned a priority and fixed. This process represents a critical component of any development project. We must not lose sight of the fact that we are ultimately attempting to design and build an intervention that is *likely to produce better student outcomes relative to current education practices*.

One of the many assessment and evaluation components to integrating virtual learning environments in distance learning is their ability to incorporate data tracking, analytics, and bots in what I have called *virtual observations*. Often in educational settings we record classes to ascertain what works and what needs more refinement. However, when teaching from a distance, this technique becomes difficult, especially in asynchronous learning. In our environments, and as part of

the design process, we create tracking systems to help us analyze data stored on servers.

Data are collected electronically using a customized tracking system. We include such variables as unique user logins (demographics), time stamps, patterns of use and interaction, chat logs, and in-world decisions (especially in simulations, field trips and labs). When students first log in to the virtual learning environment, they receive a tracking code and each decision they make as they navigate through the environment or simulation is recorded in the tracking system. Most often analyzed is each user's time stamp and chat logs in the multi-user environment. The chat logs tend to serve as an ill-structured *think aloud*. To analyze this data, we conduct several readings of the whole transcripts from the chat logs. Then we use Miles and Huberman (1984) "concurrent flows" of analyses approach to data analysis. This approach has the following phases: (1) data reduction, the transformation of raw data, and decision-making regarding data "chunking" (2) data display, the assembling of information into displays such as matrices, graphs, and charts (3) conclusion drawing, with notation of "regularities, patterns, explanations, possible configurations, causal flows, and propositions." It is important to note that this data is analyzed for specific content, such as focusing on text relating to a particular theme. These processes in the data can only be identified by several readings of the whole transcript and tracing an individual's text in the context of other participants' text.

As it pertains to simulations, in-world decisions and patterns of use become critical variables for which data mining techniques can be used. According to Ian Ayres, an econometrician and law professor at Yale, data mining analytics is a microcosm of a powerful trend that will shape the economy for years to come. He states that these data are the replacement of expertise and intuition by objective, data-based decision making made possible by a virtually inexhaustible supply of inexpensive information. Ayres calls those who use and manipulate these data streams as *Super Crunchers*, which is also the title of his book. Ayres continues by stating that Super-crunchable data can be broadly statistical or profoundly personal.

In a study in which we partnered with Dr. Chris Dede, Harvard University Professor of Learning Technologies, his doctoral student Geordie Dukas, and SAS®, we used data mining techniques to gain insight into server-side data potential as an emerging form of education analytics. In one online simulation where Algebraic concepts were being taught, server-side data were used to create a visual display of pattern tracking and in-world decisions made by students engaged in the simulation over the course of a semester.

Figure 10.4 illustrates the simulation map: the level and models of the simulation. The ultimate goal of this simulation was for each student to save the high school by defeating a witch who challenged the students to high level algebra questions.

Figure 10.5 shows the overall student decisions in this world. Note, the darker the shaded area, the more decisions made in this simulation and vice versa. The areas marked in red illustrate male decisions and those marked in blue show female decisions. There are three avenues in which a student could win this game (denoted by the "InRadius OfWitch" in the upper left of the map). Students could climb the walls (shown on the left side of the map) while answering questions of increasing

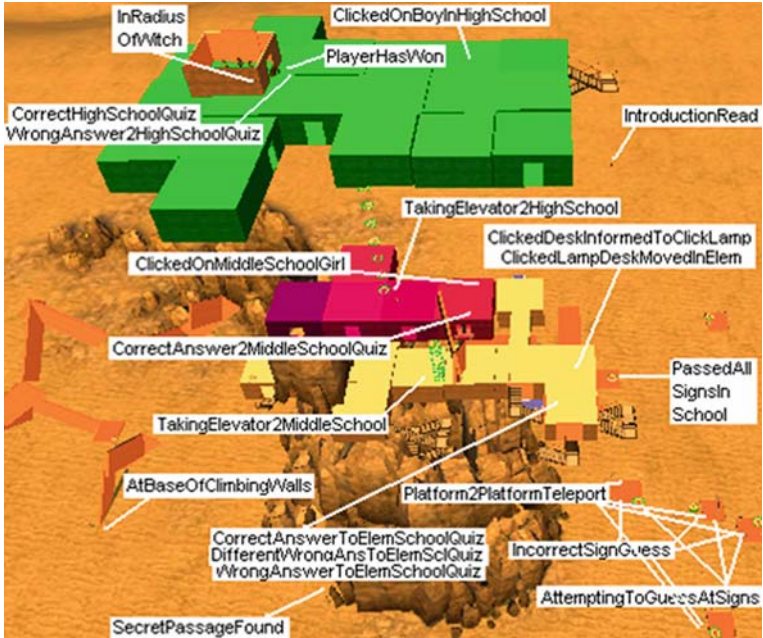


Fig. 10.4 Aerial view of algebra simulation

difficulty as they progress up the walls and through elementary, middle, and finally high school. Secondly, students could answer some high order questions that would give them a secret code that would move them to the top level. The secret code/high order question region is shown in the lower right side of the map. Finally, a student could find the secret passage to the top level. The passage is found in the lower middle of the map. What we see in Fig. 10.5 is that students spent the lion's share of their time attempting to answer the high order questions that unlocked the secret

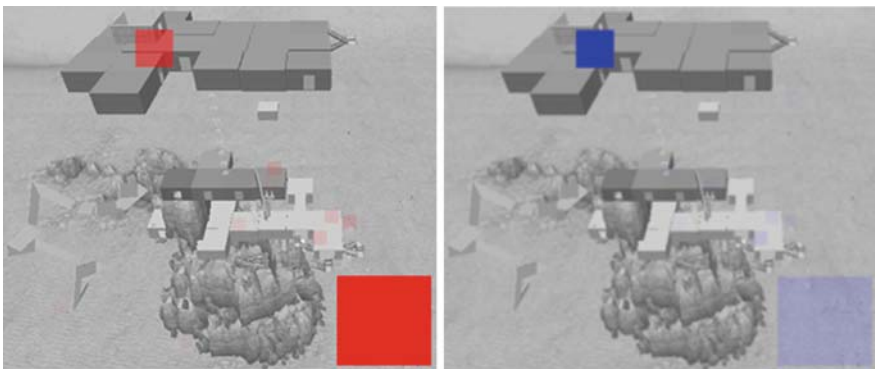


Fig. 10.5 Overall gender decisions in the algebra game

code. Interestingly enough, males spent more time in this area suggesting they had more incorrect guesses than females. Because the blue square in the upper left is darker for females, we can see that females also defeated the witch and won the game more often than males.

If we are to break these decisions down by time, another interesting event occurs. In the first 3:00 min from login, males not only had more guesses in the secret code area, but also there were some decisions being made to find the secret passage (Fig. 10.6). The secret passage is equivalent to what commercial games call *cheats*. It is an easy path to the final level. The developer of this simulation was curious if anyone would look for a cheat rather than answer the questions that were designed to teach the Algebraic concepts. Moreover, the lightly shaded red squares (other than those in the secret passage region) suggest that males not only read the game instructions, but also answered more questions through the game progression than did females. What we don't know is whether or not males just wanted to "play" the game by finding all of the game's triggers or they just could not answer the high order questions in the lower right that unlocked the secret code.

Finally, Fig. 10.7 shows that all females finished the game within 6:00–9:00 min from login. It also suggests that males were either progressing through the game at a normal rate or still trying to guess the code. Some males had finished within this time period but most have not.

These illustrations begin to shed light on the developmental aspects of the simulation and provide critical user data as to how to refine the simulation. Even through regular observational methods, the coding and analysis would create time constraints usually not worth the effort by some researchers. However, using this tracking technique ensures an immediate visual output that provides the needed data for refining the intervention.

From these images, we can work backward to discuss how the data were collected and stored in world. Figure 10.8 shows an example of a True/False type item that is built into the system. Based on the answer given by the student, the responses are numerically coded and sent to the server.

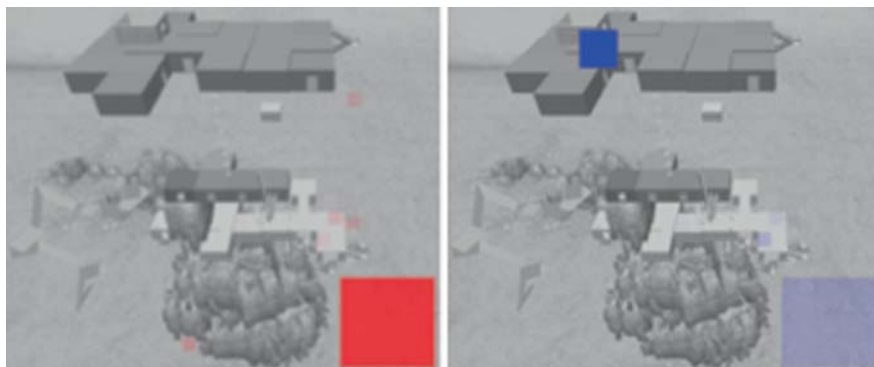


Fig. 10.6 In-world decision in the first 3:00 min

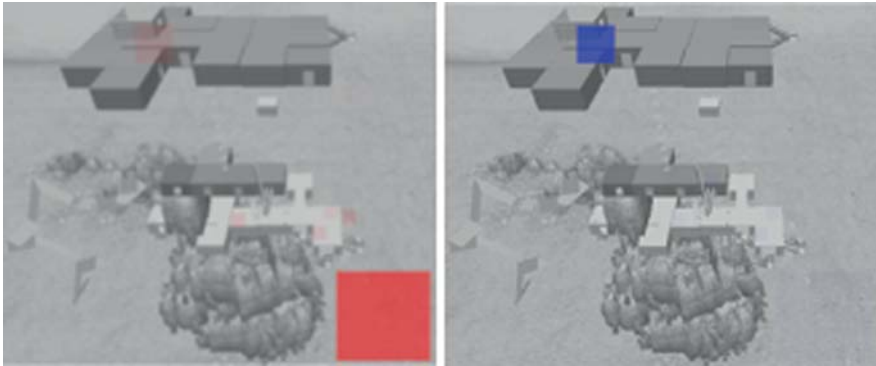


Fig. 10.7 In-game decisions from 6:00 to 9:00 min from login



Fig. 10.8 In-world assessment example

Cognitive Ethnography and Discourse Analysis

For the qualitative readers, Steinkuehler (2006) introduced the concept of cognitive ethnography (Hutchins, 1995): a “thick description” (Geertz, 1973) of the socially and materially distributed cognitive practices that constitute MMOs. She reports that the proper unit of study for work on cognition is not the individual “head” but rather the intact interactional structures of social and material activity. In most ethnographies, the researcher participates overtly, observing what goes on within the

virtual world, taking digital video recording and field notes, listening to what is said, asking questions, and generally “collecting whatever data are available to throw light on the issues that are the focus of the research”. “From these data, patterns of routine cognitive/cultural activities can be discerned. Meaning is therefore not individual but rather it is embedded in the history and social practices of the group” (Gee, 1999, p. 105) answers to the remaining research questions, such as what and where learning occurs and what it means for the identity of participants in the gaming culture, which are inaccessible without such groundwork.

In addition to routine observation and field notes, participants/students of varying ages, ethnicities, socio-economic statuses, and levels of expertise/social status within the community are recruited and interviewed repeatedly in unstructured (e.g., informal conversation within the virtual worlds), semi-structured (e.g., telephone/Skype interviews about particular topics of interest), and structured (e.g., repertory grid interviews, Fransella & Bannister, 1977) formats. Finally, chat logs form are also collected in order to capture virtual world actions.

Further, discourse analysis can take the chat logs and answer research questions beyond the scope of cognitive ethnography. Gee (1999, pp. 4–5) defined *discourse analysis* as “the analysis of language as it is used to enact activities, perspectives, and identities.” Understanding which and how particular social and material practices mark membership in the MMO communities and how participation in those practices shape, and are shaped by, participants’ identities within and beyond the game, requires understanding the *situated meanings* individuals construct (not just the information they process), the definitive *role of communities* in that meaning, and the inherently *ideological nature* of both. Coming out of the New Literacy Studies (e.g., Gumperz, 1982; Halliday, 1978; Kress, 1985; Street, 1984), d/Discourse theory (Gee, 1999) provides a way to maintain the Learning Sciences’ focus on intact interactional structures, while, at the same time, foregrounding the role of d/Discourse (language-in-use/“kinds of people”) in such interactions.

Such analyses focus on the configurations of linguistic cues used in spoken or written utterances in order to invite certain interpretive practices. Configurations of such devices signal how the language of the particular utterance is being used to construe reality in terms of the following: (1) *semiotics*, what symbol systems are privileged, how they construe the relevant context (the world), and on what epistemological basis; (2) *the material world*, what objects, places, times, and people are relevant and in what way; (3) *socio-cultural reality*, who is who and what their relationships with one another are, including the implied identity of the speaker/writer and who the audience is construed to be, all in terms of affect, status, solidarity, and (shared or disparate) values and knowledge; (4) *activities*, what specific social activities the speaker and her interlocutors are taken to be engaged in; (5) *politics*, what social goods are at stake and how they are and “ought” to be distributed; and finally (6) *coherence*, what past and future interactions are relevant to the current communication (Gee, 1999). Through microanalysis of how group members’ utterances construe the world in particular ways and not others, we are able to infer the cultural models and concomitant Discourse(s) as play. With such analyses comes explication of the full range of social and material practices with which they are

inextricably linked, since the meaning of those practices is done with and through language in-use. Through such discourse-analysis-based ethnographic work, then, we capture the sense human beings make of the social and material world and their (inter)action with it – in other words, we finally get at the phenomenon of cognition itself, in all its unbounded, situated, distributed, social, and ideological messiness.

Heuristics

Heuristics are yet another way to measure variables in virtual worlds. Five heuristics – interactive creativity; selection hierarchy; identity construction; rewards and costs; and artistic forms – form is the structural basis of Web-based communities according to Gallant et al. (2007). The heuristics were developed using a threefold process. First, they examined past research and developed a 10-item list of elements they deemed essential to online communities. Second, they ran a content analysis of written responses from 18 participants. Third, they investigated how the 10 items related to the participants' use of Web-based communities. This analysis produced the five heuristics of Web-based communities. Finally, they tested these five heuristics on three focus groups with participants who are heavy users of two popular Web-based communities: Facebook and MySpace. The five heuristics of facilitating social usability for Web-based communities were verified in the empirical analysis.

Engagement

Engagement is one of the key indicators of learning as we point out many times throughout this book. However, measuring engagement is not an easy task. In 2003, Elaine Chapman summarized successful techniques in assessing online engagement. She explained that a few studies have used summative rating scales to measure student engagement levels. Summarizing her work, she points to studies done outside of the electronic medium but studies that can be applied to online learning. Teacher report scales used by Skinner and Belmont (1993) and Skinner, Wellborn, and Connell (1990) asked teachers to assess their students' willingness to participate in certain school tasks such as effort, attention, and persistence during the initiation and execution of learning activities. They also delved into their emotional reactions to the aforementioned tasks (i.e., interest vs. boredom, happiness vs. sadness, anxiety, and anger, such as "When in class, this student seems happy"). The *Teacher Questionnaire on Student Motivation to Read* developed by Sweet, Guthrie, and Ng (1996) also asked teachers to report on factors relating to student engagement rates. These activities (e.g., enjoys reading about favorite activities), autonomy (e.g., knows how to choose a book he or she would want to read), and individual factors (e.g., is easily distracted while reading) were targeted in their analyses.

Triangulating data sources is a key to ensuring reliability. In online virtual worlds, it is difficult to observe students, especially if the course is delivered asynchronously. A number of established protocols are available for observations when the virtual observations (as previously mentioned) are not available (e.g., Ellett & Chauvin, 1991; Ysseldyke & Christenson, 1993; Greenwood & Delquadri, 1988). The *CISSAR (Code for Instructional Structure and Student Academic Response: Greenwood & Delquadri, 1988)*, for example, defines *engagement* in terms of behaviors such as attending (e.g., reading from the blackboard), working (e.g., reading aloud/silently), and resource management (e.g., looking for materials). Clearly these actions are nearly impossible to observe unless the online course closely mimics the seated classroom. What is of critical importance is observer agreement as it pertains to scoring observational protocols. Inter-rater reliability on a near-point scale provides reliability measures that validate scores from two or more different observers of the same actions. Near-point ratings account for observer agreement on a ± 1 regardless of the protocol used. This is why a common protocol is important and that the observers are properly trained on the use of the specified protocol.

You might ask how one observes classroom engagement in 3D virtual worlds? The answer is twofold. First in synchronous meetings, software packages such as *Camtasia* or the open-source equivalent *Camstudio* could be used to video capture the entire class. Just as one might review a videotaped class at a later time to view and score student dynamics, these captured videos can be saved electronically and opened on the computer. In what I have called *Virtual Observations*, the server-side data mentioned earlier can code student dynamics in real time and store the information to be mined at a later time. Thus, you don't need to video capture classes or if the class has an asynchronous component, the researcher can "observe" student engagement by looking at the mined data stored on the server.

Lessons Learned for Future Growth

This project really sheds light on how to better prepare for game and simulation development. Specifically, we learned that gender is an important variable when designing questions and understanding how males and females spend their time in an educational game/simulation. The visual model created at Harvard is a nice substitute to conventional data mining software and techniques and it provides researchers with an idea of how recorded events are effecting in-world decisions. What this model does not tell us is why these decisions were made. Future research on how to better establish an analytic model and how that model influences game and simulation architecture is sorely needed.

As described in the National Research Council report, *Knowing What Students Know* (Pellegrino, Chudowsky & Glaser, 2001), sophisticated educational media now enable the collection of very rich data streams about individual learners. As previously mentioned, each participant's utterances, interactions, and movements

in a digital educational setting are automatically time-stamped and archived in a relational database. Analyzing these rich data streams can potentially yield the following:

- Formative, diagnostic information that provides real-time feedback to teachers on which kinds of students are most at risk in a particular learning situation and what types of immediate assistance to use for each (Feng & Heffernan, 2005);
- Summative assessments about what each student has mastered, based on authentic performances, are a richer, more accurate assessment of educational outcomes than are standardized pre/post measures (Hulshof, Wilhelm, Beishuizen, & Van Rijn, 2005);
- Insights about complex patterns and dynamics of student behavior and learning related to individual characteristics such as gender, native language, and prior educational preparation (Ketelhut, Dede, Clarke, Nelson, & Bowman, in press);
- A better understanding of collaborative problem solving and team learning processes (Avouris, Margaritis, & Komis, 2004; Linton, Goodman, Gaimari, Zarrella, & Ross, 2003; Suthers & Hundhausen, 2003); and
- Insights about the microgenetics of learning by examining patterns and relationships between students' behavioral patterns and learning outcomes.

Through the use of real-time intelligent agents (virtual world-based characters programmed to respond to user actions) coupled with data mining (Seydim, 1999), this could eventually provide the basis for real-time analysis identifying comparable paths of students currently in the 3D virtual environments.

Kennerly (2003) proposed a sequence of assessing actions and mining data in game-based environments. The 6-part sequence states

1. Live: Scoop up lots of raw data in the live service.
2. Archive: From here, clean it up and store it for safekeeping in an archive.
3. Statistics: Sift through the data to create statistics, which are more informative than the raw data.
4. Analysis: Then apply the actual mining, which yields knowledge about player performance.
5. Hypothesis: Propose hypotheses about how to tune the game.
6. Test: Test each hypothesis and then introduce the new design into the live service.

The final step closes the loop.

Kennerly further proposes an alternative method to cleaning data taken from the server. Here is a simple method that economizes storage space and reduces mining computation.

This preprocess has five general steps:

1. Take a snapshot of the database.
2. Validate that the data is clean and appropriate for analysis.
3. Integrate the data into a central archive.

4. Reduce the data down to just the fields you need.
5. Transform the reduced data into a form that is easy to analyze for player performance.

Conclusion

In conclusion, assessing virtual environments is a new and critical avenue for future research and evaluation. It is important to constantly assess the effectiveness of our teaching – and this has never been as important as now – and when you create and teach in a new setting. This chapter provides some insight into how you may assess your courses in 3D virtual environments so that the data can inform practice.

References

- Avouris, N., Margaritis, M., & Komis, V. (2004, September). *Modelling interaction during small-group synchronous problem-solving activities: The Synergo approach*. 2nd International Workshop on Designing Computational Models of Collaborative Learning Interaction, 7th Conference on Intelligent Tutoring Systems, Maceio, Brasil.
- Barab, S. A., MaKinster, J. G., & Scheckler, R. (2004). Designing system dualities: Characterizing online community. In R. K. J. G. S. A. Barab (Ed.), *Designing for virtual communities in the service of learning*. Cambridge: Cambridge University Press.
- Beerer, K., & Bodzin, A. (2003). Promoting inquiry-based science instruction: The validation of the Science Teacher Inquiry Rubric (STIR). *Journal of Elementary Science Education, 15*(2), 39–49.
- Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*, 379–439.
- Brant, G., Hooper, E., & Sugrue, B. (1991). Which comes first: The simulation or the lecture? *Journal of Educational Computing Research, 7*(4), 469–481.
- Brophy, S. P. (1998). *Sequencing problem solving and hands on activities: Does it matter?* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141–178.
- Carlsen, D. D., & Andre, T. (1992). Use of a microcomputer simulation and conceptual change text to overcome students' preconceptions about electric circuits. *Journal of Computer-Based Instruction, 19*, 105–109.
- Chapman, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation, 8*(13), Retrieved August 17, 2006, from <http://PAREonline.net/getvn.asp?v=8&n=13>.
- Cobb, P. (2001). Supporting the improvement of learning and teaching in social and institutional context. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: 25 years of progress* (pp. 455–478). Cambridge, MA: Lawrence Erlbaum Associates, Inc.
- Cobb, P., Confrey, J., deSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.
- Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data analysis: Complementary strategies*. Thousand Oaks CA: Sage.

- Collins, A. (1992). Towards a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*, 13(1), 15–42.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). London: Sage.
- Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.
- Dondi, C., & Moretti, M. (2007). A methodological proposal for learning games selection and quality assessment. *British Journal of Educational Technology*, 38(3), 502–512.
- Dumas, J. S., & Redish, J. C. (1999). *A practical guide to usability testing* (2nd ed.) Bristol, England: Intellect.
- Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes*. Princeton, NJ: ETS.
- Ellett, C. D., & Chauvin, E. (1991). Development, validity, and reliability of a new generation of assessments of effective teaching and learning: Future directions for the study of learning environments. *Journal of Classroom Interaction*, 26(2), 25–36.
- Erickson, F. (1992). Ethnographic microanalysis of interaction. In M. D. LeCompte, W. L. Millroy, & J. Preissle (Eds.), *The handbook of qualitative research in education* (pp. 202–225). San Diego: Academic Press, Inc.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised edition). Cambridge, MA: The MIT Press.
- Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. American Association for State and Local History Book Series. Walnut Creek, CA: AltaMira Press.
- Feng, M., & Heffernan, N. T. (2005). *Informing teachers live about student learning: Reporting in the ASSISTment system*. In workshop on usage analysis in Learning Systems at 12th Annual Conference on Artificial Intelligence in Education, Amsterdam.
- Fransella, F., & Bannister, D. (1977). *A manual for repertory grid technique*. London, New York, San Francisco: Academic Press.
- Gallant, L. M., Boone, G. M., & Heap, A. (2007). Five heuristics for designing and evaluating Web-based communities. *First Monday*, 12(3).
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation and Gaming*, 33(4), 441–467.
- Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. New York: Routledge.
- Geertz, C. (1973). Thick Description: Toward an interpretive theory of culture. In: Geertz, C. (Ed.), *The interpretation of cultures: Selected essays* (pp. 3–30). New York: Basic Books.
- Greenwood, C. R., & Delquadri, J. (1988). Code for instructional structure and student academic response (CISSAR). In M. Hersen & A. S. Bellack (Eds.), *Dictionary of behavioural assessment* (pp. 120–122). New York: Pergamon.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Sydney: Edward Arnold.
- Heck, S., Steigelbauer, S. M., Hall, G. E., & Loucks, S. F. (1981). *Measuring innovation configurations: Procedures and applications*. Austin, TX: Research and Development Center for Teacher Education, University of Texas.
- Hulshof, C. D., Wilhelm, P., Beishuizen, J. J., & Van Rijn, H. (2005). FILE: A tool for the study of inquiry learning. *Computers in Human Behavior*, 21, 945–956.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Kennerly, D. (2003). *Better game design through data mining*. Gamasutra.com.

- Ketelhut, D. (2005, April). *Assessing science self-efficacy in a virtual environment: A measurement pilot*. Paper presented at the National Association of Research in Science Teaching Conference, Dallas.
- Kress, G. (1985). *Linguistic processes in sociocultural practice. ECS806 Socio-cultural aspects of language and education*. Victoria: Deakin University.
- LeCompte, M. D., Millroy, W., & Preissle, J. (Eds.). (1992). *The handbook of qualitative research in education* (881 pp). San Diego, CA: Academic Press.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park: Sage.
- Linton, F., Goodman, B., Gaimari, R., Zarella, J., & Ross, H. (2003). *Student modeling for an intelligent agent in a collaborative learning environment* (pp. 342–351). Proceedings of User Modeling 2003, Johnstown, PA.
- Miles, M. M., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Newbury Park, CA: Sage.
- National Research Council Committee on Information Technology Literacy. (1999). *Being fluent with information technology*. Washington, DC: National Academy Press.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Phipps, R., & Merisotis, J. (1999). What's the difference? A review of contemporary research on the effectiveness of distance education in higher education. *Journal of Distance Education, 14*, 102–114.
- Presby, L. (2001). Increasing productivity in course delivery. *Technologic Horizons in Education Journal, 28*(7).
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*(3), 337–386.
- Randolph, R. (2007). What's the difference still? A follow up methodological review of the distance education research. *Information in Education, 6*, 179–188.
- Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Seydim, A. Y. (1999). *Intelligent agents: A data mining perspective*. Dallas, TX 75275: Dept. of Computer Science and Engineering, Southern Methodist University.
- Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*(4), 571–581.
- Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: The role of perceived control in children's engagement and school achievement. *Journal of Educational Psychology, 82*, 22–32.
- Squire, K. D. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*(8), 19–29.
- Squire, K., & Barab, S. (2004, June). *Replaying history: Engaging urban underserved students in learning world history through computer simulation games* (pp. 505–512). Proceedings of the 2004 International Conference of the Learning Sciences. Santa Monica, CA.
- Steinkuehler, C. A. (2006) *Massively multiplayer online videogaming as participation in a discourse, mind, culture, and activity* (pp. 38–52). Cambridge, MA: The MIT Press.
- Street, B. (1984). *Literacy in theory and practice*. New York: Cambridge University Press.
- Suthers, D., & Hundhausen, C. (2003). An empirical study of the effects of representational guidance on collaborative learning. *Journal of the Learning Sciences, 12*(2), 183–219.
- Sweet, A. P., Guthrie, J. T., & Ng, M. (1996). *Teacher perceptions and students' motivation to read* (Reading Research Report No. 69). Athens, GA: National Reading Research Center.
- Tripp, S. C., & Bichelmeyer, B. (1990). Rapid prototyping: An alternative instructional design strategy. *ETR & D, 38*(1), 31–44.
- Tyre, P. (2007, November 16). Standardized tests in college?. *Newsweek*. Retrieved from <http://www.newsweek.com/id/70750>

- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research & Development*, 53(4), 5–23.
- Wickens, C., & Hollands, J. (2000). *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.
- Ysseldyke, J., & Christenson, S. (1993). *TIES-II: The Instructional Environment System-II. A system to identify a student's instructional needs*. Eric Document No. 367072.