Chapter 11
Alternative Assessment Strategies for Complex Problem Solving in Game-Based Learning Environments

Deniz Eseryel, Dirk Ifenthaler, and Xun Ge

Abstract The central thesis of this chapter is that emerging technologies such as digital games compel educators, educational researchers, and instructional designers to conceptualize learning, instruction, and assessment in fundamentally different ways. New technologies, including massively multi-player digital games offer new opportunities for learning and instruction; however, there is as yet insufficient evidence to support sustained impact on learning and instruction, apart from the case of military training based on large simulated war games. Technologically sophisticated design and assessment frameworks are likely to facilitate progress in this area, and that is our focus in this chapter. Specifically, we provide an integrated framework for assessing complex problem solving in digital game-based learning in the context of a longitudinal design-based research study.

Keywords Game-based learning · Assessment · Complex problem-solving

1 Introduction

In the mid-1990s, Koschmann (1996) identified computer-supported collaborative learning as a new paradigm in instructional technology, characterized by collaborative learning and integrating social issues into the foreground as a central phenomena for study, including understanding language, culture, and other aspects of the social setting surrounding the learning experience. There continues to be new technologies that offer new possibilities for computer-supported learning and instruction. For instance, massively multiplayer online games (MMOGs) combine the power of traditional forms of role-playing games with a rich, 2- or 3-D graphical simulated world allowing individuals to interact with both the

D. Eseryel (✉)
College of Education, University of Oklahoma, 820 Van Vleet Oval, Norman, OK 73019, USA
e-mail: eseryel@ou.edu

D. Ifenthaler et al. (eds.), *Multiple Perspectives on Problem Solving and Learning in the Digital Age*, DOI 10.1007/978-1-4419-7612-3_11,

simulated environment and with other game players through their digital characters or avatars. These games enable players to cooperate and compete with each other on a grand scale while solving complex problem scenarios in the game. The emergence of game spaces provides players with new opportunities for learning (Thomas & Brown, 2007). Thus, MMOGs have been receiving increasing attention from educational researchers because of their motivational power as well as their potential in promoting complex problem solving and critical thinking skills.

With the increasing popularity of games and the widespread use of emerging technologies, Koschmann's (1996) work reminds us that we should focus not only on descriptive aspects of applications involving new technologies but that we ought to pay attention to paradigm shifts and innovations in both theory and practice. Accordingly, we should ask such questions as these in addition to conducting descriptive research: (a) What does game-based learning in the digital age mean? (b) What knowledge and skills should students learn in game-based learning? (c) How should we design game-based learning environments to facilitate intended knowledge and skill acquisition? These questions are the motivation behind this chapter.

The purpose of this chapter is two-folds. First, we elaborate on the new perspectives of learning and instruction proposed by the proponents of digital game-based learning. This first section includes review of empirical studies in peer-reviewed journals to arrive at an instructional design framework for effective learning in digital game-based learning environments. Emerging computer-based technologies such as digital games enable the creation of much-needed learning laboratories to systematically study the requirements and strategies for designing learning environments that can facilitate desired higher-order learning outcomes. However, as our literature review points out, lack of validated methodologies for assessing complex problem solving processes and outcomes is what impedes progress in this regard.

Therefore, the second main goal of this chapter is to propose an integrated assessment framework to guide the research and practice on digital game-based learning. We also illustrate how this integrated assessment framework was used to help us collect and analyze data in the context of a longitudinal study, the results of which have provided us with valuable information about the feasibility and value of various instructional design strategies in a MMOG–based learning environment.

2 Game-Based Learning in the Digital Age

Let my playing be my learning and my learning be my playing.
– Johan Huizinga

It took several millennia for games to evolve from one being played in a playground to one being played in a virtual world. It has taken only a couple decades for virtual games to progress from moving dots and lines (e.g., *Pong*) to massively multiplayer online role-play games with 3-D graphical avatars (e.g., *World of Warcraft*). At one time, the term 'games' referred largely to games played in a video arcade. However, in today's context, the term 'games' may refer to different types of games, including digital games playable on a portable computer.

In some sectors, the term 'serious games' (Michael & Chen, 2006) or 'epistemic games' (Shaffer, 2006) are used to distinguish games that are designed for training and instruction from those developed for entertainment purposes. However, the concept of serious or epistemic games is not new. Elementary and secondary teachers are no strangers to using games in the classroom. For a long time, board games, card games, and role-playing games have been commonly used within a classroom setting as a teaching aide to help explain or reinforce learning.

What is new today is the *medium* of games and the arguments of the proponents of game-based learning to rebuild education for the postindustrial, high-tech world by thinking about learning in a new way (see Gee, 2003; Prensky, 2001, 2006; Shaffer, 2006). These arguments are typically based on the well-intentioned discussions in educational circles in the United States about the critical role of vision in the creation of educational system that properly address twenty-first century needs (see Papert & Caperton, 1999). twenty-first century people, it is argued, require different skills, often called the twenty-first century skills, which include (1) complex problem solving skills; (2) collaboration skills; (3) communications skills; and (4) information literacy as these skills are fundamental to the success of knowledge workers in an ever-changing, global society. Riding on these arguments of the educational policy makers, in the United States, the proponents of digital game-based learning promoted the unique affordances of the game-based learning environments to facilitate the acquisition of the twenty-first century skills (see Galarneau & Zibit, 2007).

In addition, proponents of the digital game-based learning argued that game spaces have enabled play and learning merge in fundamental ways, which is quite different from what we have come to consider as standard pedagogical practice (Thomas & Brown, 2007). In the traditional paradigms of instruction, learning concerns with *learning about* whereas these new forms of learning deal with knowledge through the dynamic of *learning to be,* that is, spaces where work and play, convergence and divergence, and reality and imagination intertwine, and where students become involved in communities of practice and learn how to be realize the things they imagine. The notion of *learning to be* is supported by the situated learning theory, in which students not only learn how to apply knowledge and solve problems, but also take on an identity by learning to think like professionals, historians, engineers, or scientists in situ, that is, in communities of learners and practice (Barab & Duffy, 2000; Brown, Collins, & Duguid, 1989). The virtual worlds of games make it possible to develop situated understanding in a *virtual practice world* (Schön, 1987) contributing to *staged learning opportunities* (Macy, Squires, & Barton, 2009) replicating real-life. When designed this way, educational games can facilitate development of twenty-first century skills in the context of real-life complex knowledge domains, such as engineering. But do they? How do we know?

Facilitating learning in and about complex knowledge domains is a challenging and fascinating instructional design problem (see Sterman, 1994 for a detailed review on the difficulties of learning in and about complex domains). Central to learning about a complex knowledge domain is the requirement for effective complex problem-solving skills.

In an attempt to arrive at a design framework for educational games that can effectively promote such higher-order learning and problem-solving skills, we reviewed the empirical research on digital educational games with specific emphasis on MMOGs. Peer-reviewed journals in ERIC and PsycINFO databases were reviewed with the search terms of 'game' and 'learning.' Most of the research conducted in the area of educational MMOGs discussed underlying design features and instructional strategies embedded in educational MMOGs, such as Quest Atlantis and River City (e.g., Squire, 2004; Tuzun, 2004; Warren & Dondlinger, 2009), focusing on engagement, relationship, social interactions, and identity. Unfortunately, little or no empirical evidence was provided in support of those proposed design features and instructional strategies.

Even fewer empirical studies exist with regard to digital game-based learning, especially in the area of problem solving acquisition. Our literature review resulted with only 24 journal articles reporting empirical studies (qualitative or quantitative) on the effect of game-based learning. Overall, the findings regarding the educational benefits of games are mixed. Some of these studies claiming positive outcomes appeared to be making unsupported claims since they only utilized learner's self-report of their learning outcomes or satisfaction surveys without a measurement of their actual learning performance.

Other measures that were used to study game-based learning included: performance on game; time to complete the game; visual attention; and knowledge tests. Only one study measured performance improvement on pre- and post-tests. It should be noted that seven of these studies focused on content understanding, one focused on collaboration, and sixteen focused on domain-specific problem solving outcomes, with one also involving on communication. No studies looked at complex problem solving skill acquisition, although it is one of the most important reasons for digital-game based learning as argued by its proponents (see Barab, Tomas, Dodge, Carteaux, & Tuzun, 2005; Dickey, 2005; Gee, 2003; Prensky, 2001, 2006; Shaffer, 2006).

Despite the arguments for the potential of digital game-based learning, the empirical evidence for their effectiveness is scant. If *learning to be* is an important game-based learning experience, then assessing the dynamic game-playing (and the underlying complex problem solving) processes is equally important as, or even more important than solely assessing the learning outcomes. Therefore, we argue for the need to systematically study, which instructional design strategies work in game-based learning environments to take full advantage of what these emerging technologies can offer for education and training. Towards this goal, a scientific attitude with regard to the design of educational MMOGs requires validated measures of learning outcomes and the associated assessment methodologies in order to determine which design elements work best, when, and why.

3 Assessment of Game-Based Learning

Educational assessment is the systematic and theory-based collection and preparation of knowledge, skills, attitudes, and beliefs with the aim of justifying, controlling and optimizing conclusions and procedures. Furthermore, the assessment of change and the detailed investigation of why and how change takes place are of particular interest. Investigating changes of cognitive structures and understanding how to influence them is the key to well-designed and effective learning environments (Ifenthaler, 2008; Scandura, 1988). Moreover, learning is not a simple matter of information retrieval. It is regarded as an active process (Ifenthaler, 2010a) and key learning goals are regarded as skills, not as facile declarative knowledge (McFarlane, 2003). Additionally, aligning learning goals with goals of the game is a challenging task for instructional designers and educational psychologists. In particular, the digital age and its new technologies have created both opportunities and areas of serious concern for learning, instruction, and assessment. Accordingly, an educational assessment in the digital age requires a flexible setting in which skills are dynamically captured, particularly those of problem solving (Almond, Steinberg, & Mislevy, 2002).

According to Newell and Simon (1972), solving well-structured problems is a linear process and consists of two distinct stages: (a) the generation of a problem representation or problem space (i.e., problem solver's view of the task environment); and (b) a solution process that involves a search through the problem space. The representation of a problem consists essentially of the solver's interpretation of the problem, which determines how easily a problem can be solved (Chi & Glaser, 1985). The solver extracts information and attempts to understand the problem or connect it to existing knowledge to form an integrated representation (Gick, 1986). If schema can be activated during problem representation, the solution process is schema driven, with little search for solution procedures. If appropriate schema cannot be activated, the problem solver goes back to an earlier stage and redefines the problem or uses another method to solve the problem (Gick, 1986). More recent research concluded that (1) processes underlying complex, ill-structured problem solving are different than processes underlying well-structured problem solving (Eseryel, 2006; Jonassen, 1997); (2) performance in solving well-defined problems was independent for ill-structured tasks (Dunkle, Schraw, & Bendixen, 1995), and (3) solving ill-structured problems required different skills than those used for well-structured problems (Hong, Jonassen, & McGee, 2003; Seel, Ifenthaler, & Pirnay-Dummer, 2009).

The digital age has originated significant developments in the field of educational assessment. However, the implementation of assessment features into game-based learning environments is only in its early stages because it adds a very time-consuming step to the design process (Chin, Dukes, & Gamson, 2009). Additionally, the impact on learning and quality criteria (e.g., reliability and validity) of technology-based assessment systems are still being questioned (Pellegrino, Chudowsky, & Glaser, 2003). Closely related to educational assessment of problem solving skills is the requirement for adequate and immediate feedback while playing a game. Feedback is considered to be any type of information provided to learners (see Wagner & Wagner, 1985). Feedback plays a particularly important role in highly self-regulated game-based learning environments because it facilitates the development of mental models, thus improving expertise and expert performance (Johnson-Laird, 1989). However, this requires the learner to be sensitive to characteristics of the environment, such as the availability of certain information at a given time, the ease with which this information can be found in the environment, and the way the information is structured and mediated (Ifenthaler & Seel, 2005). However, not only do new developments in computer technology enable us to dynamically generate simple conceptual models and expert representations, but also direct responses to the learner's interaction with the learning environment (Ifenthaler, 2009b). Nevertheless, dynamic feedback within a game-based learning environment presupposes a reliable and valid educational assessment.

Generally speaking, there are two possibilities for educational assessment within game-based learning environments: Assessment after the game has been completed (outcome) and assessment while playing the game (embedded or stealth). Assessment after learning in a game-based environment often focuses on the outcome. However, such an assessment method may neglect important changes during the learning process. Accordingly, instructors and teachers can only compare the individual outcome with previous outcomes, check against other learners or experts. Still, this assessment method does not allow conclusions on the cause of a possible incorrect result. Did the learner not understand the task? Was the task too difficult? Was he or she too excited? Was it a matter of motivation? In addition, an educational assessment after playing the game cannot involve instant feedback while playing the game.

In contrast, assessment while learning in a game-based environment mostly focuses on the process. The benefits of this assessment method are manifold. Firstly, assessing learners while playing a game will provide detailed insights into underlying learning processes. Secondly, tracking motivational, emotional, and metacognitive characteristics while playing a game will help us to better understand specific behavior and the final outcomes. Thirdly, immediate feedback based on the embedded or stealth assessment can point to specific areas of difficulties learners are having while playing the game (Shute & Spector, 2010). Finally, assessment of *clickstreams* (Chung & Baker, 2003; Dummer & Ifenthaler, 2005) could point out strengths and weaknesses of the game design. Hence, an embedded and process oriented assessment must always include multiple measurement procedures which raises the question of reliable and valid ways of analyzing such longitudinal data (Ifenthaler, 2008; Willett, 1988) and provide instant feedback based on the individual assessment (Ifenthaler, 2009b). Such an intelligent assessment and feedback would result in an adaptive game environment, which changes in response to the learner's activity.

Then, how do we assess learning within games? Basically, we distinguish between (1) external and (2) internal assessment of game-based learning. External assessment is not part of the game-based environment whereas internal assessment is part of it. Next, we discuss some selected methods for external and internal game-based assessment.

3.1 External Assessment

External assessment can be implemented before, during, and/or after playing the game. However, as external assessment is not part of the game-based environment it will interrupt playing the game, which is not desirable. A standard method of external assessment is a so-called debriefing session, which follows right after finishing the game. Debriefing sessions could be implemented as paper-and-pencil tests or as computer-based assessment scenarios (Chin et al., 2009). In order to track changes, a briefing session (also realized as paper-and-pencil test or computer-based assessment scenario) before the game functions as a pretest (Ifenthaler, 2009c; Klabbers, 2008).

Debriefing sessions use paper-and-pencil or computer-based *concept-mapping tools* in order to diagnose a person's individual structural and semantic knowledge (Mandl & Fischer, 2000). A concept map consists of different concepts that are connected with named relations. The collected data of a *concept-mapping tool* can be processed directly for further analysis (Johnson, Ifenthaler, Pirnay-Dummer, & Spector, 2009). Concept mapping is a useful method to show one's conceptual knowledge, which is an appropriate measure of domain knowledge and how the domain specific concepts are related to each other. However, concept maps are less useful as measures of problem solving performance.

One method that has been used in various studies to assess problem solver's structural knowledge about the complex knowledge domains is the *think-aloud method* (Ericsson & Simon, 1993). The think-aloud method basically involves providing solver a problem scenario and asking him or her to think-aloud while solving the problem. However, the dual requirement of both solving or reflecting on a complex problem and verbalizing the cognitive processes represents an unfamiliar situation for the test person. Therefore, the test conductor has to ask for detailed information about the test person during the experiment. The data collected represents only a small amount of the cognitive processes, which occur when one solves a complex problem. One problem with the *think-aloud protocol* method is the insufficient and imprecise verbalization of the test person. Furthermore, the quantification of the collected data and the explicit relation of verbal data to cognitive processes call the validity and reliability of this method into question (Nisbett & Wilson, 1977). However, Chi (1997) and Ericsson and Simon (1980) have developed practicable procedures for quantifying verbal data known as protocol analysis. Since then, think-aloud method has been argued as a valid and reliable method to assess a solver's problem space in various studies (e.g. Tarciani & Clariana, 2006).

The *Dynamic Evaluation of Enhanced Problem-solving* (DEEP) method was developed as a web-based knowledge mapping method for problem-based assessment (Spector & Koszalka, 2004). The test persons are provided with a problem scenario in which they are asked to represent their models of the problem space. Spector and Koszalka (2004) report five steps for constructing these models: (1) Listing key facts and causal factors influencing the problem situation, (2) documenting each factor – what it is and how it influences the problem, (3) constructing a graphical depiction of how these factors are linked, (4) making annotations of each node and link, (5) indicating other considerations or approaches. The *DEEP method* identifies gaps between novice and expert decision-making in complex domains. Spector, Dennon, and Koszalka (2005) discuss their initial findings with the *DEEP method* in three different problem domains – engineering design, environmental planning, and medical diagnosis. The findings show that the *DEEP method* enables researchers to identify differences between novice and expert thinking and problem solving.

From a methodological point of view, other forms of external assessment of game-based learning are possible (e.g. reflection papers, learning journals, retrospective interviews, etc.). Additionally, writing plays an important role for external assessment of game-based learning. Writing is not merely a strategy for externalizing and therefore fixing our current knowledge and sharing it with other learners; it also leads to the reorganization and continual construction of knowledge (Eigler, 2005). Automated knowledge assessment tools allow us to produce instant feedback on semantic and structural aspects of the written text after playing the game and thereby promote the learner's self-regulated writing skills (Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, Ifenthaler, & Rohde, 2009). However, the most important aspect of external assessment is the definition of the assessment goals: What knowledge, skills, attitudes, and beliefs should be measured after playing the game?

3.2 Internal Assessment

Internal assessment could be implemented before, during, and/or after playing the game. In contrast to external assessment methods, internal assessment is part of the game-based environment and will therefore not interrupt playing the game. Optimally, assessment is part of the action or tasks within the game-based environment. Automated assessment tools which are part of a game-based environment allow us to produce instant assessment as well as feedback and thereby promote the learner's self-regulated problem-solving skills (Ifenthaler, Pirnay-Dummer, & Seel, 2010). Depending on the assessment method, these systems are based on different scoring methods (see Baker, Chung, & Delacruz, 2008): (1) expert-based scoring (comparing experts' and learners' solutions), (2) data-driven methods (learner's performance is subjected to statistical analysis or machine learning), and (3) domain-modeling methods (learner's knowledge and skills are modeled for specific subject domains). With regard to practicability in game-based environments, we will focus only on expert-based scoring.

The first method compares an expert performance (considered as a referent standard) against a learner's performance. Frequently used applications are concept, causal, or knowledge maps which are automatically scored and compared to an expert's solution (Herl, Baker, & Niemi, 1996; Ifenthaler, 2010b). Various measures are generated on the fly which focus on structural and semantic features (Ifenthaler, 2010c). Numerous studies report the high potential of this scoring approach (e.g. Baker et al., 2008; Chung & Baker, 2003; Herl et al., 1996; Ifenthaler, Masduki,

& Seel, 2009; Ruiz-Primo, Schultz, Li, & Shavelson, 2001; Tarciani & Clariana, 2006). Recently developed tools allow an instant assessment of written text, which includes full sentences and a minimum of 350 words (Pirnay-Dummer & Ifenthaler, 2010; Pirnay-Dummer, Ifenthaler, & Spector, 2010). Written text is regarded as a more natural way of externalizing knowledge within a game-based environment than the application of concept, causal, or knowledge maps. Additionally, a comparison with an expert map or written text is also possible. Even cross-comparisons between written text and graphical representations are possible (Ifenthaler, 2009a; Ifenthaler & Pirnay-Dummer, 2009). Finally, it is important to define what is seen as an expert. Not every expert solution is a good instructional model that should be followed when it comes to training of novices. Hence, the goals of the game-based environment should be defined precisely before an expert-solution is considered for assessment purposes.

The implementation of expert-based scoring, data-driven methods, and domain-modeling methods into a game-based environment is still a challenge and most often very costly and labor intensive. Yet again, the most important aspect of internal assessment is the definition of the assessment goals: What knowledge, skills, attitudes, and beliefs should be measured after playing the game.

4 Methods for External and Internal Game-Based Assessment

Following the discussion of various possibilities and methods for assessing the effectiveness of game-based learning, we now further illustrate those concepts and methods through a specific example of a longitudinal study evaluating the effects of playing a MMOG on students' learning outcomes and complex problem-solving skills. Through this illustration, we also attempt to exemplify how various methods are used adaptively to assess digital game-based learning for various purposes and questions. Additionally, data analysis is discussed.

4.1 A Study on Surviving in Space

Surviving in Space is a MMOG designed for 8th and 9th grade students. In this game, students are asked to play the role of researchers set in a survivor story mode where gamers explore an uninhabited, uncharted island to test their skills at finding necessary resources. Through a series of task scenarios, student teams must work together to apply math, bioscience, geography and geology to maneuver through the game. The goal of the game is to successfully complete these series of task scenarios and become the winning team, which would be sent into outer space to explore colonization of other planets. The scenarios constituted complex problem tasks, which were designed to be aligned with Science, Technology, Engineering, and Mathematics (STEM) related process skills as outlined by the State Department of Education. The game was designed in such a way that it complemented State

STEM curriculum and provided teachers with a game-based learning environment to support their classroom instruction so that their students could apply and practice the process skills they learned in the class as part of the game play. Directions and hints for completing the problem solving tasks were embedded in the game.

A longitudinal study (see Eseryel et al., 2009; Miller, Eseryel, & Ge, 2009; Swearingen & Eseryel, 2010) was carried out to investigate the effects of an educational MMOG, called *Surviving in Space* in a rural high school in the Midwest of the United States. 349 ninth-grade students were randomly assigned to one of the nineteen classes. Out of these nineteen classes, ten were randomly assigned to treatment (game group) condition and nine were randomly assigned to control (no game group) condition. The students in the game group played *Surviving in Space* in teams of four for 2 days per week (50 min per day) during class hour for 16 weeks, while the students in the no-game group participated in a class that was specifically developed to facilitate students' interdisciplinary learning and improving their leadership, management, and decision-making skills. The purpose of the study was to examine the impact of Surviving *in Space* on (1) students' complex problem solving skill acquisition; (2) mathematics achievement; and (3) students' motivation. In order to achieve the assessment goals, both external and internal assessment strategies were adopted, which are illustrated in details below.

4.2 External Assessment Strategies

To assess students' learning outcomes in the areas of problem solving, mathematics, and motivation, some typically external assessment strategies were used in this study.

First, a pretest and posttest design were used to track changes over time and comparisons were made between the treatment and the control group. One week before and after the implementation of the game, students in both groups took the mathematics achievement tests and responded to three motivation surveys, which measured intrinsic motivation, competence, choice, pressure, and relatedness. These two kinds of pretest and posttests were conducted in the traditional paper-and-pencil format. The comparison of the pretest and the posttest and between the treatment group and the control group provided us with information about learners' changes in mathematics and motivation over time and allowed us to collect data on the effects of the game.

The students' complex problem-solving skills were also assessed across time by comparing the pretest scores and the posttest scores and by comparing the two conditions. Instead of simply using paper-and-pencil format, the pretest and posttest scores in complex problem solving were obtained through engaging students in a complex problem-solving task through a problem scenario. The data was assessed through two specific assessment methods: annotated causal representation and think-aloud protocols.

4.2.1 Problem Scenario

A problem scenario is an important means for eliciting students' cognitive processes and problem-solving skills for complex problem solving. For the purposes of evaluating problem-solving acquisition skills, the students were presented with the following problem scenario:

> Suppose you are chosen as the lead scientist of a team of experts by Jonathan McLarin of McLarin International. Your team is sent to space to explore Earth-like planets outside of our solar system. Your team has been traveling in space for about 6 months. Your ship's sensors have just identified a planet, which has sufficient oxygen levels in its atmosphere for humans to be able to breath. You have decided to land on this planet and survey the area. Initial explorations show that the planet has similar characteristics to an island on earth. Your team decided to live on this planet and identify resources needed to survive. This will allow you to complete your mission. As the lead scientist, your task is to guide your team. Before you proceed, you must write a report to Jonathan McLarin to inform him of your team's discovery. The guidelines you received from Mr. McLarin suggest that your mission report must have certain information. Your first step is to report your observations of the planet. Further steps and additional information will be required. Use the space provided on the next page to complete your report and submit it to Jonathan McLarin.

Students were asked to respond to this scenario playing the role of a researcher in a scientific team exploring an Earth-like planet outside our solar system. Problem solving was measured through four separate but sequential tasks: (1) writing their initial observations of the planet, (2) creating a causal representation presenting the situational analysis of the factors likely to influence human survival along with a report elaborating each of the factors that had been identified, (3) elaborating on the relationships among the factors, and (4) providing specific recommendations for ensuring survival on the planet.

4.2.2 Think Aloud Protocol

In this study, the think-aloud method was implemented by asking students to write down on the report sheet what came to their minds in Task 1, 3, and 4 by following the outlines of the report. For example, in Task 1 (Step 1), the students were asked to imagine that they are walking around on this new planet to explore its characteristics and asked them to write a report to Mr. McLarin of their initial observations of the planet. This first step was intended to provide us with information regarding students' background and assumptions related to the complex problem-solving task.

In Task 3 (Step 3), the students were asked to write a descriptive paragraph explaining why the factors they listed in Step 2 were important and how they related to each other. The data collected from Step 3 provided us with more information about students' problem space and were also used to check the alignment with the annotated causal representation developed in Step 2.

Finally, in Step 4, the students were asked to provide recommendations for solution by listing the steps to be taken to allow their team to build a settlement area for humans to live on this new planet and ensure their survival. They were also asked to write in this report to Mr. McLarin how confident they felt of their plan including any concerns they had or any possible problems they expected and what they would do if these problems would arise. Data collected from Step 4 not only provided us with the solution strategies of the students for the complex problem task scenario but we were also able to collect data on whether the students were able to justify their solution approach by providing plausible argumentation and whether they were able to assess possible weaknesses and strengths in their solution strategy.

Complex problem-solving tasks in real world contexts typically involve more than one viable solution approaches. Experts in complex knowledge domains are distinguished from the novices by their ability to view all possible solution approaches but settling on the most viable one while provide plausible arguments and a contingency plan (Eseryel, 2006). Therefore, justification and assessment of possible solution approaches are important for successful complex problem solving in real life contexts.

Students' written verbalization, that is, the think-aloud protocols, were later graded with scoring rubrics, which will be discussed specifically in 4.4.1. The purpose of think-aloud method was to elicit students' performance in problem representation, elaboration, justification and argumentation, as well as generating solutions in response to problems that have been identified.

4.2.3 Annotated Causal Representation

Annotated causal representation is the second problem-solving step or task. The students were prompted to think about what humans need to survive and were asked to make a list of each of these factors in paper. When they finished identifying all of the important factors, they were asked to develop an annotated causal representation by, first, arranging each factor so that related factors are close to each other, then by drawing arrows between the factors they thought were related; and annotating their causal representation by writing on each arrow how the two factors were related. At any point, the students were free to add new factors or delete any factor they previously listed. Students' annotated causal representation provided us with information related to their structural knowledge of the complex problem-solving task.

4.3 Internal Assessment Strategies

Since annotated causal representation and think-aloud methods were used as means to gather information about students' structural knowledge and problem-solving performance during the pretest and the posttest outside the MMOG – *Surviving in Space*, the resulting scores in the areas of annotated causal representation and think-aloud can be treated as external assessment. However, we argue that the causal representation could also be implemented as an internal assessment method. In order to provide empirical evidence for our argumentation, we compared the results of the elicited students' structural knowledge with the annotated causal representation of the problem domain with the elicited students' responses to the problem-solving task of the think-aloud method. Our preliminary results suggested that the two assessment methods provided similar results on students' structural knowledge (Eseryel et al., 2009). Accordingly, in a further step of game development we would use the advantage of the annotated causal representation by embedding it in the game-based learning environment. Moreover, HIMATT automated tools can be integrated to measure the progression of students' problem solving over time as they are engaged in the game scenario and provide them instant feedback (Ifenthaler, 2009b; Shute & Spector, 2010).

In addition, the game-environment was able to automatically track students as they were playing the game. Students' log files included details such as the path each student followed during the game-play, which tasks they completed, how much time they spent in each task, which reports they completed and so on. In addition, the game environment also automatically collected students' chat logs, the reports they wrote to McLarin International, and the spreadsheets they completed as part of their field work in the game (such as their measurements of the Ph-levels from different water sources). All of this data was automatically fed into the teacher application so that each teacher could use those to assess student performance, as they deemed necessary.

4.4 Coding and Analysis

4.4.1 Coding and Analysis of Problem-Solving Protocols

Students' protocols were analyzed by the first research team of three using problem-solving rubrics in two categories: (a) *problem representation* and (b) *generating solutions*, which served as the two dependent variables (Chi & Glaser, 1985; Ericsson & Simon, 1980). Scoring rubrics were developed, which focused on these two dependent variables. We went through an iterative process of developing and refining the rubrics. First, we had a meeting with a group of experienced K-12 teachers, who served as subject matter experts for designing *Surviving in Space*. This group of teachers discussed and worked through the same problem scenario to generate solutions. The experts' problem representation, analysis, justification, and solutions served as critical criteria for developing the rubrics for the two dependent variables. Second, based on the experts' responses we discussed and identified a list of key features important for each of the problem solving tasks, developed descriptors and criteria. Then, we tested the scoring rubrics with a small sample of students' responses, which served as important feedback for us to revise and improve the scoring rubrics. It took the research team several rounds of discussions to reach an agreement on the rubrics with 85% interrater reliability.

Regarding the criteria of problem representation, we measured (a) the number of categories of items a student listed that were needed for survival on the planet, which were compared with those categories generated by the experts (e.g., habitability, terrain, vegetation, and animal life), and (b) how well the student describe and elaborate those items in relation to survival. The criteria of generating solutions includes four areas: (a) whether the students had made a recommendation by listing the measures or actions, (b) whether those solutions were aligned with their previous analysis and problem representations, (c) whether they have made a justification of their recommendation, and (d) how confident they felt about their recommendations.

4.4.2 Coding and Analysis of Annotated Causal Representations

The second research team of three analyzed students' annotated causal representations, which served as the dependent variable of *structural knowledge* (see Funke, 1985; Ifenthaler, 2008; Scheele & Groeben, 1984; Seel, 1999). Analysis began with open coding, which is the examination of the written responses of the students' to the given problem scenario. The language of the participants guided the development of category labels, which were identified with short descriptors, known as in vivo codes, of the concepts in the causal representations. These categories were systematically compared and contrasted, yielding increasingly complex and inclusive categories. Categories were sorted, compared, and contrasted until saturated–that is, until analysis produced no new categories. The coding schema for the categories was developed from the causal representations of the group of K-12 teachers who served as subject matter experts for designing *Surviving in Space*. Experts also reviewed and participated in developing the coding schema. Thirty percent of students' causal representations were coded by three researchers with an interrater reliability (kappa) of 0.96. Then, two of the principle investigators completed the rest of the coding according to the coding schema.

Each students' coded causal representations were compared with expert causal representation on the six dimensions as suggested by the HIMATT method (Pirnay-Dummer et al., 2010) for measuring structural and semantic levels of graphical representations: (1) *surface structure*, which compares the number of propositions (concept – relation – concept) within two representation; (2) *graphical matching*, which compares the diameters of the spanning trees of the representation, which is an indicator for the range or complexity of conceptual knowledge; (3) *density of concepts*, or *gamma*, describes the quotient of terms per concept within a representation; (4) *structural matching*, which compares the complete structures of two representations (expert and subject) without regard to their content; (5) *concept matching*, which compares the sets of concepts within a representation to determine the use of terms (semantic correctness); and (6) *propositional matching*, which compares only fully semantically identical propositions between the two representation. Surface, graphical matching, gamma, and structural matching refer to organizational structure of a representation while concept, and propositional matching measures indicate how semantically similar a student's causal representation is to that of the experts' in the respective category.

5 Discussion and Future Directions

Significant developments in the field of game-based learning have been made. However, the implementation of assessment features into game-based learning environments is only in its early stages because it adds a very time-consuming step to the design process. Also, the impact on learning and quality criteria (e.g. reliability and validity) of technology-based assessment systems are still being questioned. We identified two possibilities of educational assessment within game-based learning environments: Assessment after the game has been completed (outcome) and assessment while playing the game (embedded or stealth). Additionally, we distinguish between two assessment strategies in digital game-based learning: (1) external (assessment that is not part of game-play) and (2) internal (assessment part of the game-play).

The think-aloud method is a prominent assessment approach that could be implemented to assess the outcome of a game. Additionally, the think-aloud method could also be used while playing the game. However, the data collected when playing the game represents only a small amount of the cognitive processes that occur when one solves a complex problem. Another problem of think-aloud protocols for game-based assessment is that this method cannot be embedded in the game as an internal assessment. Hence, the learner will always be interrupted for assessment when playing the game. In addition, the quantification of the collected data and the explicit relation of verbal data to cognitive processes call the validity and reliability of this method into question.

Domain-based knowledge tests are primarily used to assess the outcome of a game. Surely, embedded domain-based knowledge test exist. However, they should be designed in a way that the flow of the game is not interrupted. Accordingly, internal assessment using domain-based knowledge tests is principally possible if the system allows instant feedback on the results of the test. If this is not the case, domain-based knowledge tests are classified as an external assessment method.

Concept maps, knowledge maps, or causal representations realized as web-based knowledge mapping methods open up unique possibilities for assessment and analysis in game-based learning. They are accepted methods to illustrate the meaning of locally discussed information (Eliaa, Gagatsisa, & Demetriou, 2007; Hardy & Stadelhofer, 2006; Ruiz-Primo et al., 2001). Additionally, they can be easily embedded into a game scenario and therefore do not interrupt the game play. Of course, knowledge mapping methods can also be applied as external assessment while playing the game (e.g. for representing the understanding of the problem to be solved) or after playing the game (e.g. for representing the learning outcome). Using HIMATT, which is a newly developed analysis technology (Ifenthaler, 2010c; Pirnay-Dummer et al., 2010), provides a powerful automated and on-the-fly technique which produces results and feedback with regard to expert solutions or other game players. Embedding HIMATT into a game-based learning environment could have many benefits for learners, educators and researchers.

Click-streams or log-files assess the behavior of learners while playing the game. The extensive data has to be analyzed and reported back to the player, which makes this method impossible in most cases. Accordingly, internal assessment can be easily realized with logging all user data. However, a meaningful analysis of the collected data is often not possible. Nevertheless, this data could be a good basis for researchers to learn more about effective design of educational games.

Other promising assessment methods are chat- or forum-logs. This method tracks all communication during game-play. Accordingly, this internal assessment method provides in-depth views of the learner's thinking and learning processes. However, analysis of these logs, when done with classical qualitative approaches, are very time consuming. However, we implemented an automated text-based analysis function into HIMATT which enables us to track the association of concepts from text which contain 350 or more words directly (see Pirnay-Dummer & Ifenthaler, 2010). The algorithms produce quantitative measures and graphical representations which could be used for instant feedback within the game or for further analysis (see Pirnay-Dummer & Ifenthaler, 2010).

Accordingly, many approaches require extensive resources in time and people; therefore, scalability to large number of students within the limited resources of a single teacher is problematic. Especially in school settings, approaches to assessment of learning in complex, ill-structured domains mainly concentrate on assessing domain knowledge, largely through standardized tests. However, previous research in complex knowledge domains such as medicine show that possessing required domain-based knowledge is a necessary but not a sufficient condition for successful complex problem solving performance. Hence, our research study focused on the adaptation and verification of a method for assessing and analyzing progress of learning in complex problem solving that can be ubiquitously embedded in game-based learning environments.

We used two different methods to assess student's progress of learning in complex problem solving. The first method *adapted protocol analysis* (Ericsson & Simon, 1980) to analyze students' responses to the given problem scenario within the framework of the think-aloud method. As discussed above, this method was not embedded into the game. However, the findings will be used to validate and support the results of the second assessment method. The second method utilized *HIMATT method* (Pirnay-Dummer et al., 2010) to analyze students' causal representations. The fully automated assessment and analysis tool could be easily implemented into web-based games. External and internal assessments are both possible.

Future technological developments will enable us to easily embed assessment and analysis methods into game-based learning environment. Internal assessment and instant analysis including personalized feedback will be implemented in a new generation of educational games. However, it is up to educational research to provide theoretical foundations and empirical evidence on how these methodologies should be designed and implemented. We have just arrived in the digital age. It is up to researchers, technologists, educators, and philosophers to make sense of these powerful technologies.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four process architecture. *Journal of Technology, Learning, and Assessment, 1*(5), 3–63.

Baker, E. L., Chung, G. K. W. K., & Delacruz, G. C. (2008). Design and validation of technology-based performance assessments. In J. M. Spector, M. D. Merrill, J. van Merrienboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 595–604). New York: Taylor & Francis Group.

Barab, S. A., & Duffy, T. (2000). From practice fields to communities of practice. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments* (pp. 25–56). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315.

Chi, M. T. H., & Glaser, R. (1985). Problem solving ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 227–257). San Francisco: W. H. Freeman & Co.

Chin, J., Dukes, R., & Gamson, W. (2009). Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming, 40*, 553–568.

Chung, G. K. W. K., & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning and Assessment, 2*(2). Retrieved April 22, 2010, from http://www.jtla.org.

Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development, 53*(2), 67–83.

Dummer, P., & Ifenthaler, D. (2005). Planning and assessing navigation in model-centered learning environments. Why learners often do not follow the path laid out for them. In G. Chiazzese, M. Allegra, A. Chifari, & S. Ottaviano (Eds.), *Methods and technologies for learning* (pp. 327–334). Sothhampton: WIT Press.

Dunkle, M. E., Schraw, G., & Bendixen, L. D. (1995). *Cognitive processes in well-structured and ill-defined problem solving*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Eigler, G. (2005). Textproduzieren als Wissensnutzungs- und Wissenserwerbsstrategie. In H. Mandl & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* (pp. 187–205). Göttingen: Hogrefe.

Eliaa, I., Gagatsisa, A., & Demetriou, A. (2007). The effects of different modes of representation on the solution of one-step additive problems. *Learning and Instruction, 17*(6), 658–672.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Eseryel, D. (2006). *Expert conceptualizations of the domain of instructional design: An investigative study on the DEEP assessment methodology for complex problem-solving outcomes*. Unpublished doctoral dissertation, Syracuse University.

Eseryel, D., Ge, X., Law, V., Hayes, T., Guo, Y., & Ifenthaler, D. (2009). *The effects of an educational massively multiplayer online game on students' complex problem solving skill acquisition.* Paper presented at the Annual Convention of Association for Educational Communications and Technology, Louisville, KY.

Funke, J. (1985). Steuerung dynamischer Prozesse durch Aufbau und Anwendung subjektiver Kausalmodelle. *Zeitschrift für Psychologie, 193*(4), 443–465.

Galarneau, L., & Zibit, M. (2007). Online games for 21st century skills. In D. Gibson, C. Aldrich, & M. Prensky (Eds.), *Games and simulations in online learning: Research and development frameworks* (pp. 59–88). Hershey, PA: Information Science Publishing.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave-Macmillan.

Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist, 21*(1&2), 99–120.
Hardy, I., & Stadelhofer, B. (2006). Concept Maps wirkungsvoll als Strukturierungshilfen einsetzen. Welche Rolle spielt die Selbstkonstruktion? *Zeitschrift für Pädagogische Psychologie, 20*(3), 175–187.
Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89*(4), 206–218.
Hong, N. S., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching, 40*(1), 6–33.
Ifenthaler, D. (2008). Practical solutions for the diagnosis of progressing mental models. In D. Ifenthaler, P. Pirnay-Dummer, & J. M. Spector (Eds.), *Understanding models for learning and instruction. Essays in honor of Norbert M. Seel* (pp. 43–61). New York: Springer.
Ifenthaler, D. (2009a). Bridging the gap between expert-novice differences: The model-based feedback approach. In Kinshuk, D. G. Sampson, J. M. Spector, P. Isaias, & D. Ifenthaler (Eds.), *Proceedings of the IADIS international conference on cognition and exploratory learning in the digital age* (pp. 49–60), Rome: IADIS Press.
Ifenthaler, D. (2009b). Model-based feedback for improving expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 83–101.
Ifenthaler, D. (2009c). Using a causal model for the design and development of a simulation game for teacher education. *Technology, Instruction, Cognition and Learning, 6*(3), 193–212.
Ifenthaler, D. (2010a). Learning and instruction in the digital age. In J. M. Spector, D. Ifenthaler, P. Isaías, Kinshuk, & D. G. Sampson (Eds.), *Learning and instruction in the digital age: Making a difference through cognitive approaches, technology-facilitated collaboration and assessment, and personalized communications* (pp. 3–10). New York: Springer.
Ifenthaler, D. (2010b). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development, 58*(1), 81–97.
Ifenthaler, D. (2010c). Scope of graphical indices in educational diagnostics. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 213–234). New York: Springer.
Ifenthaler, D., Masduki, I., & Seel, N. M. (2009). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science*. Retrieved October 18, 2010, from http://www.springerlink.com/content/y62024472v394468/
Ifenthaler, D., & Pirnay-Dummer, P. (2009). Assessment of knowledge: Do graphical notes and texts represent different things? In M. R. Simonson (Ed.), *Annual proceedings of selected research and development papers presented at the national convention of the Association for Educational Communications and Technology (32nd, Louisville, KY, 2009)*, (Vol. 2, pp. 86–93). Bloomington, IN: AECT.
Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (Eds.). (2010). *Computer-based diagnostics and systematic analysis of knowledge*. New York: Springer.
Ifenthaler, D., & Seel, N. M. (2005). The measurement of change: Learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning, 2*(4), 317–336.
Johnson-Laird, P. N. (1989). Mental models. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 469–499). Cambridge, MA: MIT Press.
Johnson, T. E., Ifenthaler, D., Pirnay-Dummer, P., & Spector, J. M. (2009). Using concept maps to assess individuals and team in collaborative learning environments. In P. L. Torres & R. C. V. Marriott (Eds.), *Handbook of research on collaborative learning using concept mapping* (pp. 358–381). Hershey, PA: Information Science Publishing.
Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65–94.
Klabbers, J. H. (Ed.). (2008). *The magic circle: Principles of gaming & simulation*. Rotterdam: Sense.

Koschmann, T. (Ed.). (1996). *CSCL, theory and practice of an emerging paradigm*. Mahwah, NJ: Lawrence Erlbaum Associates.

Macy, M., Squires, J. K., & Barton, E. E. (2009). Providing optimal opportunities: Structuring practicum experiences in early intervention and early childhood special education preservice programs. *Topics in Early Childhoos Special Education, 28*(4), 209–218.

Mandl, H., & Fischer, F. (2000). Mapping-Techniken und Begriffsnetze in Lern- und Kooperationsprozessen. In H. Mandl & F. Fischer (Eds.), *Wissen sichtbar machen – Wissensmanagement mit Mapping-Techniken* (pp. 3–12). Göttingen: Hogrefe.

McFarlane, A. (2003). Editorial. assessment for the digital age, *Assessment in Education: Principles, Policy & Practice, 10*, 261–266.

Michael, D., & Chen, S. (2006). *Serious games: Games that educate, train, and inform*. Boston: Thomson Course Technology PTR.

Miller, R. B., Eseryel, D., & Ge, X. (2009). *Surviving in space: The effects of a massively multi-player online game (MMOG) on students' motivation.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA, USA.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.

Papert, S., & Caperton, G. (1999). *Vision for education: The Caperton-Papert platform*. Paper presented at the 91st Annual National Governor' Association Meeting. Retrieved April 22, 2010, from http://www.papert.org/articles/Vision_for_education.html

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2003). *Knowing what students know. The secience and design of educational assessment*. Washington, DC: National Academy Press.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77–115). New York: Springer.

Pirnay-Dummer, P., Ifenthaler, D., & Rohde, J. (2009). Text-guided automated self-assessment. In Kinshuk, D. G. Sampson, J. M. Spector, P. Isaias, & D. Ifenthaler (Eds.), *Proceedings of the IADIS international conference on cognition and exploratory learning in the digital age* (pp. 311–316). Rome: IADIS Press.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development, 58*(1), 3–18.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Prensky, M. (2006). *Don't bother me mom–I'm learning*. St. Paul, MN: Paragon House.

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching, 38*(2), 260–278.

Scandura, J. M. (1988). Role of relativistic knowledge in intelligent tutoring. *Computers in Human Behavior, 4*(1), 53–64.

Scheele, B., & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT). Eine Dialog-Konsens-Methode zur Erhebung subjektiver Theorien mittlerer Reichweite*. Weinheim: Beltz.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Francisco: Jossey-Bass.

Seel, N. M. (1999). Educational diagnosis of mental models: Assessment problems and technology-based solutions. *Journal of Structural Learning and Intelligent Systems, 14*(2), 153–185.

Seel, N. M., Ifenthaler, D., & Pirnay-Dummer, P. (2009). Mental models and problem solving: Technological solutions for measurement and assessment of the development of expertise. In P. Blumschein, W. Hung, D. H. Jonassen, & J. Strobel (Eds.), *Model-based approaches to learning: Using systems models and simulations to improve understanding and problem solving in complex domains* (pp. 17–40). Rotterdam: Sense.

Shaffer, D. W. (2006). *How computer games help children learn?* New York: Palgrave Macmillan.

Shute, V. J., & Spector, J. M. (2010). *Stealth assessment in virtual worlds*. Retrieved April 22, 2010, from http://www.adlnet.gov/Technologies/Evaluation/Library/Additional%20Resources/LETSI%20White%20Papers/Shute%20-%20Stealth%20Assessment%20in%20Virtual%20Worlds.pdf

Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2005). Individual and collaborative construction of causal concept maps: an online technique for learning and assessment. In G. Chiazzese, M. Allegra, A. Chifari, & S. Ottaviano (Eds.), *Methods and technologies for learning* (pp. 223–227). Southampton: WIT Press.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building)*. Syracuse, NY: Syracuse University.

Squire, K. (2004). *Replaying history*. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review, 10*(2–3), 291–330.

Swearingen, D., & Eseryel, D. (2010). *The effects of massively multiplayer online games on high school students' mathematics achievement.* Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO, USA.

Tarciani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology, Research, & Development, 54*(1), 65–82.

Thomas, D., & Brown, J. S. (2007). The play of imagination: Extending the literary mind. *Games & Culture, 2*(2), 149–172.

Tuzun, H. (2004). *Motivating learners in educational computer games*. Unpublished doctoral dissertation. Indiana University, Bloomington, IN.

Wagner, W., & Wagner, S. U. (1985). Presenting questions, processing responses, and providing feedback in CAI. *Journal of Instructional Development, 8*(4), 2–8.

Warren, S. J., & Dondlinger, M. J. (2009). Designing games for learning. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (pp. 1183–1203). Hershey, PA: IGI Global.

Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education, 15*, 345–422.