ANALYZING INTERACTION PATTERNS TO

VERIFY A SIMULATION/GAME MODEL

Rodney Dean Myers

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Instructional Systems Technology

of the School of Education

Indiana University

October 2012

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the

requirements for the degree of Doctor of Philosophy.

Doctoral Committee

_____

Theodore W. Frick, Ph. D.

_____

Charles M. Reigeluth, Ph. D.

_____

Michael H. Molenda, Ph. D.

_____

Jeffrey S. Bardzell, Ph. D.

September 25, 2012

This dissertation is dedicated to my parents, Barney and Karen Myers, for providing a childhood full of books and games and for fostering in me a playful and inquisitive nature; and to my brother, Robert Myers, and my sister, Angela Rittenhouse, for never tiring of playing games with me. Most of all, this dissertation is dedicated to my wife, Juliana Tagliaferri, for encouraging and supporting me, for loving and trusting me, and for letting me win every once in a while.

ACKNOWLEDGEMENTS

Rodney Dean Myers

ANALYZING INTERACTION PATTERNS TO VERIFY

A SIMULATION/GAME MODEL

In order for simulations and games to be effective for learning, instructional designers must verify that the underlying computational models being used have an appropriate degree of fidelity to the conceptual models of their real-world counterparts. A simulation/game that provides incorrect feedback is likely to promote misunderstanding and adversely affect learning and transfer. Numerous methods for verifying the accuracy of a computational model exist, but it is generally accepted that no single method is adequate and that multiple methods should be used. The purpose of this study was to propose and test a new method for collecting and analyzing users' interaction data (e.g., choices made, actions taken, results and feedback obtained) to provide quantified evidence that the underlying computational model of a simulation/game represents the conceptual model with sufficient accuracy.

In this study, analysis of patterns in time (APT) was used to compare gameplay results from the Diffusion Simulation Game (DSG) with predictions based on diffusion of innovations theory (DOI). A program was written to automatically play the DSG by analyzing the game state during each turn, seeking patterns of game component attributes that matched optimal strategies based on DOI theory. When the use of optimal strategies did not result in the desired number of successful games, here defined as the threshold of confidence for model verification, further investigation revealed flaws in the computational model. These flaws were incrementally corrected and subsequent gameplay results were analyzed until the threshold of confidence was achieved.

In addition to analysis of patterns in time for model verification (APTMV), other verification methods used included code walkthrough, execution tracing, desk checking, syntax checking, and statistical analysis. The APTMV method was found to be complementary to these other methods, providing quantified evidence of the computational model's degree of accuracy and pinpointing flaws that could be corrected to improve fidelity. The APTMV approach to verification and improvement of computational models is described and compared with other methods, and improvements to the process are proposed.

_____

Theodore W. Frick, Ph. D.

_____

Charles M. Reigeluth, Ph. D.

_____

Michael H. Molenda, Ph. D.

_____

Jeffrey S. Bardzell, Ph. D.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Statement of the Problem

In its statement on the research and design challenges in games and simulations for learning, the Federation of American Scientists (FAS) called for better guidelines and more effective tools for the design of simulation-based learning environments (FAS, 2006). Furthermore, the FAS identified as an important research task the development of knowledge regarding simulation fidelity, noting that incorrect feedback is likely to promote misunderstanding and to adversely affect learning.

In order for simulations and games to be effective for learning, instructional designers must verify that the underlying computational models being used have an appropriate degree of fidelity to the conceptual models of their real-world counterparts. Otherwise the learning experience may be, as Dewey phrased it, "miseducative" (1938/1997, p. 25). Verifying model accuracy alone is not sufficient to ensure learning, but it is necessary. While it may not be possible definitively to verify the accuracy of a model, especially one with stochastic outcomes, the method being tested in this study can be used during the design and development of a simulation as formative evaluation as well as summative evaluation of the final simulation to provide "quantified confidence" (Thacker et al., 2004, p. *v*) of a computational model's accuracy with respect to its intended use.

The purpose of this study is to propose and test a method for collecting and analyzing users' interaction data (e.g., choices made, actions taken, results and feedback obtained) to provide evidence that a simulation's underlying computational model accurately represents the conceptual model of the real-world phenomenon of interest.

Background of the Problem

The paradigm for teaching and learning in schools has changed little in the past one hundred years, even though our society has moved from an industrial age to an information age (Reigeluth, 1994). As a result, learning in the classroom bears little resemblance to learning outside the classroom, with a corresponding greater degree of disengagement among learners. According to a recent High School Survey of Student Engagement, two out of three students report that they are bored in class every day, and 17 percent say they are bored in every class (Yazzie-Mintz, 2007). Some have argued that educators need to utilize the cultural tools of contemporary society to engage students in ways that are familiar to them (Jenkins, 1992, 2006; Strommen & Lincoln, 1992). The field of instructional technology has a long history of researching emerging technologies, including mass media, and prescribing strategies and methods for employing these tools to promote learning (Molenda, 2008).

Digital games and simulations are among the latest mass media of popular culture. For the past several years, the computer and video game industry has achieved annual revenues greater than the movie industry (Chatfield, 2009). In the United States in 2009, digital games generated nearly $19.6 billion (NPD Group, 2010) compared to theatrical receipts of $10.6 billion (Corliss, 2010). These and other indicators have led Prensky (2001) to suggest that we are witnessing the rise of the Games Generation. In 2003, the Pew Internet and American Life Project released the results of a survey of college students (Jones, 2003). Everyone surveyed had played a video, computer, or online game. Seventy percent played at least once in a while, and 65% played occasionally or regularly. About one in ten admitted that playing was a way of avoiding studying. One third admitted to playing games that were not part of instructional

activities during class. A majority (69%) reported no exposure to gaming for educational purposes in the classroom.

Interest in the study and use of games and simulations for learning has visibly increased in the last decade. Rutter and Bryce (2006) compared the periods of 1995-1999 and 2000-2004 and found nearly twice as many peer-reviewed papers on digital games published during the latter period. Bragge and Storgards (2007) used the ISI Web of Science to find 2,100 studies in more than 170 categories related to digital games between 1986 and 2006, with a significant increase beginning in 2003. However, much of the reporting on the use of games for learning is anecdotal, descriptive, or judgmental and not tied to theory or rigorous research (Gredler, 2004; Kirriemuir & McFarlane, 2004; Leemkuil, de Jong, & Ootes, 2000; Washbush & Gosen, 2001; Wideman et al., 2007).

Games and simulations are of increasing interest to educators, in part because of their ability to engage and motivate players to recognize and solve authentic, difficult problems in situated contexts (Gee, 2003; FAS, 2006; Lieberman, 2006; Shaffer, Squire, Halverson, & Gee, 2005). They are compatible with many contemporary theories of learning and related methods of assessment (Becker, 2007, 2008; Gee, 2003; Mislevy & Haertel, 2006; Mislevy, Almond, & Lukas, 2004; Pellegrino, Chudowsky, & Glaser, 2001; Shaffer, 2009; Shute, Ventura, Bauer, & Zapata-Rivera, 2008, 2009). However, when games and simulations represent real-world systems and processes, designers must consider the degree of fidelity appropriate for various components, including the external representation, the underlying model, and the interaction of the components (Alessi & Trollip, 2001; Reigeluth & Schwartz, 1989).

Theoretical/Conceptual Framework

Fidelity is the degree to which a simulation is faithful to that which it simulates. A simulation cannot completely represent something or else it would become the thing itself (Korzybski, 1995). Reigeluth and Schwartz (1989) theorized that the most fundamental aspects of a simulation should have high fidelity, while lower fidelity is appropriate for the more superficial aspects that may otherwise lead to cognitive overload and impede learning and transfer (see also Huang & Johnson, 2009; Kalyuga & Plass, 2009). Some instructional situations that utilize simulations may require high physical fidelity (e.g., flight simulators) and others may require high cognitive fidelity, defined as "the degree to which the simulation faithfully represents conceptual aspects of the actual task" (FAS, 2006, p. 8). Feinstein and Cannon (2002) examined numerous studies from the 1960s and 1970s that focused on the effects of fidelity on training and education. They reported that greater fidelity did not necessarily result in greater learning and in fact may have reduced effectiveness through unnecessary complexity and overstimulation. Similarly, Winn (2002) noted that a virtual environment does not need to simulate the real world to be useful for instruction, and that high fidelity may lead to constrained and inflexible understanding and make it difficult to transfer knowledge and skills to new contexts.

A simulation's underlying model is an important aspect of cognitive fidelity and is addressed through model validation and verification (V&V). Validation is the process of assessing that a conceptual model accurately represents the real world and that the results of the simulation are similar to those of the real-world system, whereas verification is the process of determining that a computational model is calculating results that are consistent with the conceptual model (Feinstein & Cannon, 2002; Sargent, 2008; Thacker et al., 2004). The former is an aspect of external validity and the latter is an

aspect of internal validity, and both relate to the development and use of a simulation. In designing and developing a simulation, representational validation encompasses the game logic and structure (internal, construct validity) and the phenomena being modeled (external, content validity; Feinstein & Cannon, 2002). Fidelity, models, verification, validation, and related concepts are defined and discussed more thoroughly in the review of the literature in the next chapter.

Research questions

The process of developing and testing a new method (or theory or model) is necessarily iterative. This study is an early step toward validating Frick's (1990) analysis of patterns in time (APT) as a reliable method for verifying a simulation's underlying computational model. This study aims to specify a clear process for such verification. Therefore the research questions addressed in this study pertain to the effectiveness of the method and the ways in which it may be improved for subsequent testing and application. Those research questions are:

1. Is the proposed method effective in verifying and improving the accuracy of computational models created for simulations and games?

2. What does the proposed method contribute that is not available through related methods?

3. What improvements can be made to the proposed method?

The first question will be addressed by applying the method to a particular case and analyzing the results. The second question will be addressed by applying other methods that are currently used in model verification and comparing the findings of the methods. The answer to the third question will depend on the findings and on my analytical, interpretative, and imaginative faculties.

Significance of the Study

When simulations and games are designed and used for educational purposes, they must accurately represent that which is to be learned. We would not expect a textbook that is filled with factual errors to be effective in conveying knowledge; it would require significant effort on an instructor's part to correct the resultant misconceptions. Nor should we expect a simulation or game that is poorly designed to be effective in providing the desired experience for the learner. However, software may seem to work well (i.e., produce some result without crashing) yet in fact produce results that are inconsistent with observations and theories of the real world that we desire learners to master.

To provide a more concrete example, the test case for this study is the *Diffusion Simulation Game* (DSG), which is used to teach concepts and practices associated with change agency and the diffusion of innovations. Everett Rogers' (2003) description of diffusion of innovations (DOI) theory is the primary conceptual model that learners are expected to experience and understand through playing the DSG. If the results of actions taken in the DSG are not consistent with those predicted by DOI theory, the learners will form misconceptions about the theory based on their gameplay experiences.

The accuracy with which a simulation modeler translates a conceptual model into a computational model is a critical element in the effectiveness of a simulation. Numerous methods for verifying this accuracy exist and are described in this study, but it is generally accepted that no single method is adequate and that multiple methods should be used. This study proposes and tests a new method which, if shown to be effective, will provide simulation designers with additional evidence that their models are sufficiently accurate with respect to their intended purposes.

CHAPTER 2: LITERATURE REVIEW

In this chapter, I begin with a brief history of the use of games and simulations for learning in order to provide some context for understanding the significance of this study. I examine various proposed definitions for the terms "game" and "simulation" and how those terms combine to form the concept of a "simulation game." I briefly consider methods for classifying games; I then focus on the reasons for using simulations and several methods for classifying simulations. After describing how these terms and concepts will be applied in this study, I further narrow the focus to aspects that are directly relevant to understanding the proposed method, namely the concept of "fidelity" and methods of determining a simulation model's fidelity. Because the proposed method entails the collection and analysis of data generated by the execution of a simulation model, I also discuss some past and current approaches to gathering and making sense of such data.

The Educational Use of Games and Simulations

Historical Precedents

The educational use of games and simulations has been traced back to the use of war games in the seventeenth century and military training in the eighteenth century (Egenfeldt-Nielsen, 2005; Gredler, 2004), although some contend that it dates back nearly 5,000 years ago to war games in China and Japan (Smith, 2010; Wolfe & Crookall, 1998). In the 1950s the practice was adapted for business management training, and in 1956 the American Management Association produced the first widely used business game, Top Management Decision Simulation (Faria, 1998; Leemkuil, de Jong, & Ootes, 2000).

The field of instructional technology has its origins in the visual instruction movement of the early twentieth century, and for many decades it was defined in terms of instructional media. One enduring hallmark of the field is the willingness of designers to experiment with new media and technologies and to study their effectiveness for teaching and learning. Each new medium—from film to radio and sound recording to television to computers and other digital devices—presents its own possibilities and challenges to the instructional designer, and each influences the theories and practices in the field.

As the capabilities of computing technologies have grown increasingly complex and sophisticated, educators have used games and simulations for instruction in a variety of content areas, including medical education, the natural and social sciences, and corporate training. Cruickshank (1988) described several media-based (audio-visual) simulations from the 1960s and 1970s designed for preservice teachers. He noted that in the 1980s computer-based simulations became more prevalent and also more specialized with regard to content or focus. He cited as an example William Harless of the National Library of Medicine, who developed "an interactive videodisk-based simulation designed to teach clinical problem solving by enabling voice input to ask questions [and] order lab tests" (p. 151). In 1987, Faria (cited in Dempsey, Lucassen, Gilley, & Rasmussen, 1993-1994) reported that a survey revealed that 8,755 instructors in 1,900 business schools used business games in their courses. In a follow-up survey a decade later, Faria (1998) found that 97.5% of business schools that were surveyed (n=236) reported using business simulation games in their programs.

Kirriemuir and McFarlane (2004) reported that the use of mainstream games in K-12 education is and will probably remain rare for several reasons. Evaluating a game's

relevance to curriculum and accuracy of content is difficult and time-consuming. A mainstream game that is applicable to curriculum standards will likely have much irrelevant content. Furthermore, most teachers are not familiar with methods for using mainstream games in instruction. De Freitas and Oliver (2006) proposed a framework with four dimensions to guide and support the evaluation of educational games: context, learner specification, pedagogic considerations, and mode of representation. However, even with a framework, choosing a game for use in an educational setting takes time and experience. Fortunately educators are not limited to adapting mainstream games for instructional purposes. Increasingly, games are being designed explicitly for learning, and the willingness of educators to adopt them has led the New Media Consortium (Johnson, Smith, Willis, Levine, & Haywood, 2011) to predict mainstream use of games for teaching and learning within two to three years.

Gredler (2004) stated that the purposes of games and simulations in education are to practice or refine existing knowledge and skills, to identify gaps or weaknesses in knowledge or skills, to develop new relationships among known concepts and principles, and to serve as a summation or review. These are consistent with reviews of the reported use of games, in which games were most frequently used to learn new skills and practice existing skills, generally after the learners had received some introductory instruction to prepare them for the game (Dempsey et al., 1993-1994; Dempsey et al., 1996). Options for integrating games into a curriculum include use as a pre-instructional strategy, a co-instructional strategy, and a post-instructional strategy (for assessment and synthesis; Oblinger, 2006).

A review of the literature led Leemkuil, de Jong, and Ootes (2000) to conclude that there is some consensus that games and simulations will not be effective unless accompanied by instructional support, such as model progression, prompting, feedback (from the game/simulation or an instructor or peers), debriefing, and reflection. Gredler (2004) concurred that open-ended, discovery learning in a simulation is problematic. She recommended that students acquire required knowledge and capabilities (including metacognitive skills) prior to using a simulation. Research consistently concludes that students need some structure in order to learn in discovery-oriented learning experiences (Kirschner, Sweller, & Clark, 2006). Rieber (2005) recommended short explanations offered at the appropriate times within the simulation. He also suggested model progression in which the simulation becomes increasingly difficult based on the learner's mastery of required skills.

Many researchers of games and simulations have emphasized the importance of debriefing in guiding the construction and integration of new knowledge (Dempsey et al., 1996; Garris, Ahlers, & Driskell, 2002; Hays, 2006). Historically, debriefing has been used to obtain information from a participant (e.g., military debriefing of rescued hostages) and to desensitize a participant (or dehoax in the context of a psychological study involving deception; Peters & Vissers, 2004). However, debriefing in the context of experience-based learning focuses on participant reflection and learning. Because participants in a simulation game may have different experiences and therefore derive different—and possibly undesirable—understanding, debriefing is an important phase of experiential learning. Debriefing involves a joint analysis of the participants' experiences. The design of the debriefing should be tailored to the learning objectives and the

participants' characteristics (Peters & Vissers, 2004). Debriefing should focus not just on content but on process, especially when the game is played by teams rather than individuals.

Contemporary Research

Research on the use of simulations and games for learning seems to be increasing. Rutter and Bryce (2006) compared the periods of 1995-1999 and 2000-2004 and found nearly twice as many peer-reviewed papers on digital games during the latter period. Bragge and Storgards (2007) used the ISI Web of Science to find 2,100 studies in more than 170 categories related to digital games between 1986 and 2006, with a significant increase beginning in 2003. However, much of the reporting on the use of games for learning is anecdotal, descriptive, or judgmental and not tied to theory or rigorous research (Gredler, 2004; Kirriemuir & McFarlane, 2004; Leemkuil et al., 2000; Washbush & Gosen, 2001; Wideman et al., 2007). A review of the literature by Dempsey, Rasmussen, and Lucassen (1996) consisted of 99 sources from the previous twelve years. Building on an earlier article (Dempsey et al., 1993-94), the authors defined five categories of gaming articles: discussion (n=51), research (n=38), reviews (n=12), theory (n=11), and development (n=2).

Relatively few studies have been conducted on the use of games for learning in K-12 settings, and these have been primarily case studies, often involving students' and teachers' perceptions of learning (McFarlane, Sparrowhawk, & Heald, 2002; Wideman et al., 2007). Based on their review of the literature, Wideman et al. (2007) concluded that disciplines with the most research in educational gaming are medical education and business management studies. Bragge and Storgards (2007) combined the 170 categories

found in their review into larger domains to find that the three most prominent areas were

social sciences, health sciences, and information and communication technologies.

Rationale for the Use of Games and Simulations

The traditional instructional paradigm through the 1960s was information transfer

from a "knowledgeable educator who constructed and transmitted knowledge … using

the accepted instructional technologies of the day—books, articles, and lectures" (Ruben,

1999, p. 498). Foreman (2004) noted that such a model is based on scarcity of quality

materials and instructors. Furthermore this framework implies that teaching is a

prerequisite for learning; it ignores the social, collaborative, and peer-based nature of

learning outside the classroom (Ruben, 1999). As educators explored more experience-

based approaches to instruction—such as case studies, role playing, simulations, games,

and other structured exercises—the traditional, didactic model gradually ceded

prominence to a learner-centered model emphasizing active, experiential learning

(Ruben, 1999; Garris et al., 2002). This approach accommodates more complex and

diverse approaches to learning by allowing greater interactivity, collaboration, and peer

learning. Ruben (1999) noted that what is important in learning is translating knowledge

into behavior, which requires "reinforcement, application, repetition, and often practice in

a variety of settings and contexts" (p. 499).

Two main reasons for using instructional games are their power to engage and

motivate and their ability to facilitate learning through doing (Kirriemuir & McFarlane,

2004). According to Garris et al. (2002), there are several reasons why educators should

be interested in using games in instruction, including the shift to a learner-centered model

and the intensity of involvement and engagement in games. The memorization of facts

and concepts that is easily measured on a standardized test has led to the presentation of abstract, decontextualized knowledge that is divorced from purpose and instrumentality. In contrast, games require players constantly to use what they have learned to solve situated problems (Shaffer, Squire, Halverson, & Gee, 2005; Wideman et al., 2007). Findings demonstrate that the kinds of experiential learning available in games improve learners' problem-solving skills and judgment (Feinstein & Cannon, 2002). Games can serve as immersive learning environments conducive to experiential learning and can encourage exploration along the lines of guided discovery. Instead of reading about something, students can experience it. Children have also shown learning gains using games in content areas with specific stated objectives, such as math and language skills. In part this is because the active learning required in games facilitates integration of knowledge with existing cognitive structures (Randel, Morris, Wetzel, & Whitehill, 1992).

In their review of the literature, Mitchell and Savill-Smith (2004) found several frequently cited benefits of games in education. These include increases in perseverance, confidence, and self-esteem among learners; the ability to visualize, manipulate, and explore concepts; and greater academic, social, and computer literacy skills. Some studies cited improved metacognition, strategic thinking, problem recognition, and problem solving. In the health sciences, simulations enable students to diagnose and manage virtual patients' problems. In business education, teams manage virtual companies. In both areas, simulations are used to identify students' problem solving abilities and to bridge the gap between classroom instruction and real-world practice (Gredler, 2004).

Many of the attributes of games are also attributes of good instructional design. Games often involve problem solving, provide rapid feedback, and can adjust to an optimal level of difficulty (Oblinger, 2003). Gee (2003, 2005) identified dozens of learning principles that are found in good games, including manipulation and control by the learner, scaffolding and elaboration, well-ordered problems, optimal challenge, skills as strategies and cycles of expertise, information as needed (just in time), systems thinking, and learning by doing.

Many studies of the benefits of playing games to learn have emphasized the motivational or social aspects rather than knowledge acquisition. Intrinsic motivation is generally considered a prerequisite for learning. Garris et al. (2002) described the motivated learner as enthusiastic, engaged, focused, and persistent. The factors that make an activity intrinsically motivating are challenge, curiosity, fantasy, and control (Lepper & Malone, 1987; Malone, 1981; Malone & Lepper, 1987). Not surprisingly, these are all common elements of games. Garris et al. (2002) proposed an input-process-output game model that facilitates intrinsic motivation. The input is a combination of instructional content and game features. The features promote a game cycle of user judgments, user behavior, and system feedback in an iterative loop which, when successful, results in increased engagement, greater persistence of effort, and greater likelihood of achieving intended learning outcomes.

Alessi and Trollip (2001) summarized the many benefits that simulations offer when used for educational purposes. Simulations provide enhanced safety compared to environments and circumstances that may be too dangerous for practice in the real world. A flight simulator is perhaps the most familiar example of this. Furthermore, simulations

can offer experiences that are not readily available in reality, such as being miniaturized and traveling through a human body as in the *Body Wars* ride at Walt Disney World. Simulations can also enable the manipulation of time frames, so that something which takes hours or days in reality can be experienced in minutes. Simulations can even be more cost effective, especially when a large number of people need to be taught or trained. And perhaps most important, the complexity of the learning situation can be controlled in a well-designed simulation so that novices are not cognitively overwhelmed while those with more expertise can experience higher fidelity and greater challenge.

Simulations also have advantages over other media and methodologies. They can increase engagement and motivation by requiring active rather than passive learning, which may be further enhanced by incorporating elements of games like challenge, fantasy, and goals. Alessi and Trollip also cited evidence that well-designed simulations can increase efficiency of learning and improve both near transfer (similar circumstances) and far transfer (different circumstances). Finally, simulations are flexible approaches to learning. They may be used during different phases of instruction, including assessment, and they may be designed to support different philosophies of teaching and learning.

The choice to use a game or simulation should be based on "a detailed analysis of the learning requirements and an analysis of the tradeoffs among alternate instructional approaches" (Hays, 2006, p. 312). In a meta-analysis of studies that compared the instructional effectiveness of games with traditional classroom instruction over 28 years, only 68 empirical studies were found (Randel et al., 1992). Of those, 38 found no differences in effectiveness, 27 found games more effective, and 3 found classroom instruction more effective. However, the authors noted a lack of rigor in the research

designs, including a lack of random sampling, failure to report reliability and validity, and failure to control confounding variables. In a quantitative meta-analysis of simulation gaming, Van Sickle (1986) found weak support for games over other approaches. Only five studies found simulation gaming more effective for immediate recall of knowledge and only two studies found that simulation gaming improved retention of knowledge. However, Hays (2006) criticized Van Sickle's methodology, noting that 6 of the 22 studies did not compare instructional approaches.

This review of the literature on games and simulations for learning has found that their use dates back at least several centuries—if not millennia—and spans a variety of content areas and pedagogical purposes. Furthermore, the use of games and simulations in educational contexts is predicted to rise, driven by a growing body of relevant research, increasing efforts to design games explicitly for learning, and a greater willingness among the mainstream of educators to adopt them. It is therefore imperative that designers of games and simulations for learning have reliable tools and methods for ensuring that their creations lead to the desired educational objectives.

Definitions and Characteristics of Games and Simulations

A variety of definitions for "game" and "simulation" are presented in the literature. Wolfe and Crookall (1998) noted that despite several decades as a field, researchers and practitioners in simulation and gaming are still grappling to create a generally accepted taxonomy. In this section, I examine various proposed definitions and delve into the characteristics that distinguish games from simulations. My purpose is to provide a foundation for understanding when the method of simulation model verification being tested in this study is applicable.

Gredler (2004) defined games as "competitive exercises in which the objective is to win and players must apply subject matter or other relevant knowledge in an effort to advance in the exercise and win," while simulations are "open-ended evolving situations with many interacting variables … in which the participants take on bona fide roles with well-defined responsibilities and constraints" (p. 571). For Garris et al. (2002), the key distinction is that a simulation represents reality and a game does not. Heinich, Molenda, and Russell (1993) stated that games and simulations overlap, but that games tend to be goal-oriented and often have an element of fantasy while simulations offer an abstraction or simplification of real-world environments or processes. This description of simulations is consistent with Reigeluth and Schwartz (1989), who stated that the scenario of a simulation "recreates to a greater or lesser degree a real life situation" (p. 4). Klabbers (2009) synthesized from the literature a definition of a game as "any contest or effort (play) among adversaries or teammates (players) operating under constraints (rules and resources) for an objective (winning, victory, prestige, status, or pay-off)" (p. 33), whereas a simulation is "[a]n attempt to solve a problem or to work out the consequences of doing something by representing the problem or possible course of events mathematically, often using a computer" (p. 34).

De Jong & Van Joolingen (1998) defined a computer simulation as "a program that contains a model of a system (natural or artificial; e.g., equipment) or a process" (p. 180). Similarly, Alessi and Trollip (2001) stated that a simulation has "a model of some phenomenon or activity that users learn about through interaction with the simulation" (p. 213). They claimed that this definition includes microworlds, virtual reality, and case-based scenarios since they all have critical features of simulations.

Salen and Zimmerman (2004) reviewed many of the major writers on games and simulations and synthesized the following definitions: "A *game* is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (p. 80) and "[a] *simulation* is a procedural representation of aspects of 'reality'" (p. 423). They contended, as did Aldrich (2005), that some simulations are not games but that most games are some form of simulation. However, a simulation may contain game-like features and may become a game if a performance goal is set (Coleman, 1970; Prensky, 2001). For example, SimCity is an endless simulation of a city. But if the player sets a particular goal, the simulation becomes a game.

One distinction between games and simulations is the forms they take. Games are often categorized by genre, which include action, adventure, strategy, role-playing, racing, sports, shooting, word games, and puzzles. Often these genres are combined; for example, an adventure game may involve role-playing, shooting, and solving puzzles. Simulations are not categorized by these genres (simulation classifications are discussed below). Games may also be grouped by medium, such as board games, card games, video games, miniature war games, alternate reality games (ARGs), live-action role playing games (LARPs), etc. Digital games are sometimes categorized by the player's perspective: first person, third person, isometric, platform, side-scrolling, and top-down.

Rather than trying to define the distinctions between games and simulations, Aldrich (2005, 2009) pragmatically suggested that it is more productive to think about elements of the instructional experience: simulation elements, game elements, and pedagogical elements. Simulation elements include the underlying model and its representation to the user, as well as the real-world actions the user can take (e.g., by

using an input device), the effects of those actions on the simulation elements (including chain reactions and emergent behavior), and the feedback and results based on those actions. Game elements enhance the experience by promoting intrinsic motivation (e.g., by providing challenge, a narrative, and other engaging elements) and extrinsic motivation (e.g., by including scorekeeping and rewards). Pedagogical elements, such as prompts and scaffolds, promote learning from the experience, help to avoid or correct misperceptions, and facilitate transfer to different contexts.

Perhaps Parlett (1999) was correct in insisting that *game* is a "slippery lexicological customer" (p. 1) and that there is no use in trying to propose a single definition. Wittgenstein argued as much in his *Philosophical Investigations* (trans. 2001), noting that we recognize board games, card games, ball games, and other types of games as being related even though there is no one characteristic common to all of them, which is to say there is no definition that encompasses the variety of forms that games take. Instead there are "family resemblances" ("Familienähnlichkeiten", p. 27) that we recognize in the same way that we identify kinship via physical attributes that express a shared genetic source; a daughter may have her father's eyes while her brother has his father's ears, for example, yet when we see them all together we sense their relation to each other.

For the purposes of this study, formally defining games and simulations will be recognized as problematic. Nevertheless, based on the literature I propose the following working definitions to distinguish games from simulations. A game is a designed experience with a goal or objective that defines the winning state, in which a player competes (either with other players or with the game itself) to overcome obstacles and

achieve the winning state by means specified in the rules of the game. A simulation, on the other hand, represents some form of reality that is referred to as the problem entity or reference system and lacks a competitive element and winning state. Instead, the goal is to manipulate the simulation in order better to understand the underlying model and the real or proposed entity it represents.

Alessi and Trollip (2001) stated that the term *simulation game* is appropriate when the program fits the definition of a simulation and also has characteristics of a game, such as "competition, rules, winning and losing" (p. 213). This view is consistent with Coleman's (1970) perspective. It is important to note that a simulation game is primarily a simulation, that is, there is a recognizable aspect of reality that is modeled with a greater degree of accuracy than is generally found in a game. The popular computer game *Tetris*, for example, may be said to simulate gravity, yet the player who manipulates its "falling" blocks arguably learns nothing about real-world gravity.

The case used in this study (described below) is a simulation game because it models a real-world process but also contains scoring and a winning state. The primary purpose of interacting with the simulation game is to understand the conceptual model and its relationship to the real world. Henceforth I will focus on simulations with the understanding that what I say will be applicable to games to the extent that they contain elements of simulation, in particular that they have an underlying computational model that represents a conceptual model of some form of reality.

Classifications of Simulations

Scholars have proposed numerous methods for classifying simulations. However, these classifications generally do not attempt to encompass all simulations but only

particular types such as those used to make predictions or those intended to teach

something. Furthermore, different terms have been used to describe the same kind of

simulation, including "[m]icroworld, management flight simulator, business simulator,

business game, management simulator, learning environment" (Maier & Grossler, 2000,

p. 135). A classification of simulations is important in order to understand the context in

which the method of simulation model verification is being tested in this study.

Based on the literature (in particular Axelrod, 2007, although he was primarily

concerned with simulation in the social sciences), I have identified four types of

simulations categorized by purpose: entertainment, evaluation, experimentation, and

education. The primary purpose of *entertainment* simulations is to entertain, thrill,

delight, or otherwise engage the user on an emotional level. Entertainment simulations

include video games (e.g., *SimCity*, *Tiger Woods Golf*) and simulator rides (e.g., *Mission:

SPACE* at Disney's Epcot theme park). While learning from these simulations is not

precluded, they are primarily designed for fun rather than for serious purposes.

Evaluative simulations are used for testing and evaluating designs for systems or

devices. This type of simulation is used, for example, in urban planning to evaluate the

effect of a traffic management system on traffic patterns (Yang & Koutsopoulos, 1996)

and in chemistry to evaluate the biodegradability of compounds (Punch et al., 1996).

Experimental simulations are used for experimenting and exploring models of

natural and human systems. Axelrod (2007) further divided this category into simulations

for prediction, for proof (e.g., in support of a theory), and for discovery (e.g., of new

relationships and principles). Some scholars have claimed that rather than being

deductive or inductive, simulation is generative, "a third way of doing science" (Axelrod,

2007, p. 5). Where induction is "the discovery of patterns in empirical data" and deduction "involves specifying a set of axioms and proving consequences that can be derived from those assumptions," simulation—like deduction—begins with an explicit assumption, but then "generates data that can be analyzed inductively" (Axelrod, 2007, p. 5). In this way simulation enables the discovery of possibly unexpected results because of emergent interactions and behaviors, especially in agent-based simulations in which agents employ adaptive strategies as opposed to optimizing strategies (Axelrod, 2007). David (2009) noted that "generative" used in this sense seems to be synonymous with "abductive" along the lines of "Pierces' [sic] second conception of abduction, in which hypothetical explanations are inquired in order to explain a given explanandum" (p. 124).

Kuppers and Lenhard (2005) compared a simulation to a microscope that may be used to view the dynamic qualities of a generative mechanism, with the goal of determining whether the mechanism produces the results predicted by a theoretical model. Thus, in social science research, which generally seeks to develop some kind of theory or model (Gilbert, 2004), simulation experiments can help us to test, compare, understand, and even prove theories and hypotheses of complex social processes (Kuppers & Lenhard, 2005; Smith, 1996).

Educational simulations are used for teaching and training. For this type of simulation, the ultimate goal is generally "to transfer the skills gained in training to the real-world situation" (Liu, Blickensderfer, Macchiarella, & Vincenzi, 2009, p. 70). This type of simulation should have instructional functions built into it and/or be supplemented by some form of instruction (commonly debriefing) because without instruction the learner is likely to develop misperceptions (Gibbons, McConkie, Seo, &

Wiley, 2009). Reigeluth and Schwartz (1989) similarly prescribed the use of an "instructional overlay" to optimize learning. Because the test case for the current study is an educational simulation game, I will delve deeper into the literature regarding this type.

Proponents of educational simulations contend that by simplifying some aspects of reality, learners are better able to analyze and understand the fundamental principles of a system and gain the ability to control the system (Garson, 2009; Maier & Grossler, 2000). Some scholars have identified sub-types of educational simulations based on what is to be learned. Leemkuil et al. (2000) categorized simulations based on underlying models: *conceptual* models based on principles, concepts, and facts related to a system and *operational* models based on sequences of procedures applied to a system. A conceptual simulation is intended to teach "domain-specific principles or processes as well as problem-solving heuristics" (Clark, 2005, p. 605). The learner is usually expected to infer the characteristics of the underlying model (de Jong & Van Joolingen, 1998). An operational simulation is intended to teach "sequences of cognitive and noncognitive operations (procedures) that can be applied to the (class of) simulated system(s)" (de Jong & Van Joolingen, 1998, p. 180).

Alessi and Trollip (2001) took a similar approach to classifying educational simulations. They identified two groups based on educational objective, with each group having two categories:

- Simulations that teach *about* something
    - Physical simulations
    - Iterative simulations
- Simulations that teach *how to do* something

- o Procedural simulations

- o Situational simulations

They noted that many simulations combine these categories.

The first group has as its objective to teach *about* something (this is similar to the conceptual category described above). In this group, physical simulations represent a physical object or phenomenon in real or manipulated time. Iterative simulations also teach about something, but the learner runs them repeatedly and changes parameters to observe the results. These are sometimes referred to as scientific discovery simulations because the underlying model is a "black box" (i.e., it is hidden from the learner) and the learner must infer its characteristics through experimentation.

The second group of simulations has as its objective to teach *how to do* something (this is similar to the operational category described above). In this group, a procedural simulation requires a sequence of actions to accomplish a goal. These actions often include manipulating representations of physical objects, but the focus is on the procedure to be learned. The last type is a *situational* simulation, and Alessi and Trollip stated that it is the least common type because it is difficult and expensive to produce. It is highly probabilistic because it models the behaviors and attitudes of people and organizations. In a situational simulation, the learner often takes on a particular role.

To summarize, in this study I am employing a classification of simulations based on purpose, with four types: entertainment, evaluation, experimentation, and education. Educational simulations are further categorized by objective, with those that teach *about* something encompassing physical and iterative simulations and those that teach *how to do* something encompassing procedural and situational simulations. The case used in this

study to test the simulation verification method (described below) fits the description of a situational simulation, so claims of its effectiveness will be limited to that type, although I will provide a rationale for hypothesizing its effectiveness with other types of simulations given certain conditions.

Fidelity

Fidelity is the extent to which a simulation accurately represents reality (Alessi, 2000; Alessi & Trollip, 2001; Feinstein & Cannon, 2002; Liu et al., 2009). The fidelity of a simulation is important because it affects not only the learner in terms of complexity, cognitive load, and transfer from training to the real world but also the designer in terms of effort and cost (Alessi & Trollip, 2001; Liu et al., 2009; Reigeluth & Schwartz, 1989). In this study, the fidelity of a simulation's underlying model is sometimes referred to as its accuracy.

Types of Fidelity

Overall fidelity is often judged simply by how the simulation looks and feels (Liu et al., 2009). However, the literature identifies several aspects of simulations, each with its own degree of fidelity with respect to the part of reality that it represents. Liu et al. (2009) discussed a simulation's *physical* fidelity and its *psychological-cognitive* fidelity. Physical fidelity concerns how much the physical simulator looks, sounds, feels, and even smells like the real thing. Aspects of physical fidelity include visual-audio fidelity, equipment fidelity, and in some cases motion fidelity. Psychological-cognitive fidelity concerns how well the simulator produces psychological and cognitive aspects of the real-world experience. This sense of fidelity is influenced by task fidelity and functional fidelity, that is, how the simulation responds to inputs.

Gibbons et al. (2009) classified types of fidelity a little differently. Their three dimensions of fidelity encompass (1) the learner's actions, which are governed by task fidelity, environmental fidelity, and haptic fidelity; (2) the processing of the learner's actions, including the timeliness and speed of the simulation's response compared with that in the real world, the accuracy of the computations, and the exactness of the simulation's mechanisms; and (3) the simulation's representation to the learner, which includes sensory stimuli.

Several scholars have discussed fidelity in terms of the simulation's presentation, interaction, and underlying model (Alessi, 2000; Alessi & Trollip, 2001; Reigeluth & Schwartz, 1989, describe a simulation's scenario as including presentation and interaction). The presentation or appearance of the simulation is sometimes referred to as *perceptual* fidelity. In discussing the development of physical skills, Romiszowski (1999) similarly distinguishes between "technical fidelity" as defined by an expert and "perceived fidelity" as that which is simple enough to train a novice without undue complications. The interaction of a simulation's components is called *functional* fidelity, which includes both the simulation's controls and its resulting behavior (Alessi & Trollip, 2001). The fidelity of the underlying model (model fidelity) is the degree to which the simulation's mathematical or logical model represents the real-world phenomenon (Alessi, 2000). Types of models are discussed in the section below on verification and validation of models.

Degree of Fidelity

Although some researchers have studied how fidelity affects the experience and learning outcomes of using simulations, there is little guidance for determining the

necessary levels of fidelity when designing a simulation (Gibbons et al., 2009). The various aspects of simulations, including detail, resolution, error, precision, sensitivity, timing, and capacity (Liu et al., 2009), may have different levels of fidelity within a given simulation, which has been called *selective* fidelity (Andrews, Carroll, & Bell, 1995). Those levels may even change—automatically or through learner control—if the simulation is designed for *dynamic* fidelity (Alessi, 2000).

While high fidelity may be necessary or desirable for certain types of simulations, for example in engineering, it is not always the case for educational simulations. A simplified model helps learners "build their own mental models of the phenomena or procedures and provide them opportunities to explore, practice, test, and improve those models safely and efficiently" (Alessi & Trollip, 2001, p. 214). Maier and Grossler (2000) stated that a less detailed, more abstract simulation may be appropriate for business and economic systems that may otherwise be overwhelmingly complex. Furthermore, a simulation may be augmented by adding instructive feedback, hints, and other techniques to aid in comprehension (Gibbons et al., 2009; Reigeluth & Schwartz, 1989) although this approach may reduce perceived fidelity (Alessi, 2000).

Reigeluth and Schwartz (1989) theorized that the most fundamental aspects of a simulation should have high fidelity, while lower fidelity is appropriate for the more superficial aspects that may otherwise lead to cognitive overload and impede learning and transfer. They suggested that factors to consider include the complexity of the real world environment, the potential for transfer, the motivational consequence of high fidelity, and the expense of achieving high fidelity.

Peters, Vissers, & Heijne (1998) argued that the basic premise of learning from simulations is "that we are able to translate acquired knowledge and experiences from one system to another. The extent to which this translation will be successful depends, among other things, on the degree to which the game is a valid representation of the reference system" (p. 22). However, Feinstein and Cannon (2002) reviewed studies from the 1960s and 1970s that examined the effects of fidelity on learning. They reported that most studies found little or no correlation between high fidelity and learning gains; in fact, in some cases higher fidelity impeded learning. Increasing fidelity is expensive and results in minimal increases in transfer, especially in the later stages of training (diminishing returns). The appropriate amount of fidelity "depends on many factors including the individual trainees, their levels of skill, and the instructor. It also depends on the particular skills to be learned and transferred" (Liu et al., 2009, p. 71).

Evaluation of Fidelity

Fidelity is difficult to measure objectively (Feinstein & Cannon, 2002), and subjective accounts of fidelity may be influenced by an array of confounding variables, including the learner's prior knowledge and experience, mood and temperament, the environment in which the simulation is being run, etc. Comprehensive measurement is considered impossible because of uncertainty, the amount of information involved, complicated attributes and behaviors of reality, and human limitations (Liu et al., 2009).

Liu et al. (2009) identified two major methods for measuring fidelity. The first is a mathematical (objective) method that requires counting "the number of identical elements shared between the real world and the simulation; the greater the number of shared identical elements, the higher the simulation fidelity" (p. 62). The second method

involves a performance matrix that compares a human's performance in the simulation with that person's real-world performance, producing an indirect measure of fidelity.

The method of simulation model verification being tested in this study provides a measure of the fidelity of the underlying model by comparing the model's performance to a predefined goal. Defining that goal requires an understanding of the intended purpose of the simulation, the desired degree of model accuracy, and the effect of other learning objectives on the model's performance.

Validation and Verification of Simulation/Game Models

Because the purpose of this study is to propose and test a method for verifying and improving the accuracy of a simulation's computational model, it is important to understand what the terms *verification* and *model* mean in the context of simulation modeling. I will begin by defining *model* and distinguishing between conceptual and computational models. I will then define.*validation* and *verification* and describe how these terms pertain to models. I will discuss in some detail the various methods currently used for model verification and some of the strengths and weaknesses of those methods.

A simulation model is a representation of "some aspect of reality in symbolic form, usually a set of equations or logical relationships" (Alessi & Trollip, 2001, p. 313). Sargent (2008) summarized a common paradigm of simulation modeling:

> The *problem entity* is the system (real or proposed), idea, situation, policy, or phenomena to be modeled; the *conceptual model* is the mathematical/logical/verbal representation (mimic) of the problem entity developed for a particular study; and the *computerized model* is the conceptual model implemented on a computer (p. 159).

Therefore a simulation consists of two models. The conceptual model simplifies and abstracts the phenomenon of interest. The computational model is the instantiation of the conceptual model in a programming language.

Validation and Verification Defined

David (2009) noted the terminological distinctions between validation and verification in the simulation modeling sense and in the traditional philosophical sense. The two senses are often conflated, leading to confusion and arguments about whether experimental simulations can yield knowledge (see Oreskes, Shrader-Frechette, & Belitz, 1994). Therefore it is important to establish how these terms are used in the simulation modeling community.

Validation and verification are distinct yet related processes for evaluating simulation models during different stages of the development process to ensure to the highest possible degree that a simulation accurately represents its real-world counterpart and is working properly. A commonly used description of the difference between the terms is that validation is concerned with building the *right model* and verification is concerned with building the *model right* (Balci, 1997; Pace, 2004). Because different types of simulations have different purposes (entertainment, evaluation, experimentation, and education, as described above), a simulation's conceptual and computational models must be evaluated with respect to the simulation's purpose (Kuppers & Lenhard, 2005; Sargent, 2008). Methods of validation and verification are used to gather evidence that the conceptual and computational models are "sufficiently accurate" (Thacker et al., 2004, p. 1) for the simulation's intended use.

Validation of the conceptual model. In the simulation building process, a conceptual model of the problem entity must be created and evaluated before it can be instantiated as a computational model. *Validation* is the process of "determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is 'reasonable' for the intended purpose of the model" (Sargent, 2008, p. 159). This definition is consistent with definitions proposed by, among others, Whitner and Balci (1989), Balci (1997), Pace (2004), Thacker et al. (2004), and David (2009).

Because the focus of this study is verification of a computational model, I will not discuss approaches to validating a conceptual model. The interested reader is directed to Sargent (2008) for a discussion of validation techniques. I will, however, briefly highlight errors commonly associated with validation because they also apply to verification. Balci (1998) described three types of errors. A Type I error is called the *model builder's risk* and occurs when credible results are rejected by the simulation user. A Type II error is called the *model user's risk* and occurs when invalid results are accepted by the simulation user. These errors are generally associated with simulations designed for experimentation, but they also pertain to simulations designed for education as they describe the possibility of learners drawing the wrong conclusions from simulation experiences. A Type III error occurs when the simulation designer formulates and solves the wrong problem.

Verification of the computational model. Once a conceptual model has been created (or selected) and validated to ensure that it is sufficient for the purposes of the simulation, it must be instantiated as a computational model that provides the underlying

rules and functions of the simulation. This process requires translating the conceptual model into computer code, which may be accomplished using a simulation language (specific- or general-purpose) or a higher level programming language such as C++ or Java (Sargent, 2008). Model *verification* is the process of ensuring that this transformation has been done with sufficient accuracy, that is, that it has been programmed correctly and returns results consistent with those associated with the problem entity (Balci, 1997; Pace, 2004; Sargent, 2008; Thacker et al., 2004; Whitner & Balci, 1989).

Model Verification Methods

The term *method* has different meanings and so must be defined in order to avoid confusion. Landa (1999) defined a method as "a structured system of instructions and/or actions for achieving some goal" (p. 346). Similary, Edelson (2006) defined a design methodology as "a general design procedure that matches the descriptions of design goals and settings to an appropriate set of procedures" (p. 102). Following these definitions, for the purposes of this study a model verification method is defined as *a procedure for ensuring that a conceptual model has been translated into a computational model with sufficient accuracy*.

Simulation designers have applied a wide variety of methods for the verification of computational models. They usually utilize a combination of methods because no single approach has been found that provides sufficient evidence of a model's accuracy (Sargent, 2008). Models of highly deterministic and predictable phenomena may require only quantitative analysis, whereas models of complex social phenomena may also require qualitative analysis (David, 2009). Simulations of social processes are

complicated by the large number of parameters that must be estimated, the problem of identifying which variables determine behavior, and the difficulty in acquiring adequate empirical data (Garson, 2009; Gilbert, 2004). Even when variables and data collection are highly controlled, analysis of results can be challenging; the influence of initial conditions and the stochastic nature of social simulations may result in wide-ranging outcomes, and some rare events may not occur during testing (Axelrod, 2007; Smith, 1996). Qualitative analysis is best done by experts on the system who can judge the "directions of the output behaviors" and the reasonableness of their values' magnitudes (Sargent, 2008, p. 163).

Whitner and Balci (1989) developed a taxonomy of simulation model verification techniques consisting of six classifications that increase in formality, effectiveness, and complexity: informal, static, dynamic, symbolic, constraint, and formal. Balci (1997) later pared this down to four classifications by moving types of symbolic and constraint analysis into other classifications. He listed 77 techniques for conventional simulations and 38 techniques for object-oriented simulations. Conventional simulations are typically modular, hierarchical, and time-based; object-oriented (including agent-based) simulations utilize inheritance and are non-sequential (Balci, 1997; Zeigler, 1990), and verification of accuracy tends to be more difficult (Balci, 1997; Sargent, 2008). While the method being proposed and tested in this study may be applicable for the verification of object-oriented simulations, such a claim is beyond the scope of this study. Therefore I will focus on techniques for verifying conventional simulation models, based primarily on descriptions from Whitner and Balci (1989). For each classification I will briefly describe representative techniques.

Informal techniques. These methods are the most common and the most subjective as they rely on human reasoning and intuition. Ideally they are used before executable code is tested.

Desk checking. This method involves reviewing one's own work, especially in the early stages of design, although it is better if someone else does it.

Walkthrough. This method is more organized than desk checking. It is usually done by a team of peers (not managers) associated with the development of the simulation. A presenter, usually the modeler, leads the team in the walkthrough, with other members playing various roles (coordinator, scribe, user representative, etc.). Prepared test cases are used. Benefits in addition to error correction include documentation, shared responsibility, and shared expertise. Sargent (2008) claimed that this is one of the most commonly used techniques for model verification.

Code inspection. This method is a more standardized approach with a team whose members perform specialized roles. Errors are documented and classified and subsequently included in a written report of the inspection meeting. The designer then corrects defects and documents the corrections. Sometimes a follow-up inspection is necessary. Benefits are similar to the walkthrough.

Review. This method is similar to code inspection but involves management personnel and is conducted at a higher level with a focus on compliance to stated objectives, standards, guidelines, and specifications rather than accuracy of code. Otherwise the process is essentially the same as code inspection.

Audit. An audit is performed by a single person on a periodic schedule to provide a trail of decisions, actions, etc. It may include "meetings, observations, and examinations" (Whitner & Balci, 1989, p. 11).

Advantages and disadvantages. Informal techniques of verification require little in the way of computer resources but a lot of human resources. They can be subject to human error, which may be minimized if standardized procedures are used. Informal techniques are especially useful early in the design process, but they should not be the only form of verification.

Static techniques. These methods focus on the characteristics of the static model source code—including coding techniques and practices used—but do not require model execution. These may be conducted at any point in the development process and are generally more complex than informal analysis.

Syntax analysis. This method is among the most widely used of verification techniques because it is usually performed automatically by the source code compiler. It generates lots of documentation on data elements as well as submodels and their relationships and interfaces, which can be useful for debugging during dynamic analysis.

Semantic analysis. This occurs during source code translation as the compiler "attempts to determine the modeler's intent in writing the code" (Whitner & Balci, 1989, p. 13), which the modeler should then verify. This method also provides lots of documentation, including uninitialized variables and unintended side effects. It is also useful when combined with dynamic analysis techniques.

Structural analysis. This method focuses on determining whether best practices for software design are being used, especially for "sequence, selection, and iteration"

(Whitner & Balci, 1989, p. 15) control structures, which are the fundamental building blocks of conventional simulations. A structural graph of control flow may show "anomalies, such as multiple entry and exit points, excessive levels of nesting within a structure, and questionable practices such as the use of unconditional branches" (Whitner & Balci, 1989, p. 15) which can be examined in detail later.

Data flow analysis. This method involves examining model variables and their values, dependencies, and transformations. Data flow graphs assist in visualizing relationships and sequences and in troubleshooting problems.

Consistency checking. This method requires examining the model description for contradictions and for consistent use of data (usually documented in a data dictionary). Consistency checking usually utilizes documentation produced during syntax and semantic analysis.

Advantages and disadvantages. Static analysis techniques require moderate computer resources (automated tools but no model execution) and fewer human resources than informal techniques. Static analysis complements other methods, in particular by providing extensive documentation.

Dynamic techniques. These methods involve executing the model and observing its behavior. The complexity of models (in terms of possible execution paths) can make it difficult to determine the best objectives and methods of testing. Decisions can be informed by static and symbolic analyses. Dynamic techniques usually require model instrumentation, which involves inserting code into the executable model to collect information about model behavior.

Top-down testing. Using this method, developers begin by testing the global model, then the submodels. Calls between levels utilize "stubs" or dummy models which do nothing except allow the model to complete its call.

Bottom-up testing. In this method, submodels are completed and tested first, often incrementally. Then they are integrated and the integration is tested.

Black-box testing. This technique is also known as functional testing and is concerned with *what* is produced by the model. Inputs are fed into the model and outputs are analyzed based on model specifications. This approach is typically used for the global model after all submodels have been tested with another approach.

White-box testing. In contrast to black-box testing, this method is concerned with *how* outputs are produced—execution paths based on data flow and control flow graphs (see data flow analysis above).

Stress testing. This technique uses extreme values for inputs. If the model performs well under stressful conditions, it is more likely to be correct (although not necessarily). Errors that do occur might indicate potential problem areas under normal conditions as well.

Debugging. Debugging is the process of finding and removing errors that were discovered during testing. Debugging is time-consuming and expensive but necessary.

Execution tracing. This is a form of model instrumentation that uses traces to follow model execution, data flow, variable declaration, etc. It is typically used for debugging purposes rather than on its own for verification.

Execution monitoring. This method is less detailed than execution tracing. It may degrade model performance but can benefit from the use of statistical methods by sampling data at fixed intervals.

Execution profiling. This is similar to monitoring but at yet a higher level.

Symbolic debugging. Using this technique the modeler uses a tool to set break points in the model execution for manual control, for example, to set variable values or replay execution. Symbolic debugging is often used after other dynamic techniques have been used to identify a problem.

Cause-effect graphing. Based on the model specification, a decision table is created in which "[c]auses are input conditions, effects are transformations of output conditions" (Whitner & Balci, 1989, p. 28). This decision table is the basis for constructing high-yield test cases.

Regression testing. This practice involves rerunning previous tests after mistakes have been corrected to ensure that there are no side effects of the corrections.

Statistical analysis. This is not itself a verification technique but is often used to analyze the large amounts of data generated by dynamic analysis techniques. Smith (1996) argued that because the outputs of micro-simulation runs differ due to their stochastic nature, methods such as analysis of variance may be used to understand these empirical data. Similarly, in discussing social science simulations, Axelrod (2007) recommended multiple runs based on identical parameters followed by statistical analyses to determine "just which results are typical and which are unusual" (p. 7). Examining the correlations of the runs can provide estimates of the reliability of the model (Garson, 2009).

Advantages and disadvantages. Dynamic analysis techniques have a high cost in human resources as much data may be generated that must be examined. Also, for models of any complexity, not all possible execution paths may be tested. Only *adequate* test coverage is possible. Execution history can serve as documentation of model structure and performance, which is especially important for regression testing when tests are rerun (Krahl, 2005).

Formal analysis. These techniques are based on formal mathematical proof of correctness that is usually not attainable. Therefore I will only briefly summarize this approach. "Correctness" in this sense means "that the model meets its specifications" (Whitner & Balci, 1989, p. 35). The Lambda calculus (λ-calculus) "is a system for transforming the programmed model into formal expressions … so that mathematical proof techniques can be applied" (p. 35). Another alternative is to express the programmed models as predicates and use the predicate calculus. Other options include inference, logical deduction, and induction.

Methods of Collecting and Analyzing Interaction Data

Because the method proposed in this study entails the collection and analysis of data generated by the execution of a simulation model, I will discuss some past and current approaches to gathering and making sense of such data. In the past twenty years, interest has grown regarding the need to collect data on users' interactions with hypermediated environments. Hypermedia is defined as "the use of text, data, graphics, audio and video as elements of an extended hypertext system in which all elements are linked so that the user can move among them at will" ("Hypermedia," n.d.). A user's particular sequence of choices from among a prescribed set of alternatives in a hypermediated environment has been referred to metaphorically as an audit trail, an

information trail, a solution path, a navigation path, and in an educational context, an instructional path. Schwier and Misanchuk (1990) attributed to M. W. Petruk the term "audit trail" as a description of "the instructional path taken by each learner" (p. 1). Williams and Dodge (1993) propagated the use of the term in their description of methods for collecting and analyzing data from computer-based instruction, as did Judd and Kennedy (2004). More recently, Loh proposed "information trail" as "a series of agent-detectable *markings* left by another *moving agent* within an information ecology" (2007, p. 329, emphasis in the original). The term "navigation path" is generally used in the context of Web analytics to describe the sequence of links followed by a user, which is sometimes called a "clickstream."

In a series of papers, Schwier and Misanchuk (1990; Misanchuk & Schwier, 1991, 1992) discussed the reasons for collecting audit trail data, methods for analyzing those data, and problems encountered during collection and analysis. They noted the increasing interest in naturalistic observation and the concomitant feature of the audit trail as an unobtrusive method for obtaining data. They proposed that descriptive data captured at decision points could be used for a variety of purposes. During formative evaluation of specific instructional products, such data may indicate where and how learners make errors. Audit trail data may also be used to derive generalizable rules about the construction of paths through interactive or hypermediated instruction, especially when combined with other data "to explain the influences of social variables, individual differences, and cognitive style on the paths taken through the instruction" (Schwier & Misanchuk, 1990, p. 3). McEneaney (2001) extended this idea by suggesting that the association of paths and outcomes may be used to design better paths for different user

needs and objectives. Finally, audit trail data may be used for conducting basic research on the design of computer-mediated environments.

Schwier and Misanchuk (1990) described both descriptive and inferential approaches to analyzing audit trail data. For the former, the simplest method proposed was a raw (unsummarized) data matrix, which would be relatively easy to construct yet difficult to interpret. A summarized version of this matrix would show nodal frequencies and proportions, but it would be difficult to interpret relationships across nodes. Judd and Kennedy (2004) also recognized the limitation of simple counts for discovering and interpreting meaningful patterns in the navigation of complex environments. Williams and Dodge (1993) recommended selecting a small sample of raw data due to limited resources and utilizing qualitative methods to seek phenomena of interest. Given subsequent advances in computing power and software capabilities, this advice may no longer be relevant.

Methods proposed by Schwier and Misanchuk (1990) for visually representing audit trail data include a "petit-point pattern" inspired by Tukey's stem-and-leaf representation of data, and an audit trail tree that uses line thickness to represent path frequency. They discussed using inferential approaches for comparisons between groups and, like Frick (1990), realized the potential for confusion with path analysis in multiple regression. They noted that a normal distribution cannot be assumed and proposed collecting data for a large group of users and regarding that as the "usual distribution" (p. 7) for comparison with data for an individual. This approach might lend itself to the use of a chi-square one-sample goodness of fit test to determine "statistical significance of observed deviations from 'usuality'" (p. 7). Williams and Dodge (1993) proposed time

series and non-linear regression as viable statistical techniques for analyzing trend patterns in data gathered over time. McEneaney (2001) proposed formalizing the node-and-link model of hypertext as both an adjacency matrix for computational analysis and a di-graph for visual analysis.

There are many potential problems related to the collection and analysis of audit trail data. In particular, many interactive or hypermediated environments are multilinear, which can make it difficult to define the differences between individual audit trails and to combine and compare groups of audit trails (Schwier & Misanchuk, 1990; Misanchuk & Schwier, 1991). Furthermore, audit trails may be of different lengths and may loop back to previously visited nodes. Depending on how navigation decisions are coded, data may lack context, which Schwier and Misanchuk call "the dependency problem" (1990, p. 3). In brief, given a sequence of numbers representing choices made at each decision point, the meaning of a number depends on the previous numbers (decisions). That is, if two learners each have "1" in the second position, that "1" means something different if one learner chose "2" in the first position and the other chose "3" (because they followed different paths and presumably where given different choices at the second position).

Williams and Dodge (1993) addressed the dependency problem by capturing contextual data along with user actions. They described the programming constructs used to capture categorical data of user actions that included the action taken by the user (e.g., mouse click, menu selection, keyboard response), the location of a mouse event (X/Y coordinates for both click and release when relevant), and the time at which the interaction occurred. At the time, they were working with HyperCard, a programming application for the Macintosh computer. Objects created in HyperCard could contain

scripts (handlers) that responded to events. They created four major handlers to track learner actions (trails). These handlers initiated the trail variable in memory, added events to the trail, saved the trail in a text file on disk, and output the trail. They identified several possible types of measurement errors that may be caused by general disruptions, accidents and carelessness, offline behaviors, confusion leading to unintended mistakes, disinterest, perceived intrusiveness, and individual differences in learners.

More recently, Judd and Kennedy (2004) adapted exploratory sequential data analysis (ESDA) techniques for use with audit trail data from multimedia or hypermedia environments. They described four techniques for sequence analysis, one of which deals with state transitions. This technique is "based on Guzdial's (1993) adaptation of Markov chain analysis" (p. 479) and results in diagrams with probabilities of sequences. Judd and Kennedy concluded that EDSA techniques are useful in understanding users' interactions with multimedia applications and may prove fruitful in the analysis of computer-mediated discussions (cf. the use of APT in Barrett, Howard, & Frick, 2011 and Howard, Barrett, & Frick, 2010).

While "audit trail" may be sufficiently descriptive of the researcher's intent in collecting these data, it carries financial connotations that may confuse or mislead the reader. "Information trail" is in some ways preferable to "audit trail," but the use of the word "information" is too imprecise for our purposes. Instead, at the risk of muddying the waters with yet another term, I propose "interaction trail" to stand for *the sequence of actions performed by a user in a hypermediated environment*. These actions may be programmatically coded or manually coded by an observer.

Summary

The purpose of this literature review was to provide a foundation for the reader to understand the context of the problem and how this study seeks to address that problem. The design and use of simulations and games for experiential learning is on the rise, in large part because of their potential to provide safe, authentic, and active learning; to engage and motivate through immersive scenarios with optimally challenging tasks; and to facilitate transfer of knowledge and skills to the real world. However, an inappropriate degree of fidelity can overwhelm novices, frustrate experts, and—most importantly—potentially lead to misunderstandings and adversely affect learning, especially when flaws in a computational model lead to results and feedback that are inconsistent with the real world.

It is therefore critical that designers of games and simulations for learning be able to gauge whether their computational models have achieved the appropriate degree of accuracy given the intended use of their creations. In the following chapters, I will propose and test a new method for verifying whether a computational model has achieved the desired degree of accuracy, for identifying aspects of the model that may be improved, and for measuring the impact of design decisions on the accuracy of the model's performance.

CHAPTER 3: ANALYSIS OF PATTERNS IN TIME
FOR MODEL VERIFICATION

The purpose of this study is to describe, test, and improve a method for verifying and improving the accuracy of a simulation's computational model with respect to the conceptual model of a real-world phenomenon. The proposed method is an application of Frick's (1990) analysis of patterns in time (APT). In this chapter I provide a detailed description of the method, which I refer to as "analysis of patterns in time for model verification" (APTMV) to make clear the particular application of APT. The goal of APTMV is to provide quantified evidence of the accuracy of a computational model by comparing the behavior of and results from a computational model with predictions made based on the related conceptual model.

In this chapter I provide an explanation of analysis of patterns in time (Frick, 1990), the foundational method that is being applied in a heretofore untested context. I also provide examples of studies which have used APT to analyze diverse systems of interest. This chapter concludes with a description of the proposed process for applying APT for model verification and improvement, including the necessary conditions under which the method should be used and considerations for alternative use.

Analysis of Patterns in Time

Map & Analyze Patterns & Structure Across Time (MAPSAT) is a set of relation mapping and analysis methods developed by Dr. Theodore Frick and Kenneth Thompson. Frick's MAPSAT research group at Indiana University is developing software to conduct MAPSAT analyses of systems, including a database designed to facilitate the analysis of temporal and structural patterns of systems. The MAPSAT method for analyzing temporal patterns—known as Analysis of Patterns in Time

(APT)—will be utilized in this study. Therefore we will begin with a brief overview of MAPSAT, focusing on APT.

Frick (1983) proposed a method for analyzing temporal patterns in his doctoral dissertation. Originally conceived "in the mid-1970s as a methodology of classroom observational research to investigate patterns of transactions among students, teachers, curricula, and educational settings," (Frick, 1990, p. 181) nonmetric temporal path analysis (NTPA) was based on concepts from set, probability, information, and general systems theories, in particular as synthesized by Maccia and Maccia as the SIGGS (set, information, graph, and general systems) theory model (Frick & Thompson, 2008). Frick concluded that this approach was superior to the linear models approach (LMA) when studying stochastic educational relations.

In 1990, Frick changed the name of the method to Analysis of Patterns in Time (APT) to avoid confusion with statistical path analysis. In the late-1990s, Frick began collaborating with Thompson, who had been working on Axiomatic Theories of Intentional Systems (ATIS) as an extension of SIGGS (Thompson, 2005, 2008a, 2008b), a "logico-mathematical theory model for analyzing and predicting behavior of systems that are goal-directed or intentional" (Frick & Thompson, 2008, p. 71). This led to a method of measuring system structure which, when combined with Frick's method of measuring system dynamics (temporal patterns), became Analysis of Patterns in Time and Configuration (Frick, Myers, Thompson, & York, 2008). Because both methods may be conceived as the mapping of patterns, the combined methods became known as MAPSAT: Map and Analyze Patterns and Structures Across Time. Therefore MAPSAT consists of two methods: Analysis of Patterns in Time (APT) and Analysis of Patterns in

Configuration (APC). It is theoretically possible to combine these methods to examine changes in the structural properties of a system over time.

APT is an empirical approach to observing and coding phenomena as mutually exclusive and exhaustive categories within classifications. In effect, researchers create measures of temporal patterns by counting the occurrences of these categories (Frick, 1990). Once these data have been collected, researchers specify APT queries to calculate the probability of joint and/or sequential patterns of interest. This is different than the linear models approach of measuring variables separately and using statistics to analyze their relations (Frick et al., 2008). For example, an early study (Frick & Rieth, 1981, as cited in Frick, 1990) applied APT in the classroom observation of academic learning time of handicapped students. Of the numerous classifications used in the study, the analysis of two time measures—the amount of available instruction and the amount of student engagement—revealed a high proportion of student engagement when direct instruction occurred (0.967) as opposed to a much lower proportion when nondirect instruction occurred (0.573). Frick (1990) noted that "students were 13 times more likely to be off task during nondirect instruction than during direct instruction," whereas "the linear correlation between direct instruction and student engagement was about 0.57 … [which] does not reveal the clear pattern indicated by the APT time measure functions for joint events" (p. 184).

In recent years, several studies with diverse research questions have utilized APT maps as a way of understanding temporal patterns of data. An (2003) applied APT in usability tests of software to analyze the frequency of various types of mode errors and their relationship to categories of design incongruity. Frick, Chadha, Watson, and

Zlatkovska (2010) analyzed student evaluations of courses using APT. Among their findings was that "students were three to five times more likely to agree or strongly agree that they learned a lot and were satisfied with courses when they also agreed that [Merrill's] First Principles of Instruction were used *and* students were frequently engaged successfully." Koh (2008) used APT to study the relationship between scaffolded instruction and computer self-efficacy among pre-service teachers. Howard, Barrett, and Frick (2010) used APT to determine the frequency of desirable discourse patterns in the asynchronous computer-mediated communications of pre-service teachers who were critiquing each other's work. Barrett, Howard, and Frick (2011) used APT to investigate the relationship between the quality of a student's product (a website created by the student) and the types of comments that the product received as well as the types of comments that the student made about other students' products.

These varied studies indicate (but do not encompass) the range of applications for APT in educational research. Because my research agenda currently focuses on the design and use of games and simulations for learning, I was drawn to MAPSAT because its methods enable the study of events and relationships in systems. As Salen and Zimmerman (2004) and others have argued, games—and especially contemporary digital games—are complex systems and should be studied as such. I see great potential for applying MAPSAT to the study of games, simulations, and other hypermediated environments.

Description of the APTMV Method

Purpose

The purpose of APTMV is to provide quantified evidence of a computational model's accuracy with regard to the conceptual model that it represents. I use the term

"quantified" because the data that are gathered—the map of joint and sequential patterns—are not quantitative in nature; instead, a logical analysis of an APT map yields a calculated probability of an event (Frick, Barrett, Enfield, Howard, & Myers, 2009). In APTMV, one calculates the probability of the specified interaction patterns leading to a desired outcome referred to as the *threshold of confidence*.

Situational Factors

Timing. APTMV may be used for formative evaluation of a computational model as long as it is possible to generate gameplay data for analysis. Complex simulations with stochastic outcomes will probably require a large amount of data, more than may be feasible to generate through manual gameplay. Instead, designers should include the capability to simulate gameplay through an automated means. The amount of data required for a thorough analysis of a computational model will depend on a number of factors, including the number of decisions that a player must make, the number of variables influencing those decisions, and the number of possible outcomes of those decisions.

APTMV may also be used for summative evaluation of a computational model. In this case, users will have played the simulation and generated gameplay data. A benefit of this approach is that users will sometimes make illogical, unpredictable choices that can uncover errors in the computational model that may not be found through simulated gameplay. Additionally, experienced gamers in particular are likely to detect and exploit flaws in the game, and the use of those strategies should stand out because they will be inconsistent with the strategies predicted to be successful based on the conceptual model.

Type of simulation. Classification of simulations is discussed in detail in the review of the literature. APTMV should be useful in testing any simulation that has expected outcomes (i.e., is based on theory and/or empirical observation) that can be expressed as patterns of joint and/or sequential occurrences of events.

Outline of the APTMV Method

The following outline includes the major steps of the procedure along with guiding questions and explanations. I recommend that a subject matter expert—ideally someone who was not closely involved in the design of the simulation—lead the verification effort. This approach is consistent with several other verification methods described above.

1. Formulate the conceptual model as patterns of temporal events.

    a. What actions does the model specify or imply?

    b. What conditions influence or mediate those actions?

    c. What are the probable results of those actions?

2. Map those events to actions that may be taken in the simulation.

    a. What are the mechanics/controls in the simulation?

    b. How can those mechanics/controls be used, and what conditions influence or mediate their use?

    c. What are the possible outcomes?

3. Validate mappings of events to simulation actions.

    a. Review by selected experts who are knowledgeable about the conceptual model.

b. Possible approaches (depending on the nature of the mappings) include but are not limited to asking experts to:

    i. Agree or disagree with mappings and provide rationale and alternatives.

    ii. Assign probabilities to possible results.

    iii. Rank order possible results.

4. Identify the data associated with those actions that are required for analysis.

    a. Specify in terms of mutually exclusive and exhaustive classifications and categories, as described by Frick (1990).

5. Specify the threshold of confidence.

    a. Given the intended use of the simulation, what are the required probabilities for the model to be considered sufficiently accurate?

6. Programmatically collect and store the data.

    a. Data requirements are based on classifications and categories.

7. Format the data for insertion into the specialized MAPSAT database.

8. Specify queries of the data for patterns of interest using MAPSAT software.

    a. Patterns of interest are those actions in the simulation that map to events indicated in the conceptual model.

    b. Patterns may be joint occurrences of categories and/or sequences of categories.

Note: Because the necessary MAPSAT software is not yet fully operational, manual examination of the data is necessary.

9. Analyze the results of queries to ascertain the probability that the computational model accurately represents the conceptual model. Depending on the type of data collected from experts, this may require statistical analysis or a more qualitative method.

CHAPTER 4: METHODS

The purpose of the following overview is to describe the emergence of

educational design research (EDR) as a valid research design. As Reeves (2006) has

argued, "Educational technology is first and foremost a design field, and thus design

knowledge is the primary type of knowledge sought in this field" (p. 61). Because this

study seeks to test a method for improving the design of simulations and games for

learning, a design-oriented approach is appropriate for generating this kind of design

knowledge. Following this overview, I will describe how EDR will be applied in this

study.

Overview of Educational Design Research

My particular field of interest is instructional design, which goes by many names

including instructional technology, instructional systems technology, and educational

technology. During my initial studies in instructional design, I was introduced to the

major theories, models, and processes that constitute the standard, systematic approach to

designing instruction. Molenda (2008) provides a concise history of the modern field's

progression through several eras in the 20th century—from the Visual and Audio-Visual

Instruction Movements, through Educational Radio, Educational Television, and

Computing and Programmed Instruction. An important aspect of this history is the

development of the systems approach to instructional design, which grew out of

operations research and systems analysis, techniques that originated during World War II

for the design of training in man/machine operations. Molenda emphasizes that the

adoption of the systems approach in the field of instructional technology was less

rigorous than its application in the military; instead it was recognized as "a loose set of

guidelines that were applicable to the complex problems of human learning only by

analogy and not the sort of completely deterministic and tightly controlled methodology described by some of its detractors." (p. 13)

Nevertheless, the systems approach gave birth to a plethora of instructional systems design models in the 1970s and 1980s, and many researchers today continue to develop prescriptive models and processes for the design of instruction. These models and processes are routinely taught to instructional design students and employed—sometimes slavishly—by less experienced designers who seem to find comfort in the certainty of a prescribed approach when faced with the complex problem of helping people to learn (Gordon & Zemke, 2000; Zemke & Rossett, 2002).

However, there seems to be a growing interest in the field in recognizing the complex, situated nature of learning in naturalistic settings (Barab & Squire, 2004), in acknowledging that this complexity creates the potential for unexpected, emergent behavior (Sims & Koszalka, 2008), and in characterizing the design of instruction as a "wicked problem" (Rittel & Webber, 1973) with many potential solutions and a vexing resistance to reduction and analysis. This has led educational researchers to adopt design research methods that were pioneered in other design fields such as architecture and industrial design and to venture from their laboratory settings in order to study and refine their designs and related theories amid the confounding variables of the real world.

Educational design research (van den Akker, Gravemeijer, McKenney, & Nieveen, 2006) encompasses many related approaches or frameworks such as design experiments (Brown, 1992; Collins, 1992), development research (van den Akker, 1999), formative research (Reigeluth & Frick, 1999), design-based research (Barab & Squire, 2004; DBRC, 2003), and design and development research (Richey, Klein, & Nelson,

2004; Richey & Klein, 2007). These approaches have subtle differences but similar goals and methods (van den Akker et al., 2006; see Wang & Hannafin, 2005, for a comparison of these approaches).

Educational design research has its roots in the "design experiments" of the 1980s created by educational researcher Ann Brown, who was trained as a "classic learning theorist prepared to work with 'subjects' (rats, children, sophomores), in strictly controlled laboratory settings" but who ended up studying learning "in the blooming, buzzing confusion of inner-city classrooms" (Brown, 1992, p. 141). Brown says that her design experiments were informed by the procedures of design sciences and that her goal was "to engineer innovative educational environments and simultaneously conduct experimental studies of those innovations" (p. 141).

A decade later, researchers interested in studying learning in naturalistic settings and inspired by Brown's approach began a concerted effort to define the standards and argue the legitimacy of this type of *research through design*. For example, the Design-Based Research Collective defined design-based research (DBR) as "an emerging paradigm for the study of learning in context, through the systematic design and study of instructional strategies and tools" that involves "continuous cycles of design, enactment, analysis, and redesign" (DBRC, 2003, p. 5). Similarly, Reigeluth and Frick (1999) described formative research as "a kind of developmental research or action research that is intended to improve design theory for designing instructional practices or processes" (p. 633). In general, these various approaches to EDR attempt to understand and explicate the relationships among theory, artifacts, and practice while creating or refining generative or predictive theories of learning. In addition to contributing generalizable

knowledge, EDR seeks to produce desired changes in the setting, and it considers the success of these changes as evidence in support of theory. This pragmatic approach draws on the philosophies of Dewey and Peirce, "both of whom have provided systems of inquiry rooted not in claims of truth, but rather in the viability of theories to explain phenomena and produce change in the world" (Barab & Squire, 2004, p. 7).

Research methods in EDR may be quantitative or qualitative or mixed (Kelly, 2006) and often involve rich descriptions of events as well as detailed documents intended to connect processes to outcomes. As with most qualitative methods, in EDR trustworthiness and credibility are related to reliability and validity (Glesne, 2006), while usefulness is similar to generalizability and external validity. Barab and Squire (2004) equated usefulness with consequentiality. In discussing DBR, Hoadley (2004) suggested that rather than measurement validity, DBR is concerned with treatment validity (treatments align with the theories they represent) and systemic validity (the theories are "communicated in a way that is true to the inferences used to prove them"; p. 204). In discussing formative research, Reigeluth and Frick (1999) suggested that instead of validity, the major concern in developing design theory is *preferability*, which they defined as "the extent to which a method is better than other known methods for attaining the desire outcome" (p. 634). The degree to which effectiveness, efficiency, and appeal are valued in a given situation is important in determining the preferability of a method.

The generalizability of outcomes in EDR is problematic because context is recognized as a significant factor. Because the researcher is usually a participant-observer, he or she must make clear that claims are founded on researcher-influenced situations and may not transfer readily to other contexts. This creates threats to credibility

and trustworthiness that must be addressed. One approach is to intervene when necessary in the implementation of a design, and then to study the impact of the intervention as it relates to theoretical issues. As with other kinds of case studies, a design case study can add support to generalizations derived from a collection of related cases (Edelson, 2002).

Research Design

This study may be characterized as a single-case study (Yin, 2009) within the paradigm of educational design research (van den Akker et al., 2006). Many approaches to EDR focus on improving theory, but this study seeks to improve design practice by providing a new method for verifying and improving the fidelity of a simulation's computational model with respect to the desired degree of accuracy. The process of developing and testing a new method (or theory or model) is necessarily iterative. This study is an early step toward validating Frick's (1990) analysis of patterns in time (APT) as a reliable method for verifying a simulation's underlying computational model. This study aims to specify and test a process for such verification. Therefore the research questions addressed in this study pertain to the effectiveness of the method and the ways in which it may be improved for subsequent testing and application. To reiterate, those research questions are:

1. Is the proposed method effective in verifying and improving the accuracy of computational models created for simulations and games?

2. What does the proposed method contribute that is not available through related methods?

3. What improvements can be made to the proposed method?

The first question will be addressed by applying the method to a particular case (described below) and analyzing the results. The second question will be addressed by

applying other methods that are currently used in model verification and comparing the findings of the methods. The answer to the third question will depend on the findings and on my analytical, interpretative, and imaginative faculties. Before describing in detail how I will attempt to answer those questions, I will first provide a description of the case to be used.

Description of the Case

The case selected to test the proposed method is an existing online simulation game, the *Diffusion Simulation Game* (DSG), which has as its primary model Rogers' (1962/2003) description of diffusion of innovations (DOI) theory (Molenda & Rice, 1979). The game is available at https://www.indiana.edu/~simed/istdemo/. This section will begin with a detailed description of diffusion of innovations theory to give the reader a thorough understanding of the conceptual model. This will be followed by descriptions of the original board version of the DSG and the online version of the DSG to give the reader some insight into relevant design decisions of the simulation game.

The Diffusion of Innovations Theory

Everett Rogers (1931-2004) grew up in Iowa, where he became interested in the adoption of agricultural innovations and earned a bachelor's degree in agriculture from Iowa State University (Rogers, 2003; unless otherwise noted, all information in this section is derived from this source). After a stint in the Air Force, he returned to Iowa to pursue graduate studies in rural sociology. While working on his doctoral dissertation, he became convinced that the diffusion of innovations followed a general pattern regardless of the type of innovation or the culture in which it was spreading. He began developing a general model of diffusion and published the first edition of his book, *Diffusion of Innovations*, in 1962. Each subsequent decade he published an updated edition as he

reviewed the latest research and theoretical developments and refined his model. At the time of publication of the fifth edition (2003), Rogers estimated that there were about 5,200 publications on diffusion, with roughly 120 new diffusion publications each year. He claimed that "[n]o other field of behavior science research represents more effort by more scholars in more disciplines in more nations" (p. xviii).

Rogers defines "diffusion" as a social process "in which an innovation is communicated through certain channels over time among members of a social system" (p. 5). The goal of communication with respect to an innovation is to reduce uncertainty by sharing information and subjective evaluations of the innovation. Rogers' definition contains four main elements that are key to understanding the model, including

1. the nature and attributes of the innovation;

2. the communication channels through which information is disseminated;

3. the time required for individuals to make a decision regarding the adoption of the innovation;

4. the social system through which the innovation is diffused.

Perceived attributes of innovations. Rogers defines an innovation as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p. 12). He identified five attributes of innovations which he claims "explain about half the variance in innovations' rates of adoption" (p. 222).

Relative advantage. Rogers defines this as "the degree to which an innovation is perceived as better than the idea it supersedes" (p. 15) and states that this is "one of the strongest predictors of an innovation's rate of adoption" (p. 233). Influences on the perception of relative advantage include economic factors such as initial cost and

subsequent changes in cost; status aspects such as prestige, which may be more important for earlier adopters; and an individual's inclination toward "overadoption," which is adoption without sufficient cause and may be driven by status seeking.

Compatibility. This is "the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters" (p. 15). Rogers suggests that a change agent (defined below as part of the social system) should have a high degree of empathy and rapport in order to understand a client's needs, sociocultural values and beliefs, and previous exposure to related ideas.

Complexity. Rogers uses this term to mean "the degree to which an innovation is perceived as difficult to understand and use" (p. 16). The perception of complexity is negatively related to the rate of adoption of an innovation (i.e. the more complicated something seems, the slower it will be adopted).

Trialability. This is "the degree to which an innovation may be experimented with on a limited basis" (p. 16), ideally in the potential adopter's own setting and circumstances. Rogers says that peers can serve as "a kind of vicarious trial for later adopters" (p. 258).

Observability. This is "the degree to which the results of an innovation are visible to others" (p. 16). The opportunity to observe an innovation in use can stimulate peer discussion and motivate people to seek evaluative information.

Communication channels. Rogers places great importance on the sharing of information about an innovation. He defines a communication channel as "the means by which messages get from one individual to another" (p. 18). He describes two general

mediums of communication—mass media channels and interpersonal channels—and two scopes of channels—localite and cosmopolite.

Mass media channels. These channels include television, radio, newspapers, magazines, and so on, which enable a small number of people to spread their messages to a large audience. Mass media channels are generally effective in creating awareness about the existence of an innovation, especially among earlier adopters who tend to pay more attention to external sources of information. Rogers claims that mass media may also be effective in changing "weakly held attitudes" (p. 205).

Interpersonal channels. Rogers says that these "involve a face-to-face exchange between two or more individuals" (p. 18). Interpersonal communication is less effective in creating awareness or interest in an innovation and more effective in persuading someone to try an innovation about which they are already aware, especially if the message is coming from someone who is "similar in socioeconomic status, education, or other important ways" (p. 18). Rogers defines such similarity as "homophily," the opposite of which is "heterophily." While homophily is common in interpersonal networks, heterophily plays an important role because such links usually span dissimilar cliques of individuals and facilitate the diffusion of an innovation across disparate groups in a system.

It is worth noting that in recent years certain communication channels have blurred the distinction between mass media and interpersonal communications. The Internet in particular facilitates one-to-one, one-to-many, and many-to-many communications via e-mail, blogs, and various social network applications such as Facebook updates and "tweets" to followers on Twitter. The impact of these

communication technologies on the diffusion of innovations seems likely to be a fertile area for research.

Cosmopolite and localite channels. Cosmopolite channels are "those linking an individual with sources outside the social system under study" (p. 207). Mass media channels are largely cosmopolite as they tend to originate outside the system, while interpersonal channels may be either local or cosmopolite. Rogers states that cosmopolite channels are generally more effective in creating awareness about an innovation, and localite channels are more important at the persuasion stage. The relationship between these types of channels and types of adopters will be discussed below.

The various communication channels described above form a communication network of "interconnected individuals who are linked by patterned flows of information" (p. 337). There is much overlap between the study of diffusion and the study of communication networks, which has come to be known as social network analysis; however, only the most salient aspects of the latter will be touched upon here. The interested reader is referred to Scott's (2000) handbook for an accessible introduction to the historical development and principal concepts of contemporary social network analysis.

A key concept of communication networks is the idea of communication proximity, which Rogers defines as "the degree to which two linked individuals in a network have personal communication networks that overlap" (p. 338). In an *interlocking personal network*, all individuals interact with each other (i.e., form a clique); in a *radial personal network*, individuals are linked to a focal individual but do not interact with each other, leading to more openness (information exchange with the environment). The

close proximity of interlocking networks is associated with homophily, while radiality is associated with cosmopolite channels and is positively related to innovativeness. Thus, while heterophilous links of low proximity—which Granovetter (1973) called "weak ties"—are relatively rare, they create important bridges between cliques and facilitate the communication of information. These individuals can also serve as "gatekeepers" (p. 155) and prevent or delay the flow of information.

Time. Based on decades of observation and research, Rogers developed a model of the innovation-decision process, which he defines as "the process through which an individual (or other decision-making unit) passes from first knowledge of an innovation, to the formation of an attitude toward the innovation, to a decision to adopt or reject, to implementation and use of the new idea, and to confirmation of this decision" (p. 20). Rogers describes five stages in this process. In the first edition of his book (Rogers, 1962), these stage were: awareness, interest, appraisal, trial, and adoption. By the fifth edition (Rogers, 2003) these stages had become: knowledge, persuasion, decision, implementation, and confirmation—and he contends that they usually occur in this specific sequence unless, for example, the decision stage precedes the persuasion stage because adoption was declared mandatory by an authority figure.

Stage 1: Knowledge. At this stage the individual becomes aware of the existence and function of the innovation and seeks to reduce uncertainty about what it is and how it works. Mass media channels are especially effective at this stage. An individual may seek information about an innovation to reduce dissonance. Rogers identifies three types of knowledge about an innovation:

- Awareness knowledge (What is it?). The individual becomes aware that the innovation exists, which may lead to the next two types of knowledge. Awareness knowledge is usually best achieved through mass media channels.

- How-to knowledge (How does it work?). This information is necessary to use an innovation properly and is probably most essential to individuals in the decision stage when they become willing to try an innovation.

- Principles knowledge (Why does it work?). This information deals with the functioning principles underlying how an innovation works.

A change agent's efforts to spread information about an innovation may be thwarted by (a) an individual's tendency toward *selective exposure*, by which the individual is less likely to attend to messages that are inconsistent with his or her existing attitudes and beliefs, and (b) *selective perception*, by which the individual is likely to interpret messages in terms of his or her existing attitudes and beliefs.

Stage 2: Persuasion. At this stage the individual forms an opinion about the innovation by seeking to reduce uncertainty about the consequences of adopting the innovation. Interpersonal communications with near peers (based on their personal experience with adoption of the innovation) provides information that is more relevant to the individual's particular situation. The main type of thinking at this stage is affective (or feeling). Selective perception is especially important here as the individual seeks new information and decides what messages seem credible. The perceived attributes influence the individual's opinion of the innovation.

The individual may experience dissonance as a result of the growing gap between his or her attitude toward the innovation and use of the innovation. Rogers refers to this

as the *KAP-gap*, where "KAP" stands for "knowledge, attitudes, practice" (p. 176). This gap may motivate the individual to try the innovation. At this stage, the individual "seeks social reinforcement from others of his or her attitude toward the innovation" (p. 175) so mass media messages are ineffective in providing personalized and credible reinforcement of the individual's beliefs about the innovation.

Stage 3: Decision. At this stage the individual makes a choice regarding adoption of the innovation. Individuals often try to reduce uncertainty by trying the innovation on a limited basis, if possible. For some individuals and some innovations, trial by a peer may suffice. "A demonstration can be quite effective in speeding up the diffusion process, especially if the demonstrator is an opinion leader" (p. 177).

The knowledge-persuasion-decision sequence may be influenced by whether the culture is primarily individualistic or collectivistic. In the latter case, the sequence may be knowledge-decision-persuasion for certain individuals if the collective decides to adopt before the individual is convinced.

Stage 4: Implementation. This stage involves overt behavior change as the individual puts the innovation into use. Even though the individual has made the decision to adopt (or reject), there is still some degree of uncertainty, especially concerning operational problems that may be encountered. Here the change agent may serve as a technical consultant and provide knowledge about procedures and principles relevant to the innovation. Otherwise the individual may experience dissonance and seek to reduce it by discontinuing use of the innovation.

During implementation, the individual may "re-invent" the innovation to suit better his or her needs and conditions. Rogers defines *re-invention* as "the degree to

which an innovation is changed or modified by a user in the process of its adoption and implementation" (p. 180). Reinvention may increase the rate of adoption and sustain the use of the innovation over a longer period of time.

Stage 5: Confirmation. At this stage the individual seeks reinforcement for the decision or reverses the decision, which Rogers calls *discontinuance*: "a decision to reject an innovation after having previously adopted it" (p. 190). Disenchantment discontinuance results from dissatisfaction with an innovation, while replacement discontinuance involves the adoption of a new and better idea. Later adopters (laggards) are more likely to discontinue the use of an innovation. Rate of adoption (defined below) and discontinuance are inversely proportional. "Innovations that have a high rate of adoption should have a low rate of discontinuance" (p. 191).

There are a few key concepts which are related to and/or influence the innovation-decision process. The first is *rate of adoption*, which Rogers defines as "the relative speed with which an innovation is adopted by members of a social system" and which is usually measured "by the length of time required for a certain percentage of the members of a system to adopt an innovation" (p. 23). The cumulative frequency of adoption results in an S-shaped curve, but the slope of the curve varies depending on the perceived attributes of the innovation and differences in social systems. The main goal of the change agent is to accelerate the rate of adoption.

According to Rogers, a change agent can have the greatest impact on rate of adoption by persuading opinion leaders to adopt, which is likely to occur "somewhere between 3 and 16 percent adoption in most systems" (p. 223). When the diffusion curve reaches about 10 percent to 20 percent adoption, critical mass is achieved. Rogers defines

critical mass as "the point after which further diffusion becomes self-sustaining" (p. 343). This point is especially significant for communication technologies, which benefit from "a process of *reciprocal* interdependence" by which later adopters increase the value of the innovation for earlier adopters. Rogers' prescribed strategies for reaching critical mass include

- targeting highly respected individuals;

- influencing individuals' perceptions (e.g., convincing them that critical mass has already been achieved;

- targeting intact groups who are more likely to adopt (e.g., a research and development unit);

- providing incentives for early adoption.

Rogers distinguishes between critical mass and *individual threshold*, which he defines as "the number of other individuals who must be engaged in an activity before a given individual will join that activity" (p. 355). The threshold is at the individual level, while critical mass is at the system level. An individual's threshold may be lower if previous adopters include several people who are in the individual's personal network.

Social system. The fourth and final element in Rogers' model is the social system in which the innovation is to be diffused, which Rogers defines as "a set of interrelated units that are engaged in joint problem solving to accomplish a common goal" (p. 23). A social system may have a formal structure, such as a bureaucracy with hierarchical positions and rules, or an informal structure, such as interpersonal networks with patterned communications among primarily homophilous individuals.

A *change agent* is "an individual who influences clients' innovation-decision in a direction deemed desirable by a change agency" (p. 27). A change agent may be a member of the social system but often comes from outside the system. He or she usually possesses a high degree of expertise, which makes him or her heterophilous with respect to most others in the system. Rogers recommends that the change agent establish rapport with individuals in the system in order to understand their needs and create an information exchange relationship. This may involve diagnosing problems in the system and developing in people a sense of need for change, which should then be cultivated into intent to change. Rogers notes that focusing on early adopters (defined below) can be especially effective as they tend to share certain traits with change agents and also have more influence with their peers. Utilizing their interpersonal networks can move their near peers through the persuasion and decision stages toward adoption.

Rogers categorizes the individuals who form a social system according to their *innovativeness*, which he defines as "the degree to which an individual or other unit of adoption is relatively earlier in adopting new ideas than the other members of a system" (p. 22; see Figure 1).
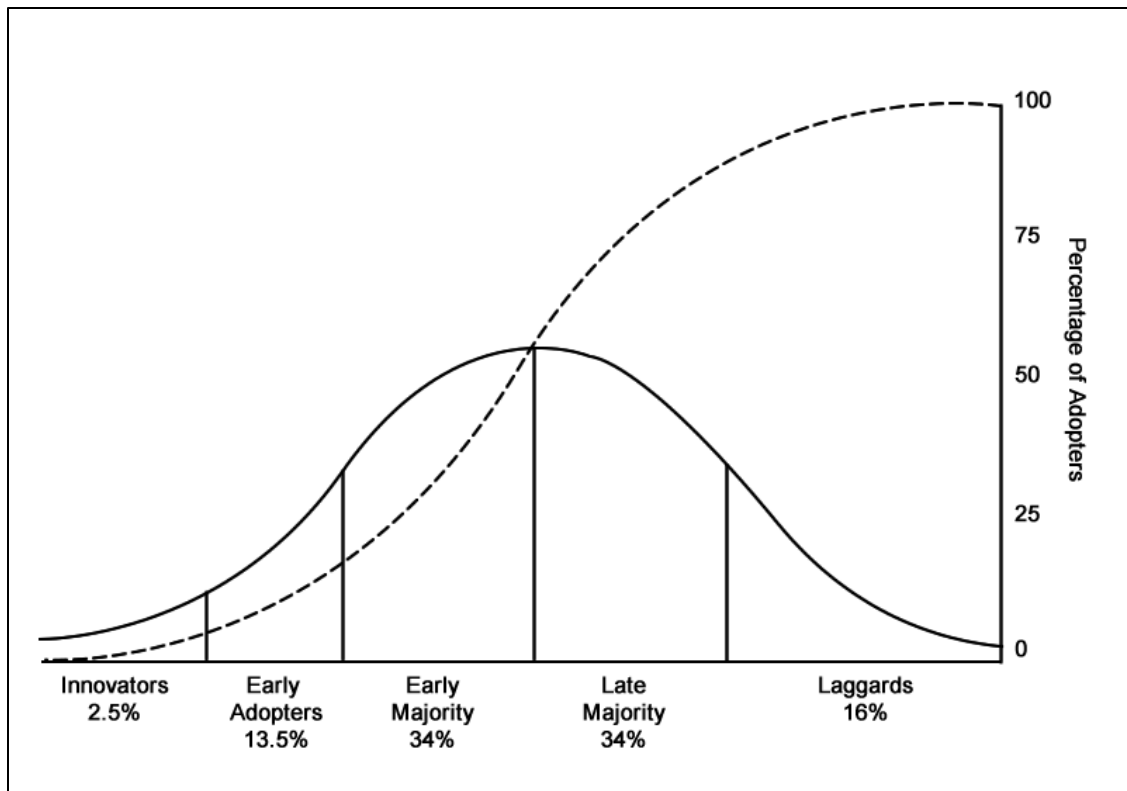
Figure 1. Distribution of adopter types in a population following a bell curve. The S-shaped curve indicated by the dashed line shows one possible rate of adoption for the cumulative number of adopters.

Rogers claims that an individual's network interconnectedness "is positively related to the individual's innovativeness" (p. 330) so that the less connected an individual is, the slower he or she will be in adopting an innovation. Rogers acknowledges that innovativeness is a continuous rather than a discrete variable; nevertheless his categories provide a means of distinguishing among individuals based on their openness to innovations and other related attributes. Such audience segmentation enables the change agent to use different communication channels with different types of adopters at different stages in the innovation-decision process to accelerate the overall rate of adoption. Rogers' five adopter categories in order of innovativeness are:

Innovators. These individuals actively seek information about new ideas through relatively greater exposure to mass media and interpersonal networks that extend well beyond their local system. Therefore they generally play "a gatekeeping role in the flow of new ideas into a system" (p. 283). They tend to be more technologically adept and have a greater tolerance for uncertainty, characteristics which make them less homophilous with others in the system and thus less credible as role models.

Early adopters. These individuals are more localite than innovators and have "the highest degree of opinion leadership in most systems" (p. 283). Rogers defines *opinion leadership* as "the degree to which an individual is able to influence other individuals' attitudes or overt behavior informally in a desired way with relative frequency" (p. 27). Early adopters' and opinion leaders' innovativeness is tempered by the *system norms*, which Rogers defines as "the established behavior patterns for the members of a social system" (p. 26). When a system is more open to change, early adopters and opinion leaders will exhibit a higher level of innovativeness. Other characteristics of opinion leaders include more exposure to mass media, greater contact with change agents, greater social participation, a more central role in the system's communication structure, and a somewhat higher socioeconomic status.

Because of these distinguishing characteristics, opinion leaders—who tend to be early adopters—play a crucial role in the diffusion of an innovation, especially in reaching critical mass and influencing later adopters. However, a change agent may overuse an opinion leader in diffusion activities so that the opinion leader becomes perceived as too similar to the change agent, thereby losing credibility. In general, though, "opinion leadership structures are stable in the relatively short term" (p. 312).

Early majority. These individuals tend to interact frequently with their peers but generally aren't considered opinion leaders. Rogers has noted that "[a]dopter distributions follow a bell-shaped curve over time and approach normality" (p. 275) and that the early majority accounts for roughly a third of the individuals in a system.

Late majority. These individuals tend to be skeptical and cautious, but they are susceptible to peer pressure once uncertainty is sufficiently reduced. Like the early majority, they constitute a third of the population in a system.

Laggards. These individuals are the most localite of all types and have almost no opinion leadership. They are the least connected to others in the system with many being near isolates, making them difficult to influence.

Rogers provides several generalizations regarding the differences between "early knowers" and "late knowers." Research on diffusion has consistently found that "individuals' socioeconomic status is highly related to their degree of change agent contact" (p. 159) which is in turn related to their degree of innovativeness. Early knowers tend to be better educated, wealthier, and have a higher social status and greater upward mobility. They are likely to be less dogmatic, better able to cope with uncertainty, and have a more favorable attitude toward science. In terms of their communication behavior, they have more exposure to mass media, more social connections and participation, more contact with change agents, and a higher degree of opinion leadership. Knowledge of these differences can aid the change agent in identifying and influencing members of the various adopter types.

The Diffusion Simulation Game

The original DSG was conceived and created "in 1975-76 at Indiana University by an Instructional Development Center team composed of professor Michael Molenda and six IST [Instructional Systems Technology] graduate students, led by Patricia Young and Dale Johnson" (M. H. Molenda, personal communication, May 9, 2011). The board game was to be used during a day-long workshop, and Molenda and Rice (1979) reported that it underwent extensive formative evaluation and refinement to ensure that the affective and cognitive objectives were achieved. Among these objectives were the ability to classify individuals by adopter type and communication role (e.g., opinion leader) based on described attributes, to identify the stages of the innovation-decision process, and to select the most effective diffusion activities based on the available information.

In the DSG, the player takes on the role of a change agent whose task is to influence the principal and teachers at a junior high school to adopt peer tutoring. The player may gather information about each staff member and also view diagrams of professional and interpersonal networks, including who eats lunch with whom, who serves on committees together, and who socializes outside of school.

The player may also choose from a variety of diffusion activities, some of which target a single individual or up to five people. For example, the player may use the "Talk To" activity to have a face-to-face discussion with one staff member; the "Print" activity to distribute written materials to as many as five staff members; or the "Local Mass Media" activity to influence those who pay attention to the mass media. The full list of activities with their descriptions is shown in Appendix A. Each activity requires from one

to six weeks to complete, and the player has two academic years (72 weeks) to persuade as many staff members as possible to move through the stages of the innovation-decision process and adopt peer tutoring.

The results of a player's choices are determined by an "algorithm board" (Molenda & Rice, 1979, p. 462) shown in Figure 2.

Figure 2. Algorithm board for the *Diffusion Simulation Game* showing every diffusion activity. The circled numbers indicate sets of cards containing possible outcomes of the activity.

Based on the chosen activity, the affected staff members, and in many cases previously chosen activities, the game monitor consults the algorithm board to determine the outcome, which is a number that refers to a set of feedback cards. For example, if the

"Talk To" activity is selected along with one of the opinion leaders (represented in the game by the letters F, H, and M), the game monitor is instructed to refer to the card set represented by the number 7. This particular card set contains six cards, five of which provide positive feedback and reward points, such as:

> He/she listens attentively to your ideas and shares them with his/her out-of-school compatriates. GAIN 2 POINTS FOR HIM/HER and ONE POINT FOR EACH OF HIS/HER SOCIAL CONTACTS.

The sixth card also provides positive feedback but does not reward points:

> A potentially useful contact; if he/she adopts, a number of others will be favorably disposed. Unfortunately, this is the week his/her family was moving into a new home…no time for serious talk. May be worth trying again later. NO POINTS.

The slight possibility of unfavorable results for what should be effective strategies is meant to model the stochastic nature of dealing with human beings in the real world. One of the affective goals of the game is to foster appreciation for the difficulty of diffusing an innovation.

In 2002, Frick supervised a development team in the creation of the DSG as an online simulation game (Frick, Kim, Ludwig, & Huang, 2003). Figure 3 shows the interface for this online version, which was developed using HTML, CSS, and XML for information display and storage, and PHP for interaction programming. This version of the game may be accessed at https://www.indiana.edu/~simed/istdemo/.
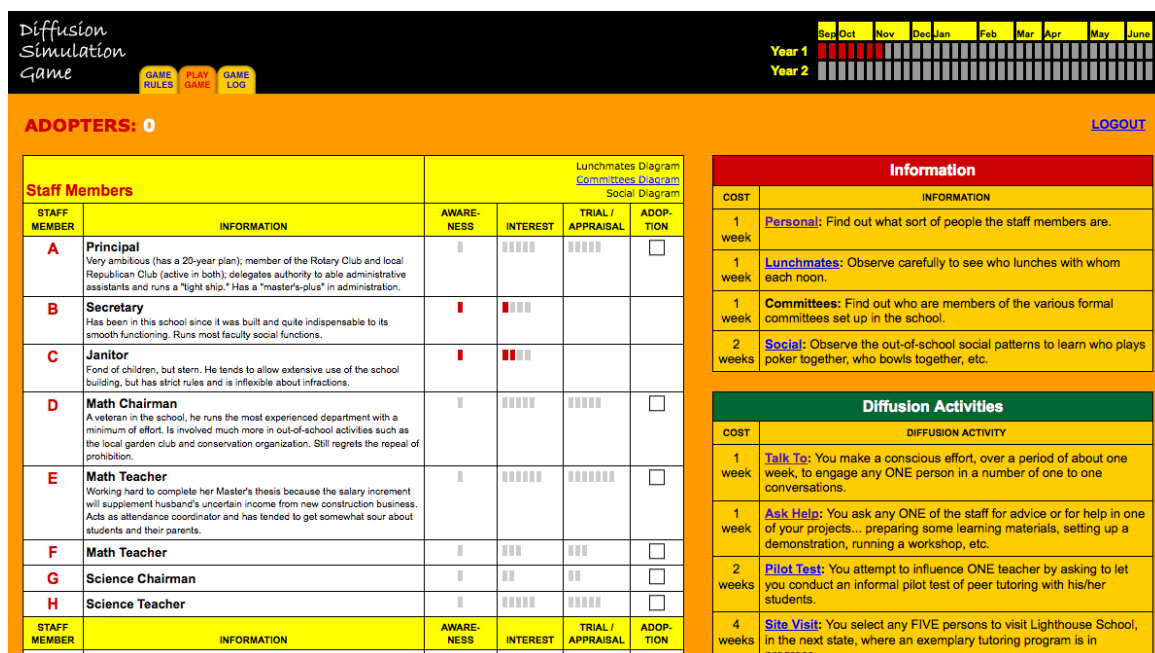
Figure 3. A partial screenshot of the *Diffusion Simulation Game* showing information about potential adopters on the left and activities to choose from on the right.

The Hypertext Markup Language (HTML) is used to describe the structure of Web pages, while Cascading Style Sheets (CSS) describe the visual representation of Web pages (W3C, 2010a). The Extensible Markup Language (XML) is "a simple text-based format for representing structured information" (W3C, 2010b). XML and its related technologies are commonly used to store, retrieve, and manipulate data. PHP Hypertext Processor (PHP) is a flexible scripting language with a parser that resides on a server and handles calls from PHP pages. For more information about the history and use of PHP, the reader is referred to *Programming PHP* (Lerdorf, Tatroe, & MacIntyre, 2006).

Since 2006, when Frick released a public version with anonymous login, data from more than 10,000 game sessions have been collected. In the public version of the DSG, every time a player logs in anonymously a new folder is created on the server that

hosts the game. Two XML template files are copied into the new folder (see Appendix C and Appendix D for samples). One of these files, currentinfo.xml, contains data about the current game state, including the number of weeks elapsed, the total number of adopters, and the number of points obtained in each innovation-decision stage for each staff member. The other file, history.xml, stores information about each turn in the game session, including the activity selected, staff members selected, and the text of the feedback that the player received.

Sample

In this study, I examined two types of data: computer-generated games and player-generated games. As described above, APTMV may be used for formative evaluation during the design and development of a simulation's computational model. To test this, I modified the PHP code that runs the DSG so that it would automatically play a specified number of games using optimal strategies that are described below. To test the usefulness of APTMV for summative evaluation, I compared strategies used in historical gameplay data with the optimal strategies. The historical data were a subset of the 10,000 game sessions that were played between October 7, 2006 and April 4, 2009. This subset consisted of 2,361 *finished* games, which were defined as those games in which the player achieved all 22 possible adopters or used all 72 weeks on the game calendar. These finished games included 107,294 turns in which players selected information or diffusion activities and, where applicable, staff members to participate in those activities.

Data Collection and Analysis Procedures

For a previous study (Enfield, Myers, Lara, & Frick, 2012), we wrote a PHP script to gather data from the currentinfo.xml and history.xml files for each game session. These data were written to a MySQL database. Data fields included a game session

identifier, a number for each turn in a game session (so that the sequence of turns could be analyzed), the number of adopters at the end of each turn, the activity selected for each turn, identifiers for each staff member selected (if any) for each turn, and the text feedback. We found that critical data were not collected. In particular, the data did not indicate at which stage of the innovation-decision process staff members were when the player selected activities. These data are necessary to ascertain the accuracy of the simulation model with regard to the conceptual model and therefore to answer the first research question concerning the effectiveness of the proposed method.

For the current study, I wrote a program to "replay" the selected game sessions based on the extant data and to collect the missing data. For convenience I will hereafter refer to these games as "replayed games." In writing the program to replay the original games, I included the ability to automatically play games using a strategy-selection algorithm based on DOI theory. I will refer to these games as "roboplayed games" to distinguish them from the replayed games. Given the stochastic nature of the result selection in the DSG—whereby even optimal choices occasionally lead to poor results, mimicking the influence of situational factors that aren't and possibly can't be modeled— the roboplay game program was run multiple times to generate data for 100 games during preliminary analyses and 500 games for final analysis. Data from the replayed games and from the roboplayed games were written to a database where they were later retrieved for analysis.

For replayed games, I evaluated the reliability of my data collection process by collecting redundant data to compare with original data. I compared the total number of adopters from the original game session data with the total number of adopters in the

replayed game data to ensure that my program generated the same final results. I also

compared the total number of adopters at the end of each turn in the two datasets, the

feedback text generated by the game algorithm, the activity selected, and when

applicable, the staff members selected. The additional data that I collected were:

1. The innovation-decision stage for each staff member at the beginning of the

   turn.

2. The innovation-decision stage for each staff member at the end of the turn.

3. The number of points awarded to each staff member for the turn.

4. The cumulative number of points each staff member had at the end of the turn.

These data enabled me to examine at which stage staff members were when a particular

activity was selected and how effective that selection was in terms of moving staff

members toward adoption.

Data for both roboplayed games and replayed games were stored in a MySQL

database. The historical data for the replayed games included the following fields. The

"recordID" column contained a numeric identifier for the game session; the "step"

column contained the turn number within a game session; the "activity" column

contained the name of the activity selected; the "totalAdopters" column contained the

cumulative number of adopters achieved in the game by the end of the turn. Other

columns ("S1" through "S5") indicated which staff members were selected during those

turns that used targeted activities. Data for additional columns were generated during

automated replay of the game sessions. The "activityCost" column contained the number

of weeks spent on the selected activity; the "weekTotal" column contained the

cumulative number of weeks used as of the end of the turn; the "turnScore" column

contained the total number of points earned for all staff members during the turn. In addition to these data, four columns were created for each staff member (the "*" in the column name varies from "A" to "X" depending on the staff member's identifier in the game): the "*turnScore" column contained the points gained for the staff member during the turn; the "*turnPhaseStart" column contained the staff members innovation-decision stage at the start of the turn, while the "*turnPhaseEnd" column was the innovation-decision stage at the end of the turn; and the "*totalScore" column contained the staff members cumulative points gained as of the end of the turn. Figure 4 shows a representative part of the data table.

recordID	step	activity	activityCost	weekTotal	turnScore	scoreTotal	totalAdopters	s1	s2	s3	s4	s5	AturnScore	AturnPhaseStart	AturnPhaseEnd	AtotalScore
4	1	personal	1	1	0	0	0	A	B	D	G	J	0			0
4	2	personal	1	2	0	0	0	C	F	R	S	X	0			0
4	3	talk_to	1	3	2	2	0	B	.	.	.	.	0			0
4	4	talk_to	1	4	1	3	0	A	.	.	.	.	1		awareness	1
4	5	talk_to	1	5	0	3	0	F	.	.	.	.	0	awareness	awareness	1
4	6	personal	1	6	0	3	0	E	H	K	M	P	0	awareness	awareness	1
4	7	personal	1	7	0	3	0	I	L	N	O	Q	0	awareness	awareness	1
4	8	personal	1	8	0	3	0	T	U	V	W	.	0	awareness	awareness	1
4	9	talk_to	1	9	1	4	0	X	.	.	.	.	0	awareness	awareness	1
4	10	talk_to	1	10	1	5	0	D	.	.	.	.	0	awareness	awareness	1
4	11	talk_to	1	11	7	12	0	G	.	.	.	.	1	awareness	interest	2
4	12	pilot_test	2	13	4	16	1	G	.	.	.	.	0	interest	interest	2
4	13	talk_to	1	14	1	17	1	F	.	.	.	.	0	interest	interest	2
4	14	talk_to	1	15	3	20	1	P	.	.	.	.	0	interest	interest	2
4	15	pilot_test	2	17	1	21	1	P	.	.	.	.	0	interest	interest	2

Figure 4. Screenshot of part of the data table.

If I had fully operational APT software, I could treat each column as a classification with the cells tracking category changes. I could then specify queries that look for patterns of interest and tally probabilities of success in the game sessions. For example, Strategy 3 (below) predicts that using the *Local Mass Media* and *Print* activities with Innovators and Early Adopters who need points in Awareness or Interest will be successful. An APT query would look for turns containing joint occurrences of those activities, adopter types, and innovation-decision stages. It would return the probability of that pattern occurring in a successful game versus in an unsuccessful game.

I worked around the absence of APT software primarily by using functions in Excel to create conditional statements. I added columns to the table for each strategy. For each turn and each strategy, I calculated a numeric value as described below. These calculations quantified the use of the strategies in each turn.

Following the proposed procedure for APTMV, I began by expressing Rogers' description of DOI theory as a series of statements (strategies) describing actions that should lead to success in the diffusion of an innovation. For example, Rogers says that mass media should be effective in spreading knowledge about an innovation, especially among innovators and early adopters. I then mapped these statements to actions that may be taken in the DSG, which involve combinations of activities, adopter types, and innovation-decision stages (see Appendix A for descriptions of activities available in the game). Next I identified data associated with these actions and designed a database for data collection in which the columns are event classifications (e.g., activity selected, current stage in the innovation-decision process for each staff member) and the rows contain the relevant categories in each classification for each turn in a game.

For this study, I specified two general kinds of strategies. The first kind of strategy involved the selection of an activity available in the game at an appropriate time to influence staff members at particular stages of the innovation-decision process. Some activities, here referred to as *targeted* activities, require the selection of one or up to five staff members. For example, the *Talk To* activity requires the selection of one staff member, while the *Site Visit* activity allows the selection of up to five staff members. The second kind of strategy involved the selection of particular staff members based on their attributes, which include adopter type, opinion leadership, and interpersonal

relationships. Appendix A contains the information that was sent to the expert reviewers, which included descriptions of the available activities in the DSG and the rationale for each strategy predicted to be successful. In this section, I describe how I quantified and analyzed the use of each strategy.

I specified nine strategies that should lead to success in the DSG if the computational model is consistent with the conceptual model (DOI theory). Each of these strategies consisted of a pattern of joint occurrences of categories within the various classifications. To continue the previous example, a turn in the game should be successful if the [classification: *activity*] is [category: *Local Mass Media*] and the majority of staff members who are [classification: *adopter type*] either [category: *innovator*] or [category: *early adopter*] and are in the [classification: *innovation-decision stage*] of either [category: *awareness*] or [category: *interest*]. Each of these strategies is described below, followed by my rationale for the strategy based on Rogers (2003) with references to relevant generalizations he made (which are listed in Appendix B). I also describe how a score for each strategy was calculated.

I should note that because the DSG was developed in the 1970s, it is based on an earlier version of DOI theory than the 2003 edition of Rogers' book with which it is being compared. The rationale for this is that the 2003 version of the theory is currently being taught (i.e., it is the current conceptual model), and the DSG is being used to teach it.

Strategies Based on DOI Theory

Strategy 1: Target earlier adopters and opinion leaders early in the game to work toward critical mass.

Rationale: Achieving a critical mass of adopters can increase the rate of adoption in a system. Innovators are among the first to try innovations and tend to act as gatekeepers in introducing new ideas into a system, but they exert little influence over others. Early Adopters tend to have greater social participation and more central positions in communication networks. Early Adopters also have the highest degree of opinion leadership, and adoption by opinion leaders can have the greatest impact on the rate of adoption. Relevant generalizations: 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 7.18, 7.19, 7.26, 8.2, 8.6, 8.13, 9.11.

Scoring: Unlike most of the strategies described here, this strategy is independent of the activity used. The score for this strategy is added to the activity-based strategy score in a given turn. The earlier adopters and opinion leaders in the game are identified by the letters F, G, H, L, M, P, and X. Targeting them "early in the game" is operationalized as the first 15 turns in a particular game session. The Strategy 1 score is the number of these earlier adopters who are appropriate targets for the activity (usually based on their current innovation-decision phase) divided by the total number of earlier adopters (7). For example, if *Local Mass Media* (which should be effective in raising awareness or interest in an innovation) is the selected activity, the Strategy 1 score is the number of earlier adopters who need points in the Awareness or Interest phases divided by 7.

Strategy 2: Use *Personal Information* and *Talk To* activities.

Rationale: The change agent should establish empathy and rapport in order to understand a client's needs, sociocultural values and beliefs, and previous exposure to related ideas. Relevant generalizations: 9.1, 9.2, 9.3, 9.4.

Scoring: To simplify the strategy-selection algorithm, I forced it to use the *Personal Information* activity (for which the game awards no points) for the first five turns to gather information on all staff members. The descriptions obtained contain clues about adopter type, opinion leadership, and other characteristics that assist players in targeting individuals who will increase the rate of adoption. Many activities in the DSG require the player to gather this information before selecting a particular staff member. In addition, the next three turns used the *Talk To* activity with the gatekeepers—the Principal (A), the Secretary (B), and the Janitor (C) because some activities require this. After that, the decision to use the *Talk To* activity was calculated by dividing the number of staff members who had not yet participated in the activity by the total number of staff members.

Strategy 3: Use *Local Mass Media* and *Print* activities to gain points in the Awareness and Interest phases among earlier adopters.

Rationale: Mass media communication channels are especially effective in creating awareness and interest among earlier adopters, who tend to pay more attention to external sources of innovation than later adopters. Print is not exactly a mass medium in the DSG since its use is targeted at five staff members at a time; however, because earlier adopters tend to have more in common with change agents (Rogers says they are more

"homophilous"), they should be receptive to this approach. Relevant generalizations: 5.3, 5.5, 5.7, 5.13, 5.14, 5.15, 5.16, 7.20, 7.21, 7.22, 9.8, 9.9.

Scoring: The earlier adopters in the game are F, G, L, M, P, and X. The score for Strategy 3 is the number of these staff members divided by 6.

Strategy 4: Use the *Presentation* activity to gain points in the Awareness and Interest stages among earlier adopters.

Rationale: Because the change agent is giving the presentation, earlier adopters are more likely to be influenced. Nevertheless, once potential adopters (including later adopters) are aware of an innovation, they tend to be receptive to information about the innovation as they seek to reduce uncertainty about the consequences of adoption. Relevant generalizations: 7.21, 7.24, 9.9.

Scoring: The earlier adopters in the game are F, G, L, M, P, and X. The score for Strategy 4 is the number of these staff members divided by 6. Note that this score calculation is identical to that for Strategy 3. When these strategies have the highest score of all strategies, the strategy-selection algorithm randomly selects the *Local Mass Media*, *Print*, or *Presentation* activity.

Strategy 5: Use the *Demonstration* activity, especially by an opinion leader, to gain points in the Interest phase for other potential adopters.

Rationale: Allowing potential adopters to see the innovation in use provides how-to knowledge and can reduce uncertainty about the consequences of adoption, especially if the demonstration is by a respected peer. Relevant generalizations: 6.5, 8.2, 8.13, 9.11, 9.12.

Scoring: The *Demonstration* activity must be conducted by someone who is already an adopter. The score for Strategy 5 is the number of staff members of have at least one point in the Interest phase but still need points in that phase divided by 23 (total staff members minus the one doing the demonstration). The staff member selected to conduct the demonstration is one of the opinion leaders (F, H, or M), randomly selected if more than one is an adopter.

Strategy 6: Use the *Site Visit* activity to gain points in the Interest phase and move into the Trial phase.

Rationale: The strengths of this activity are similar to those of the *Demonstration* activity except the visitors see the innovation in use on a larger scale. The five selected visitors are likely to see the innovation used in classrooms similar to their own and be persuaded to try the innovation on a limited basis. Relevant generalizations: 6.5, 9.12.

Scoring: The score for Strategy 6 is the number of staff members who need points in the Interest phase divided by the total number of staff members.

Strategy 7: Use the *Pilot Test* activity to gain additional points for those with some points in the Interest stage or in the Trial stage.

Rationale: The ability to try the innovation on a limited basis can further reduce uncertainty about the consequences of adoption. If used with opinion leaders or other highly connected individuals, it may have a similar effect on others in their social networks. Later adopters are unlikely to be receptive to participating in a pilot test due to their lack of affinity with the change agent. However, Rogers says that trial by peers can serve as a vicarious trial for later adopters. Relevant generalizations: 5.13, 5.14, 6.4, 7.19, 8.2, 8.10, 8.11, 8.13, 9.11, 9.12.

Scoring: Later adopters and staff members without a classroom are excluded from consideration. The score for Strategy 7 is the number of remaining staff members who have at least one point in the Interest phase but are not yet adopters divided by 10.

Strategy 8: Target highly connected individuals to gain additional points in the Interest stage among later adopters in their social networks and move them into the Trial stage.

Rationale: Mass media channels are less effective at later stages of the innovation-decision process; instead, interpersonal channels can be quite effective in persuading someone to try an innovation as people tend to trust the opinions of their near peers. Relevant generalizations: 5.13, 8.2, 8.10, 8.13.

Scoring: Like Strategy 1, this strategy is independent of the activity used, and its score is added to the activity-based score. Six of the staff members have 10 or more interpersonal connections (F, G, H, J, V, and W). A score is calculated for each based on how many of their connections need points in the Interest phase. Depending on the activity selected, one or more of these scores are added to the activity-based score.

Strategy 9: Use the *Training Workshop (Self)* and *Materials Workshop* activities to gain points in the Trial stage.

Rationale: How-to knowledge is most essential when someone becomes willing to try an innovation. The change agent should provide knowledge and assistance regarding procedures and principles to further reduce uncertainty and increase confidence. Relevant generalizations: 6.3, 6.4, 9.2, 9.4.

Scoring: The score for Strategy 9 is the number of staff members who have at least one point in the Interest phase but are not yet adopters divided by the total number of staff members.

For each strategy described above, a score is calculated for each turn in a game. For strategies that are effective in raising awareness and interest, these scores are relatively large near the beginning of a game because all staff members first need a point in the Awareness phase before proceeding to the Interest phase; however, these scores progressively decrease as more staff members move into the Trial phase later in the game. Instead, the scores for activities that are more effective in the later phases of the innovation-decision process get larger. Adding the scores for Strategy 1 and Strategy 8 to the activity-specific strategies results in a strategy-selection algorithm that is optimized to target influential people.

I sought confirmation of these associations through expert review, which is sometimes referred to as face validity (Garson, 2009) or psychological validity (Kuppers & Lenhard, 2005; Peters et al., 1998), by searching the relevant literature to identify and then survey scholars who are knowledgeable about DOI theory. Appendix E contains the text of the email that I sent.

Measuring Success in the DSG

To answer the first research question regarding the effectiveness of APTMV in verifying a computational model, I needed to establish criteria for determining whether the fidelity of the DSG with respect to DOI theory was sufficient for its intended use. The DSG gives players 72 weeks to obtain 22 adopters. The time it takes to implement an activity ranges from 1 to 6 weeks. The number of points necessary to turn a particular

staff member into an adopter depends largely on his or her adopter type, with innovators requiring as few as 5 points and laggards as many as 14 points. The points are distributed across the Awareness, Interest, and Trial phases that lead to Adoption. Obtaining all 22 adopters requires 220 points (although this includes 5 points each for two gatekeepers, the Secretary and the Janitor, neither of whom can become an adopter). When measuring success in the DSG, the number of points obtained is arguably a better metric than the number of adopters obtained. To understand this, imagine a game in which the player obtained 8 adopters while the rest of the staff members were still in the Awareness or Interest stages. Compare this with a game in which the player obtained only 5 adopters while the rest of the staff members had moved through Awareness and Interest and were in the Trial stage. Overall the latter player gained many more points toward adoption even though fewer adopters were obtained.

The APTMV method prescribes specifying a threshold of confidence for determining the accuracy of a computational model. The feedback at the end of a game states that achieving 11 or more adopters is above average performance. The DSG was designed with a major objective of promoting recognition that "innovation diffusion is a complex, difficult, time consuming, and frustrating process (since even well-planned campaigns may result in less than total acceptance)" (Molenda & Rice, 1976, p. 461). In the pilot study of historical gameplay data (Enfield et al., 2012) the mean number of adopters was 13.08 (n=2,361) and the mean number of adoption points was 164.25. However, these figures may have been inflated by the programming errors described below.

Given the stochastic nature of situational simulations, along with the purposeful difficulty of the DSG, the definition of success in the game for model verification purposes should be reasonably broad. If we divide the total possible adoption points into quartiles and project a negatively skewed distribution given the use of optimal strategies, we might expect the majority of games to obtain at least 166 points and no games to score fewer than 110 points. Just how many games should fall into the third and fourth quartiles depends on the designers' goals, which in the case of the DSG include striking a balance between rewarding good choices and conveying the difficulty of change agentry. Designers using the APTMV method should decide on the desired distribution of scores given optimal strategies before undertaking formative evaluation. Table 1 shows my estimate of a desirable outcome given my understanding of the goals of the DSG.

Table 1
Projected Distribution of Games by Number of Adoption Points

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%

To address the second research question, which concerns the strengths of the proposed method in comparison with other available methods, I employed several other model verification methods in addition to APTMV. The decision regarding which model verification methods to use is a matter of judgment. Factors to be considered include the type of simulation; the time, money, and human resources available; the simulation objectives and acceptability criteria; and the consequences of modeling errors (Sargent,

2008; Youngblood, 2006). The methods are usually chosen to balance the strengths and weaknesses of each method.

The first verification method used was an individual walkthrough that involved a line-by-line review of the original DSG code. The designers of the online DSG had little experience with the PHP language, so I expected to find errors that affected scoring. I have over 20 years of experience in programming with various languages, including several years of experience with PHP. Indeed, I found numerous errors (see Appendix F). By correcting and documenting these errors, I ensured to the best of my ability that the computational model was working as intended before running the roboplayed games.

I spent several weeks writing, debugging, and refining the strategy-selection algorithm until I obtained satisfactory results. Generally this involved running the program with execution tracing to play from 5 to 20 games and examining the results. This verification method also enabled me to test the original DSG code. I periodically used desk checking to review my code for consistency and to update the documentation of my code. In addition, I used Adobe Dreamweaver software to write my code, in part because it provides automatic syntax checking for PHP. To further reduce the chance of programming errors, I separately wrote formulas in Microsoft Excel to calculate the strategy scores and then imported the roboplay data to compare with the PHP calculations. When I found discrepancies between the strategy-selection algorithm and Excel score calculations, I tracked down the cause and corrected the code.

The final verification method used was statistical analysis of game data. The previous study (Enfield et al., 2012) employed independent samples *t* tests to compare successful games (achieving all 22 adopters) with unsuccessful games (achieving fewer

than 16 adopters). The current study used descriptive statistics such as counts, percentages, and means as well as t tests to compare means between groups of games to determine whether changes to the computational model made statistically significant differences in game outcomes.

Throughout data collection and analysis, I kept a record of my process including problems encountered, solutions attempted, and ideas for improving APTMV. These notes provided the raw material for addressing the third research question regarding potential improvements to the procedure.

CHAPTER 5: RESULTS

Research Question 1: Is the proposed method effective in verifying and improving the accuracy of computational models created for simulations and games?

Two approaches are necessary in order to answer the first research question. The first approach is to examine games in which optimal strategies were employed. If the computational model is sufficiently accurate, the majority of games should achieve a high number of adoption points and satisfy the threshold of confidence. If they do not, further analysis will be necessary to identify the discrepancies between the computational model and the conceptual model. The first part of this chapter examines roboplayed games in which a strategy-selection algorithm calculated scores for optimal strategies based on DOI theory and employed the highest scoring strategy.

The second approach is to examine games in which non-optimal strategies were employed. Non-optimal strategies may be devised by an imaginative programmer or they may result from a randomized strategy-selection algorithm. For example, a programmer might employ only strategies that are not predicted to work based on the conceptual model; in the case of the DSG, an example would be the use of mass media communication channels to influence laggards. Another approach might be to write a program that seeks the best strategies in the game through trial-and-error. In the case of the DSG, such a program would at first randomly select diffusion activities and staff members, and it would keep track of game states and results. Over time, it would begin to use strategies that previously resulted in a high number of adoption points when it encountered a similar game state.

In this study a third option was chosen because thousands of DSG game sessions were available for analysis. Given the range of outcomes in these games, I assumed that a sufficient number of players used strategies that were not aligned with DOI theory. For games that resulted in high scores, further analysis was necessary to determine whether this outcome was due to the use of optimal strategies, a problem with the computational model, or a flaw in the conceptual model. As described above in the literature review, simulations are sometimes used to discover new relationships and principles. Given an accurate computational model, a non-optimal strategy that proved successful may suggest that further research is needed in the real world to revise the conceptual model.

Analysis of Roboplayed Games

As described above, the measure used to verify the accuracy of the computational model was the percentage of roboplayed games that fell into the fourth quartile of adoption points. I divided the 220 possible adoption points into four quartiles and decided that, based on my understanding of the designers' goals for the DSG, at least 60% of games that used optimal strategies should fall into the fourth quartile (Table 2).

Table 2
Projected Distribution of Games by Number of Adoption Points

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%

I began testing by running the strategy-selection program to play 100 games (Table 3) to check for obvious anomalies. The results were reasonable although I was surprised that no games resulted in 220 points.

Table 3

First Dataset: Descriptive Statistics for Games (n=100)

Variable	Mean	SD	Min	Max
Turns	33.79	2.59	28	40
Total Points	148.74	17.51	96	187
Total Adopters	14.67	2.28	10	19

I decided to examine these data in more detail because as a simulation designer conducting formative evaluation, my next step would be to look for unusual results such as low point scores for strategies. But first I checked the distribution of adoption points to determine whether the threshold for model verification was met (Table 4). The majority of games (82%) fell into the third quartile, while only 17% fell into the fourth quartile.

Table 4

First Dataset: Distribution of Roboplayed Games by Number of Adoption Points (n=100)

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%
First Dataset	0%	1%	82%	17%

These results suggested to me that some DOI strategies were not being as effective as predicted and hence the computational model could be improved. Because most of the strategies are associated with particular activities, I decided to examine the number of points being awarded for each activity (Table 5).

Table 5

First Dataset: Frequency and Scores of Activities for 100 Roboplayed Games

		Turns	
Activity	n	%	M Points
Personal Information	500	14.8%	0.0
Talk To	1153	34.1%	1.8
Local Mass Media	176	5.2%	8.4
Print	171	5.1%	2.4
Presentation	185	5.5%	11.0
Demonstration	404	12.0%	3.5
Site Visit	78	2.3%	8.3
Pilot Test	139	4.1%	0.3
Training Workshop (Self)	299	8.8%	8.0
Materials Workshop	274	8.1%	16.0
Grand Total	3379	100.0%	4.4

I was not surprised to find that *Talk To* was the most frequently used activity nor that it had one of the lowest mean scores per turn. It is designed to achieve points mostly in the Awareness stage, and in some cases (e.g., when Laggards are targeted) no points are scored, but according to the feedback text their antagonism toward the innovation has been diffused.

For the *Demonstration* activity, the high frequency of use (12% of all turns) combined with a relatively low mean score (3.5 points per turn) warranted further examination. The low mean scores for the *Print* and *Pilot Test* activities also suggested problems. The high mean score for the *Materials Workshop* activity also warranted investigation.

Examining the original DSG code for the *Demonstration* activity revealed a discrepancy in my implementation of DOI Strategy 5. This illustrates the benefit of having someone outside the design team lead the verification effort. My interpretation of DOI theory stated that once potential adopters were aware of the innovation, a demonstration by an opinion leader would increase their interest. In my code, I required

staff members to have completed the Awareness phase, but the original DSG code

additionally required that they have at least one point in the Interest phase. My approach

widened the opportunity for using the *Demonstration* activity, but the original DSG code

did not award points for the additional use. So which approach should be considered

correct? I consider it a matter of design judgment, but I think judgment can and should

utilize the available information. In this case, I ran the program two more times with

slight adjustments. The first time I changed my implementation of DOI Strategy 5 so that

it also required at least one point in the Interest phase, making it consistent with the

original DSG code. The second time I changed the DSG code so that it required only

completion of the Awareness phase, making it consistent with my original strategy. Using

this approach, I was able to assess the impact of the design decision as it related to the

design goals. Table 6 shows the results of these changes.

Table 6
Second Dataset: Frequency and Scores of Activities for 100 Games

Activity	DOI Strategy Changed			Original DSG Changed		
	n	%	M Points	n	%	M Points
Personal Information	500	14.9%	0.0	500	15.3%	0.0
Talk To	1167	34.8%	1.8	1159	35.5%	1.7
Local Mass Media	206	6.1%	7.6	190	5.8%	8.0
Print	207	6.2%	2.3	181	5.5%	2.3
Presentation	175	5.2%	10.9	179	5.5%	11.1
Demonstration	69	2.1%	7.2	143	4.4%	24.2
Site Visit	138	4.1%	10.1	79	2.4%	8.5
Pilot Test	206	6.1%	0.5	102	3.1%	0.6
Training Workshop (Self)	387	11.5%	7.9	359	11.0%	9.1
Materials Workshop	301	9.0%	15.0	370	11.3%	15.2
Grand Total	3356	100.0%	4.6	3262	100.0%	5.8

Changing my strategy for the *Demonstration* activity to require at least one point

in the Interest phase caused it to become the least used activity (from 12% to 2.1%), and

its mean score per turn went from 3.5 to 7.2, making it more in line with most other

activities. At the same time, the change resulted in some differences for other activities,

most noticeably those that are commonly used after the *Demonstration* activity when

staff members are in the Interest and Trial phases. The use of the *Site Visit* activity nearly

doubled (from 2.3% to 4.1%) and its mean score per turn increased from 8.3 to 10.1. On

the other hand, changing the original DSG code to require only completion of the

Awareness phase caused a sevenfold increase (from 3.5 to 24.2) in the mean score for the

Demonstration activity. This dramatic increase was balanced by the fact that its

frequency of use was 4.4% instead of the original 12%.

Next I examined the effects of these changes on the distribution of adoption

points to determine whether the threshold for model verification was met (Table 7). In

both cases the majority of games fell into the fourth quartile, although changing the DOI

strategy to match the original DSG code (with both requiring at least one point of

Interest) resulted in fewer games than desired (53%) while changing the DSG code to

require only Awareness resulted in 85%.

Table 7
Second Dataset: Distribution of Games by Number of Adoption Points

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%
DOI Strategy Changed	0%	0%	47%	53%
Original DSG Changed	0%	0%	15%	85%

I should note that even though the DSG is an educational simulation, by

roboplaying games to test the accuracy of the computational model, I was in essence also

using it as an experimental simulation. If the computational model is well designed, unpredicted results might indicate potential improvements to the *conceptual* model. Such an outcome in a simulation would need to be confirmed through empirical research in the real world. Furthermore, I needed to bear in mind the affective goal of the DSG to convey the difficulty of diffusing an innovation among members of a system. By changing the game's criteria for successfully using the *Demonstration* activity, the percentage of games in the top quartile was much higher than the established threshold of confidence, which suggested to me that it was now easier than the designers intended.

Given the systemic nature of games in general and the DSG in particular, in which small changes can lead to significant differences in results, I opted to examine the other potential problems that I identified in the results of the first dataset before deciding how the *Demonstration* activity ought to be implemented. The supplementary materials for the DSG state that the *Print* activity, which requires the selection of up to five staff members, is "rather weak, useful mainly for arousing Awareness" (Molenda, 1976, p. 7). My DOI Strategy 3 says that it should raise Awareness or Interest when used with Innovators and Early Adopters. According to the original DSG code and feedback text, *Print* usually results in one point in Awareness or Interest for a few of the five selected staff members, regardless of adopter type. Limiting its effectiveness to earlier adopters would further reduce its impact, yet it seems unlikely that a brochure would generate interest among later adopters. However, as one of the expert reviewers pointed out, the effectiveness of the *Print* activity may be influenced by the attributes of the particular innovation and the context in which the change agent is working. Nevertheless, the APTMV method identified this as an aspect of the game that may need improvement.

Next I focused on the *Materials Workshop* activity to see why it was so successful and whether it might require modification. The supplementary materials for the DSG state that this activity, which does not target particular staff members, "consistently gives a high payoff for those in the latter phases" (Molenda, 1976, p. 7). This is consistent with my strategy-selection algorithm based on DOI theory in which this activity scores highly when staff members are in the Interest or Trial phase (Strategy 9). In the original DSG code, scores for this activity are increased significantly (often doubled) if the Principal is already an adopter. This is an example of a context-specific adjustment to DOI theory based on empirical research regarding the influence of principals on adoption rates in schools (Molenda, 1976).

The final aspect to examine was the *Pilot Test* activity (Strategy 7), which had a much lower mean score per turn (0.3) than other activities. The supplementary materials for the DSG state that the *Pilot Test* activity "could be used to arouse Interest for all but the foot draggers … but payoffs are moderate" [ellipsis in the original] (Molenda, 1976, p. 7) *Pilot Test* is an activity for which one staff member is selected. In writing the strategy-selection algorithm, I excluded Laggards because DOI theory says they are not likely to be receptive to direct appeals from a change agent. I also excluded staff members who don't have a classroom to conduct a pilot test. Of the remaining staff members, two stood out as anomalies. The Social Studies Chairwoman (J), whose adopter type is Late Majority but who has a large social network (Strategy 8), was selected for 64 of 139 turns (46%) but scored a total of only 4 points; and a Science Teacher (H), whose adopter type is Early Majority and who is an opinion leader with a large social network, was selected for 22 of 139 turns (15.8%) and scored no points.

Together they comprised nearly 62% of the turns that used *Pilot Test* yet they accounted

for only 4 of the 38 points scored (10.5%).

The original DSG code for the results of the *Pilot Test* activity puts the staff

members into five groups with a different set of possible results for each group (Figure

5).
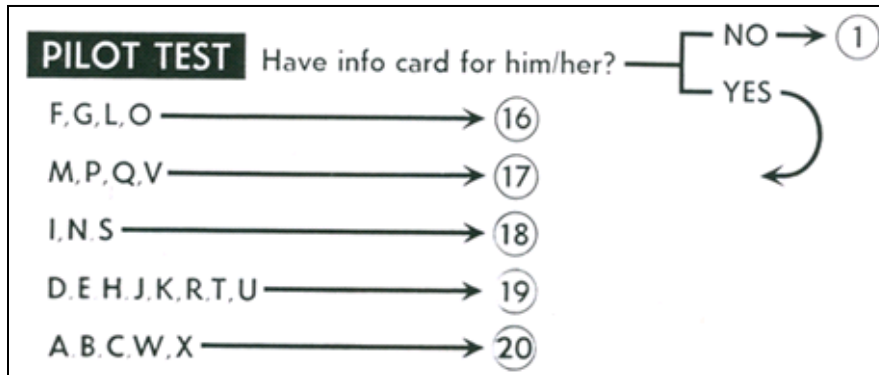
Figure 5. Groupings of staff members for determining the results of *Pilot Test* activity.

The largest group contains the Laggards and several of the Late Majority (including J, the

Social Studies Chairwoman), but it also contains the Science Teacher (an Early

Majority), which explained her dearth of points for this activity. I contacted one of the

original designers of the DSG, Dr. Michael Molenda, and he agreed that she seemed out

of place in that group (personal communication, March 7, 2012). I decided that she would

be better placed in a group with other Early Majority staff members, and I made a couple

of other changes to the groups so that the Innovators and Early Adopters were together as

were most of the Early Majority. I also changed the strategy-selection algorithm to make

Late Majority undesirable for selection.

Dr. Molenda confirmed that the original designers intentionally required at least

one point in the Interest phase when scoring the *Demonstration* activity (personal

communication, March 7, 2012). I decided to continue my testing using the original DSG

code for *Demonstration* (requiring at least one point in Interest) and changing my

strategy-selection algorithm accordingly. I also added the Principal to the group of

opinion leaders who are specifically targeted in several strategies.

A final run of 500 games (Table 8) found no significant difference in the mean

number of turns per game between the first dataset and this third dataset, $t_{.05}(598) =$

0.735, p=0.463, which was expected. However, there was a significant difference in mean

total points, $t_{.05}(598) = 10.839$, p<0.000, and in mean total adopters, $t_{.05}(598) = 7.252$,

p<0.000, both of which indicated that the modifications to the game described above

improved game scores when optimal strategies were used.

Table 8
Comparison of Descriptive Statistics for Games

Variable	First Dataset (n=100)				Third Dataset (n=500)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Turns	33.79	2.59	28	40	34.01	2.76	28	44
Total Points	148.74	17.51	96	187	170.72	18.70	103	220
Total Adopters	14.67	2.28	10	19	16.54	2.37	6	22

Furthermore, the results were very close to satisfying the threshold for model

verification that was specified at the beginning of analysis, with the percentage of games

in the fourth quartile exceeding the 60% target (Table 9).

Table 9
Third Dataset: Distribution of Games by Number of Adoption Points (n=500)

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%
Third Dataset	0%	0.4%	38.6%	61.0%

The frequency of activity use and the mean scores per turn also showed some desired improvements (Table 10). The use of the *Demonstration* activity decreased from 12% to 8% while its mean score per turn rose from 3.5 to 9.99. The *Pilot Test* activity still had the lowest mean score per turn, but at 1.66 it was much higher than 0.3 in the first dataset and was now comparable to other activities such as *Print* and *Talk To* that were designed to provide only small gains.

Table 10
Comparison of Activity Selection by Frequency and Average Scores

	First Dataset			Third Dataset		
Activity	n	%	M Points	n	%	M Points
Personal Information	500	14.8%	0.0	2500	14.7%	0.00
Talk to	1153	34.1%	1.8	6526	38.4%	1.69
Local Mass Media	176	5.2%	8.4	915	5.4%	7.81
Print	171	5.1%	2.4	859	5.1%	2.43
Presentation	185	5.5%	11.0	950	5.6%	10.46
Demonstration	404	12.0%	3.5	1354	8.0%	9.99
Site Visit	78	2.3%	8.3	717	4.2%	9.32
Pilot Test	139	4.1%	0.3	172	1.0%	1.66
Training Workshop (Self)	299	8.8%	8.0	1545	9.1%	8.51
Materials Workshop	274	8.1%	16.0	1467	8.6%	14.68
Grand Total	3379	100.0%	4.4	17005	100.0%	5.02

Analysis of Replayed Games

For comparison purposes, Table 11 shows the descriptive statistics for both the roboplayed and replayed games. It is important to keep in mind that when the historical games were played (and replayed), the DSG code contained numerous errors that resulted in extra points being awarded. Also, because the replayed games consisted only of *finished* games, they did not include games that may have been aborted by players near the end due to low scores.

Table 11
Comparison of Descriptive Statistics for Games

Variable	Roboplayed Games (n=500)				Replayed Games (n=2361)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Turns	34.01	2.76	28	44	45.44	8.72	18	84
Total Points	170.72	18.70	103	220	164.25	39.25	41	220
Total Adopters	16.54	2.37	6	22	13.53	5.48	1	22

I also examined the distribution of adoption points to see in which quartiles the historical games fell (Table 12). Nearly half the games (47.65%) fell into the fourth quartile, and I examined these games in more detail to see whether any unexpected strategies led to success.

Table 12
Distribution of Games by Number of Adoption Points

	Adoption Points			
	0 – 55 pts.	56 – 110 pts.	111 – 165 pts.	166 – 220 pts.
Verification Targets	0%	0%	<40%	>60%
Roboplayed Games (n=500)	0%	0.4%	38.6%	61%
Replayed Games (n=2361)	0.04%	9.19%	43.12%	47.65%

First I compared activity selection to see if any activities in the top quartile of replayed games were more successful than expected (Table 13).

Table 13
Top Quartile Games: Comparison of Activity Selection by Frequency and Scores

Activity	Turns in Roboplayed Games (n=305)			Turns in Replayed Games (n=1125)			t Test (α=.005) (DF) t, p, d
	n	%	M Points	n	%	M Points	
Personal Information	1525	14.5%	0.000	5009	10.7%	0.000	
Talk To	4188	39.8%	1.723	17115	36.6%	2.262	(21301) -12.3, <0.000, -0.21
Local Mass Media	572	5.4%	8.047	3309	7.1%	6.094	(3879) 8.033, <0.000, 0.36
Print	460	4.4%	2.424	3425	7.3%	4.535	(3883) -35.304, <0.000, -1.75
Presentation	567	5.4%	10.903	1705	3.6%	6.807	(2270) 13.527, <0.000, 0.66
Demonstration	892	8.5%	10.450	2377	5.1%	13.892	(3267) 6.682, p<0.000, 0.26
Site Visit	444	4.2%	9.592	3562	7.6%	11.583	(4004) -7.449, <0.000, -0.37
Pilot Test	90	0.9%	1.678	2038	4.4%	1.758	(2126) -0.331, 0.741, -0.04
Training Workshop Self	847	8.1%	10.004	1309	2.8%	12.879	(2154) 5.42, <0.000, -0.24
Materials Workshop	926	8.8%	15.336	1521	3.3%	21.917	(2445) 10.51, <0.000, -0.44
Subtotal	10511	100.0%	5.282	41370	88.4%		
DIAGRAMS							
Social				399	0.9%	0.000	
Committees				402	0.9%	0.000	
Lunchmates				463	1.0%	0.000	
UNDESIRABLE							
Training Workshop Prof				1058	2.3%	4.455	
Ask Help				2888	6.2%	1.083	
Compulsion				79	0.2%	9.684	
Confrontation				138	0.3%	16.551	
Grand Total	10511	100.0%	5.282	46797	100.0%	4.807	

Note that the replayed games included turns in which activities were selected that were not included in the strategy-selection algorithm for the roboplayed games. The

Social, *Committees*, and *Lunchmates* diagrams in the DSG show relationships among the staff members. The player must spend two weeks to view the *Social* diagram while the other two diagrams cost only one week; once these diagrams have been obtained, they may be viewed by the player at any time during the game. In the Enfield et al. (2012) study, we found that these diagrams were less likely to be viewed in the more successful games, and we suspected that these players had already played the DSG several times and were familiar with the diagrams. These relationships among staff members are addressed by Strategy 8, which seeks to target highly connected individuals in the hope that they will increase interest among their peers. I chose not to include the selection of these as part of the roboplayed games because they would have only a minimal impact on the number of weeks used in the games.

I excluded the *Training Workshop (Prof)*, *Ask Help*, *Compulsion*, and *Confrontation* activities from roboplayed games because I could find no rationale for their use based on my understanding of DOI theory. However, the mean points per turn were higher than I expected for all but the *Ask Help* activity, suggesting to me that these activities may merit further investigation.

Independent samples *t* tests (Table 13) found significant differences between mean points per turn for all of the remaining activities except *Pilot Test*. However, most of the effect sizes (Cohen's *d*, a measure of differences in terms of standard deviation based on pooled variances) were small, with the exceptions being *Presentation* (0.66) and *Print* (-1.75). I decided to focus on the activities in the replayed games that had both a relatively high frequency of use and high mean points per turn. The *Site Visit* activity was used 3,562 times (7.61% of turns) and scored an average of 11.583 points, making it the

most successful activity in terms of total points (18.34% of all points in replayed games). The second most successful activity was *Talk To*, which was used less frequently than in roboplayed games yet had a higher mean points per turn, resulting in 17.21% of all points in replayed games. Similarly, the *Materials Workshop* activity was used less frequently than in roboplayed games yet had significantly higher mean points per turn, garnering 14.82% of all points in replayed games. Combined, these three activities accounted for nearly half (47.84%) of all points in replayed games.

The *Site Visit* activity, in which five staff members are selected to visit a nearby school where peer tutoring is being used, is specified in Strategy 6 as a way to increase interest among potential adopters by allowing them to see the innovation in use, thereby providing how-to knowledge and reducing uncertainty. The original DSG code contains four possible outcomes, all of which award from 5 to 20 points including additional points if the Principal is already an adopter. Other than the Principal, the activity does not reward the use of opinion leaders. One outcome awards additional points to the six members of the curriculum committee. This outcome also awards one point for three of the five visitors with an additional two points for each of these three visitors if the Principal is already an adopter. However, the code was written so that all five visitors received points, resulting in from two to six additional points added erroneously. I fixed this error before running the roboplayed games, but it impacted the results of the replayed games. The flawed outcome was randomly selected 26% of the time that *Site Visit* was used (928 out of 3,562 turns). However, not all points were always awarded. If, for example, a selected staff member or curriculum committee member was already an adopter, he or she would receive no points. Using the available data, I calculated that an

additional 2,808 points had been erroneously awarded. Subtracting these from the 41,258

total points awarded for *Site Visit* reduced the mean points per turn to 10.794, which is

closer to the 9.592 in the roboplayed games but still significantly different, $t_{.005}(4004) =$

4.792, $p<0.000$, with a small effect size of -0.24 (Cohen's *d*). However, it is impossible to

know what choices players might have made and how the overall game results would

have changed had they not received those additional points.

The *Talk To* activity requires the selection of one staff member. It has 45 possible

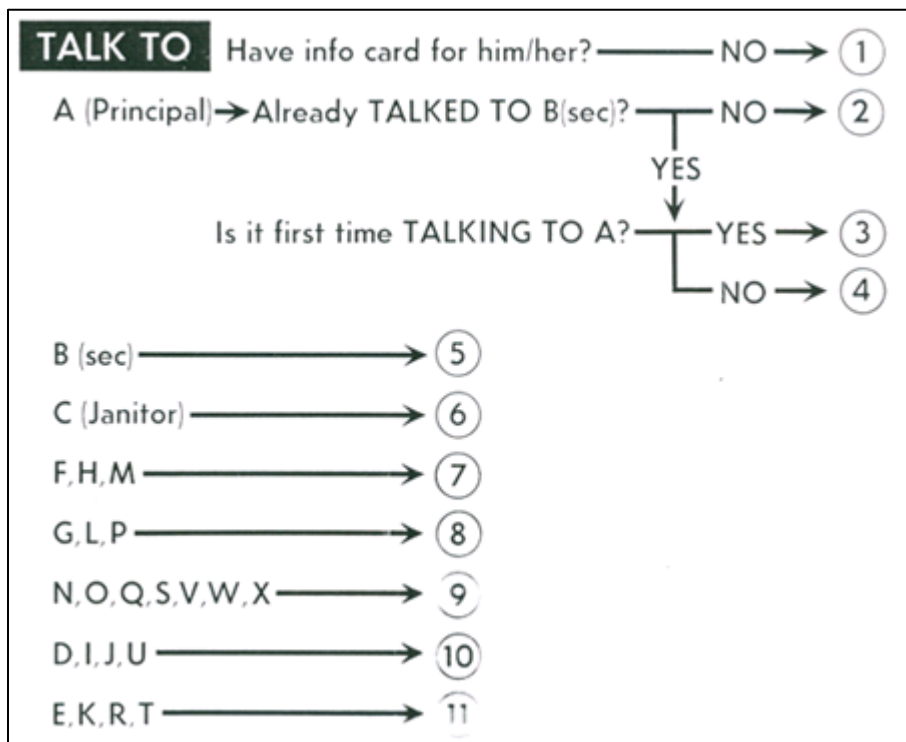outcomes that comprise 11 groups of outcomes (Figure 6).

Figure 6. Groupings of staff members for determining the results of the *Talk To* activity.

The Principal alone has 10 possible outcomes divided into three groups. The Secretary

and the Janitor, who are gatekeepers to resources, have four possible outcomes each. The

rest of the outcomes are essentially grouped by adopter type, with the opinion leaders (F,

H, and M) forming their own group. Given the variety of outcomes for this activity, I

decided first to examine the selection of staff members (Table 14).

Table 14
Top Quartile Games: Comparison of Staff Selection for the Talk To Activity by Frequency and Scores

Staff	Turns in Roboplayed Games			Turns in Replayed Games			Diff
	n	%	M Points	n	%	M Points	
A	305	7.28%	1.000	3514	20.53%	1.226	0.226
B	305	7.28%	2.505	1502	8.78%	2.214	-0.291
C	305	7.28%	2.498	1274	7.44%	2.272	-0.226
D	192	4.58%	1.219	654	3.82%	1.326	0.107
E	178	4.25%	2.494	952	5.56%	3.639	1.144
F	305	7.28%	1.607	1010	5.90%	2.677	1.071
G	304	7.26%	0.898	809	4.73%	3.227	**2.329**
H	305	7.28%	3.449	567	3.31%	3.344	-0.105
I	125	2.98%	1.896	381	2.23%	1.556	-0.340
J	304	7.26%	1.382	424	2.48%	1.078	-0.304
K	100	2.39%	2.930	677	3.96%	3.879	0.949
L	305	7.28%	0.623	364	2.13%	2.997	**2.374**
M	69	1.65%	0.870	885	5.17%	3.586	**2.717**
N	69	1.65%	1.710	328	1.92%	1.872	0.162
O	57	1.36%	1.123	291	1.70%	1.832	0.709
P	45	1.07%	0.311	352	2.06%	2.943	**2.632**
Q	54	1.29%	0.537	415	2.42%	1.723	1.186
R	56	1.34%	2.696	454	2.65%	2.808	0.112
S	55	1.31%	1.655	257	1.50%	1.763	0.108
T	44	1.05%	2.705	547	3.20%	3.461	0.756
U	52	1.24%	0.846	193	1.13%	1.269	0.423
V	305	7.28%	1.993	366	2.14%	1.710	-0.283
W	305	7.28%	1.469	469	2.74%	1.565	0.096
X	44	1.05%	0.114	430	2.51%	1.363	1.249
Grand Total	4188	100.00%	1.723	17115	100.00%	2.262	0.540

The staff members are referred to by the letters used in the game to identify them,

with A being the Principal, B the Secretary, and C the Janitor. Four of the staff members

stood out as anomalies because of the large differences in their mean points per turn

between the roboplayed and replayed games: the Science Chairman (G), a Social Studies

Teacher (L), the Language Arts Chairman (M), and a Foreign Language Teacher (P).

The Language Arts Chairman is an Early Adopter and opinion leader, while the other

three are Innovators. I isolated these four staff members and analyzed the outcomes when

they were selected in the games (Table 15).

Table 15
Top Quartile Games: Comparison of Outcomes by Frequency and Scores when Talk To was Selected

| | Roboplayed Games | | | | | | | |
| | G | | L | | M | | P | |
Feedback	Turns	M Points	Turns	M Points	Turns	M Points	Turns	M Points
1								
7a					16	2.438		
7b					23	0.565		
7c					21	0.381		
7d					9	0.000		
8a	64	0.078	57	0.088			2	1.500
8b	53	0.019	64	0.125			13	0.231
8c	75	3.493	56	2.750			7	0.286
8d	49	0.041	58	0.241			10	0.600
8e	63	0.048	70	0.129			13	0.000
Grand Total	304	0.898	305	0.623	69	0.870	45	0.311

| | Replayed Games | | | | | | | |
| | G | | L | | M | | P | |
Feedback	Turns	M Points	Turns	M Points	Turns	M Points	Turns	M Points
1	30	0.000	30	0.000	21	0.000	16	0.000
7a					251	6.502		
7b					220	4.068		
7c					200	3.235		
7d					193	0.000		
8a	156	2.500	70	2.400			60	2.400
8b	146	2.322	64	2.234			66	2.470
8c	158	6.247	68	5.956			67	4.358
8d	156	2.359	67	2.403			68	2.485
8e	163	3.233	65	3.292			75	3.573
Grand Total	809	3.227	364	2.997	885	3.586	352	2.943

For staff members G, L, and P, I reviewed the feedback text for the five outcomes (8a through 8e). The text for 8e says that the selected staff member receives all necessary points to become an adopter, while the text for the other four outcomes says to award the staff member three points. The text for 8c additionally awards one point to the staff member's lunchmates. Because the mean points per turn in roboplayed games was substantially lower than expected based on the feedback text, I examined each staff member's stage in the innovation-decision process during the turns in which they were selected. I found that G, L, and P were already adopters over 90% of the time when there were selected for the *Talk To* activity. Therefore they could not receive the points that were awarded to them. On the surface, this suggests that the strategy-selection algorithm might be improved so that adopters would not be selected for the *Talk To* activity. However, many activities require the use of *Talk To* with a staff member before he or she can be selected.

I used the same approach described above for staff member M, who has a different group of outcomes. Other than 7d, which awards no points, all of the feedback texts award either one or two points to the staff member and an additional one or two points to each of his or her lunchmates (7a) or social contacts (7b, 7c). Here too, in roboplayed games M was already an adopter 88% of the time and so could not receive points. Furthermore, his two social contacts were almost always adopters or near adoption. While analyzing these data, I noticed that two outcomes (7e and 7f) did not appear in my summaries. A query of the databases for both roboplayed and replayed games revealed that these outcomes had never been selected. I reviewed the original DSG

code and found that the function for randomly selecting from this group of outcomes had an error that excluded these two outcomes.

The *Materials Workshop* activity has seven possible outcomes. Two of these award no points because either the Principal or the Janitor has not yet been selected for the *Talk To* activity. The feedback texts for the remaining five outcomes say to double the points awarded if the Principal is already an adopter. I figured this might explain the difference in mean points per turn between roboplayed and replayed games. In roboplayed games, the Principal was an adopter 45.4% of the time when *Materials Workshop* was selected, while in replayed games he was an adopter 64.1%. This explains some of the difference between the mean points per turn. However, when I looked at only those turns in which the Principal was an adopter, I still found large differences (Table 16).

Table 16
Top Quartile Games: Comparison of Outcomes when Materials Workshop was Selected and the Principal was an Adopter

	Roboplayed Games			Replayed Games		
Feedback	Turns	%	M Points	Turns	%	M Points
32				16	1.64%	0.000
33				23	2.36%	0.000
36a	116	27.62%	8.690	191	19.59%	12.613
36b	98	23.33%	19.867	282	28.92%	36.404
36c	110	26.19%	16.155	262	26.87%	33.889
36d	96	22.86%	5.719	201	20.62%	7.393
Grand Total	420	100.00%	12.574	975	100.00%	23.631

The outcomes for 36b and 36c award the highest points, and the feedback texts for these indicate that two points are awarded (doubled to four because the Principal is an adopter) to every staff member in the Interest phase (36b) and the Interest or Trial phase (36c). Therefore the difference in points awarded must be attributable to the fact that

more staff members were in those phases when *Materials Workshop* was selected in replayed games. An analysis of these turns (Table 17) revealed that for 36b, only 12.54% of the staff members were in the Interest phase for roboplayed games compared with 29.47% for replayed games. For 36c, a total of 26.10% of staff members were in the Interest or Trial phase in roboplayed games while replayed games had 48.51%. Instead, the roboplayed games had more staff members in earlier phases, with 25.75% for 36b and 23.29% for 36c versus replayed games with 6.47% and 8.8%.

Table 17
Top Quartile Games: Comparison of Staff Members' Innovation-Decision Phase when Materials Workshop was Selected and the Principal was an Adopter

Phase	Roboplayed Games		Replayed Games	
	36b	36c	36b	36c
[No Phase]	15.89%	14.94%	2.04%	2.53%
Awareness	9.86%	8.35%	4.42%	6.27%
Interest	**12.54%**	**8.53%**	**29.47%**	**26.70%**
Trial	17.83%	**17.58%**	22.29%	**21.81%**
Adoption	43.88%	50.61%	41.78%	42.69%
Grand Total	100.00%	100.00%	100.00%	100.00%

The explanation for this is complicated. The score calculated by the strategy-selection algorithm for the *Materials Workshop* activity (Strategy 9) is based on the percentage of staff members who have at least one point in the Interest phase but still need points in either the Interest or Trial phase. At the same time, Strategy 1 targets earlier adopters as much as possible, so that they reach the Interest phase well before other staff members. This resulted in the *Materials Workshop* activity being selected earlier in the roboplayed games (the mean turn is 20.88) than in replayed games (the mean turn is 28.83), $t_{.05}(1402)=-14.34$, $p<0.000$, with a large effect size (Cohen's *d*) of -0.93.

Together, the differences in frequency of use when the Principal was an adopter and the number of staff members eligible to receive points accounted for the higher mean points per turn for the replayed games. This suggests a potential improvement to the strategy-selection algorithm. One approach would be not to let later adopters fall too far behind by targeting them once a certain percentage of staff members had reached the Interest phase. This seems reasonable from the perspective of DOI theory because activities that don't target particular individuals should be more effective if more people are prepared to engage in them. Rogers recommends trying to achieve a critical mass of adopters so that further adoption becomes somewhat self-sustaining, but perhaps a corollary should be not to lose focus of the more resistant members of the system.

Research Question 2: What does the proposed method contribute that is not available through related methods?

Comparing APTMV with other verification methods was a subjective task. Using multiple methods of verification is a recommended practice, but the choice of methods depends on many variables, including the context, the content, and the knowledge and preferences of the person(s) conducting the verification. I began with two programming goals: first, to adapt the DSG code so that I could replay the historical games and generate the missing data necessary for an analysis of gameplay patterns; second, to write a program that would use the adapted code to play new games using optimal strategies based on DOI theory. Accomplishing these goals required careful inspection of both the DSG code and my additional code, and this involved an iterative cycle of coding, testing, analyzing, and debugging.

In addition to APTMV, I used several methods (often simultaneously) to audit my code and verify the computational model. These methods, which are described in the literature review, were: walkthrough, execution tracing, desk checking, syntax checking, and statistical analysis. The first four methods focused mostly on the structure and accuracy of the code. Because the ultimate purpose of the roboplay code was essentially to instantiate a separate computational model of DOI theory with which to test the DSG's model, these methods were critical in reducing the chance of errors in my code. I also established face validity by asking experts in DOI theory to review the strategies that were the basis for my computational model. APTMV, sometimes used in conjunction with statistical analysis, was used to examine the results of turns and outcomes of games. However, anomalies in the results always led back to examination of the code to determine whether they were caused by coding errors or model design flaws.

Writing the code to replay the historical games required a thorough understanding of the original DSG code, so I began with an individual walkthrough of the code to become familiar with the structure, functions, and variables. I examined the code using Adobe's Dreamweaver software, which does automatic syntax checking for PHP and thereby made it easier to identify programming errors. The DSG was structured such that every activity had its own Web page containing hundreds of lines of redundant PHP, HTML, and CSS code. Replaying the historical games did not require displaying these pages; instead I only needed the functions for calculating scores and updating records. Therefore I created a single page of PHP code with a function for each activity, and I reduced all of the redundant code to a set of functions that could be called to handle input, calculate scoring, and write results to both XML files and a database. These

functions retained as much of the original code as possible because I needed to duplicate the results of the historical games, errors and all. Additional functions retrieved original gameplay information and sequentially passed it to the game engine in order to replay every turn of every game. During the process of writing and testing this code, I found most of the coding errors that would comprise my "bug report" (see Appendix F); however, some significant errors did not show up until I began analyzing the data from roboplayed games. I discovered programming errors of the following types:

- Typographic errors: These were usually incorrectly spelled variable names, such as "$tiral" instead of "$trial" and "$totoalpoints" instead of "$totalpoints."

- Code syntax errors: These contained invalid combinations of characters according to the programming language's rules, such as omitting the index in an array reference (e.g., "$interest" instead of "$interest[$i]" to specify a particular element in the $interest array).

- Semantic errors: These were syntactically correct yet produced the wrong results without crashing the program. These usually awarded the wrong number of points or awarded points to the wrong staff member; they were found by comparing the code with the feedback text for every possible turn result.

- Logic errors: These were a special type of semantic error involving incorrect logical operators. For example, one turn result specified awarding points only if the selected staff member was not the Secretary OR was not the Janitor, which should have been not the Secretary AND not the Janitor.

Some of these errors explained unexpected findings in the Enfield et al. (2012) study. For example, in that study we found that games in which players achieved all 22

adopters were 4.6 times more likely to use the *Print* activity. I discovered that the code

for every possible result of using the *Print* activity contained semantic errors that

awarded more points than intended. Savvy players may have recognized and exploited

this flaw. These errors were not corrected prior to replaying the historical games as that

would have made it impossible to match the scoring and results of original gameplay

decisions. All known errors were fixed before running the roboplay program using

optimal strategies so that anomalous results would likely be due to discrepancies between

the computational and conceptual models.

In this case, conducting an individual walkthrough of the original DSG code was

necessary because I had not written it, yet I needed to adapt it for my purposes. The

walkthrough revealed numerous coding errors that probably would have been found

through subsequent testing and debugging or eventually through APTMV analysis, but

those methods are generally more laborious and less efficient.

Execution tracing, desk checking, and syntax checking generally were used in

tandem while writing both the replay and roboplay code. I periodically used desk

checking while writing my code to ensure that modifications I made—especially to the

strategy-selection algorithm—were consistent across functions. I wrote my code in

Adobe's Dreamweaver software, which provides some simple, automatic syntax

checking for PHP and helped me to avoid small yet costly errors like misspelling variable

names and omitting closing braces for functions. Execution tracing enabled me to follow

the conditional paths of the code and the variable declarations to ensure, for example, that

the score for Strategy 1 was being added to other strategy scores only during the first

fifteen turns of a game. Again, final APTMV analysis might have uncovered these errors,

but the process would have been less efficient. Overall, these methods increased the likelihood that the implementation of the optimal strategies as a computational model was accurate and would provide reliable data for evaluating the DSG's computational model.

APTMV analysis of data from replayed games confirmed that significant programming errors found through code review were manifested in game results that were inconsistent with predictions based on using optimal strategies. In the replayed games, for example, the errors in the score calculations for the *Print* activity made it more successful than predicted, and APTMV identified this discrepancy when comparing replayed games to roboplayed games. Other unpredicted results were traced back to errors in the code that weren't found during code review, demonstrating that APTMV may be used to support code review through analysis of gameplay results.

APTMV proved especially useful in identifying potential discrepancies between the computational and conceptual models and in quantifying the results of modifications to the code in order to improve the computational model. I was able to fine-tune the DSG code for the *Demonstration* and *Pilot Test* activities so that they resulted in mean points per turn that were reasonable with respect to other activities yet improved overall game results to meet the threshold of confidence for model verification. Statistical analyses in the form of *t* tests supported the assertion that these changes made significant differences in the mean points per turn.

Research Question 3: What improvements can be made to the proposed method?

By documenting problems encountered, solutions attempted, and ideas for improving APTMV throughout data collection and analysis, I identified potential

improvements to the method. To answer this research question, I will discuss each step of the APTMV procedure.

Step 1: Formulate the conceptual model as patterns of temporal events.

In this study, the conceptual model was the diffusion of innovations theory as described by Rogers (2003). Rogers provides a number of generalizations about what actions should be effective with various adopter types at different stages of the innovation decision process (see Appendix B). This made it fairly easy to identify important data classifications and their categories. However, this task requires a solid understanding of APT, and reading Frick's explanations (e.g., Frick, 1990) may not be sufficient preparation. Because this step is crucial in using APT—for model verification or any other purpose—I suggest developing instructional resources to prepare designers and researchers to use APT.

Step 2: Map those events to actions that may be taken in the simulation.

Game mechanics are methods for taking actions in a game. See Reigeluth and Myers (2012) for a discussion of game mechanics and their relationship to instructional design for serious games. The DSG's core game mechanic is selecting diffusion and information activities. Some of these activities also employ a mechanic that requires the player to choose one or up to five staff members to participate in the activity. The success of an activity should depend largely on its alignment with each staff member's adopter type and current phase in the innovation-decision process. The fundamental purpose of APTMV is to determine whether the results of using the game mechanics are consistent with predictions of real-world results based on the diffusion of innovations theory.

However, part of the process of mapping events to actions should include identifying aspects of the conceptual model that are not included in the computational model, and that was not done in this study. For example, Rogers discusses "reinvention" in which potential adopters adapt the innovation to suit their particular situations and needs. If for some reason people are not allowed or able to reinvent the innovation, they are less likely to adopt it. This aspect of diffusion of innovations theory was added after the DSG was developed, so it is not part of the computational model. As another example, Rogers discusses the use of incentives to entice later adopters to adopt earlier and accelerate the rate of adoption, but incentives are not part of the DSG. Whether these and other aspects of the diffusion of innovations theory that are missing from the DSG should be included are design decisions influenced by the learning goals and the desired fidelity and complexity of the DSG. Nevertheless, differences between the conceptual and computational models should be noted and discussed during the APTMV process.

Step 3: Validate mappings of events to simulation actions.

At this point we try to eliminate as much bias as possible from the verification process by asking experts to review the assumptions made about the relationship between the conceptual and computational models. In this study, I requested assistance from six people who either had published papers related to diffusion of innovations theory or had taught related classes. Of the two who agreed to participate, one had never played the DSG while the other used it regularly in his classes. The first reviewer did not directly address the assumptions I put forth, commenting that it was difficult to understand them without experiencing the simulation. He said that what I proposed seemed reasonable, but he continued by stressing the importance of context for the success of diffusion activities.

119

The second reviewer directly addressed each assumption and in some cases suggested design alternatives based on his understanding of the theory. Both of these reviewers provided useful feedback, but it seems that validation of mappings might be improved by asking reviewers to play the simulation to understand its context.

Step 4: Identify the data associated with those actions that are required for analysis.

Because the DSG consists of discrete turns with just a few game mechanics, identifying the required data seemed like it would be fairly simple. However, it ended up being an iterative process that involved multiple attempts until I finally had all of the variables necessary to answer my research questions. Each of the final datasets (one for the replayed games and one for the roboplayed games) consisted of 150 variables, including 5 variables for each of the 24 staff members.

The approach I took with roboplayed games (which were analyzed before replayed games) was to run a small number of games and then see if I was able to calculate the values needed to answer my research questions. In some cases I was able to calculate new variables in Excel rather than rewrite my code to calculate and store those values during automated gameplay. For example, the variables S1, S2, S3, S4, and S5 were used to store the codes for the staff members who were selected to participate in targeted activities. However, I found that for data analysis it was better to have a Boolean variable for each staff member indicating whether he or she was selected in a given turn. It was easier to write a formula in Excel to calculate this rather than to modify the DSG code and database and rerun the games.

Step 5: Specify the threshold of confidence.

This was perhaps the most difficult step in the procedure. Determining the required probabilities for the model to be considered accurate depends on a number of variables, but the primary consideration should be the risks associated with using a flawed model. Simulations used to train pilots or doctors, for example, should have extremely high verification thresholds due to the real-world dangers that may result from poor learning experiences. Situational simulations that attempt to model human behavior are generally more stochastic, so deciding on an appropriate verification threshold requires balancing the unpredictability of human nature with the desire to convey the attributes of the underlying theory that is to be learned. More specific guidelines for specifying the threshold of confidence for model verification may emerge over time as APTMV is used with a variety of types of simulations in different contexts.

Step 6: Programmatically collect and store the data.

Data requirements are based on classifications and categories as described above. I designed a database specifically for this project based on the data needed to verify the model, and I modified the DSG to capture those data and insert them into the database. However, ideally the APTMV approach to model verification should be part of the planning and design process of a simulation so that the collection and storage of the necessary data is included in the original code. Modifying the code after the simulation is finished increases the risk of introducing errors that may affect the operation of the simulation.

Step 7: Format the data for insertion into the specialized MAPSAT database.

I included this and the next step in the procedure anticipating the day when MAPSAT software will be available. We currently have a MAPSAT database that can store both temporal and structural data for analysis, and we built a prototype tool that can calculate the structural properties of systems as specified by MAPSAT. Once we have a tool for querying temporal patterns and have tested and modified the database as necessary, we will document the format and procedure for loading data into the database. Until that time, model verification must take an approach similar to the one used in this study.

Step 8: Specify queries of the data for patterns of interest using MAPSAT software.

Lacking MAPSAT software, I instead analyzed roboplayed games by assigning scores for each strategy to every turn, selecting the optimal strategy, and looking at the results to determine whether these strategies—which were predicted to lead to success based on DOI theory—actually resulted in high scores. For replayed games, I looked for instances in which successful game results were achieved when optimal strategies were not used. However, I was able to take this approach only because a large amount of historical gameplay data was available. An alternative approach not taken in this study would be to program roboplaying so that sub-optimal strategies or strategies predicted not to be successful were used.

Step 9: Analyze the results of queries to ascertain the probability that the computational model accurately represents the conceptual model.

MAPSAT software would have calculated the probability of achieving a successful game given the appropriate use of each strategy. An optimal strategy with a lower than expected probability would have indicated a flaw in the computational model. In this study, I instead used statistical methods to find strategies that were not performing as expected as well as to find unpredicted strategies that were successful. When MAPSAT software is available, it should be useful to compare both approaches to understand the strengths and weaknesses of each.

CHAPTER 6: CONCLUSIONS

Summary

When a simulation or simulation game is designed for educational purposes, the designers must ensure that they achieve the appropriate degree of fidelity with respect to the conceptual model on which it is based. Otherwise the learning experience may lead to misconceptions based on incorrect feedback. Model verification is a procedure for ensuring that the conceptual model of a real-world phenomenon has been translated into a computational model with sufficient accuracy. The purpose of this study was to test whether the proposed method for the verification of a simulation's computational model, referred to as analysis of patterns in time for model verification (APTMV), is effective in providing quantified evidence of and in improving a model's performance.

The APTMV method of simulation model verification is based on a concept that is easy to understand: pattern matching. Simulation modeling begins by selecting or creating a conceptual model of a system, process, entity, or other phenomenon of interest. If this conceptual model can be used to predict outcomes based on joint and/or sequential patterns of the model's components, it is likely that APTMV can be used to verify the related computational model by calculating the occurrence of those predicted outcomes in the simulation given the same patterns. Outcomes for highly deterministic models are reliably predictable and should require a high threshold of confidence for verification, while more stochastic models should have a lower threshold to accommodate the variability of outcomes.

Balci (1997) identified four types of model verification methods: informal techniques, static techniques, dynamic techniques, and formal analysis. APTMV is a dynamic technique because it requires model execution in order to examine the results of

the model's performance. In simulation modeling it is generally acknowledged that no single verification method can provide sufficient evidence of a model's accuracy. Instead, multiple methods are employed and are selected based on a variety of considerations, including the purpose and type of simulation and the knowledge and skills of the designer(s) doing the model verification. In this study, the methods used in addition to APTMV were walkthrough, execution tracing, desk checking, syntax checking, and statistical analysis. Experts on the conceptual model provided a degree of face validity by reviewing the proposed strategies for testing the computational model.

The case used in this study to test the APTMV method was the *Diffusion Simulation Game* (DSG), which has as its primary conceptual model the diffusion of innovations theory (DOI). In the DSG, the player takes on the role of a change agent in a junior high school who must persuade as many of the staff members as possible to adopt peer tutoring before the end of two academic years. The player may select information activities to obtain descriptions of the staff members and observe their social connections. The player may also select diffusion activities, some of which require the selection of one or more staff members, to spread knowledge about peer tutoring and move individuals toward its adoption.

Rogers (2003) makes a number of generalizations based on DOI theory about what should be effective in the real world to promote the adoption of an innovation within a social system. Those generalizations can be expressed as temporal patterns of events and attributes of system members. For example, Rogers says that mass media communication channels can be especially effective in raising the awareness and interest of earlier adopters. For this study, I expressed generalizations from DOI theory as

strategies consisting of joint and sequential patterns of components in the DSG. Those components included activities (e.g., using mass media); the innovation-decision phase for each member of the system (which changes over time as members become aware of the innovation and move toward adoption); and attributes of system members, such as adopter type, opinion leadership, and interpersonal connections. Each turn in the DSG requires the player to decide on a course of action, and these choices in total comprise an *interaction trail* that may be compared with optimal strategies based on DOI theory.

Two approaches were used to examine interaction trails. The first approach entailed using a computer program to automatically play the DSG, and I called these "roboplayed" games. At the heart of this program was a strategy-selection algorithm that examined the game state during each turn, calculated a score for each optimal strategy, and implemented the highest scoring strategy. In this way, every turn of every game employed the best possible strategy based on DOI theory. The hypothesis was that this approach should result in a majority of games (more than 60%) in the top quartile in terms of possible points obtained, an outcome I called the *threshold of confidence* for model verification.

The computational model of the DSG initially failed to achieve the desired threshold of confidence for verification. Examination of APTMV data revealed aspects of the model that were not performing as expected. For example, the *Pilot Test* activity, in which one staff member is selected for a hands-on trial of peer tutoring, had a significantly lower mean points per turn than any other activity. The strategy specified that the staff member must already be in the Interest phase, and usually an opinion leader with a large social network was selected. Investigation of the game algorithm revealed

that one staff member who was frequently selected for this activity—an earlier adopter and opinion leader—was grouped with later adopters, resulting in almost no points for those turns. I made some small changes to the feedback groupings so that, in terms of adopter type, similar staff members were grouped together. This resulted in improved performance for this aspect of the computational model. I made several other modifications to the model based on examination of similar discrepancies in the data and tested the results. These small changes were enough to affect game outcomes so that the threshold of confidence was met.

The second approach entailed examining games that had been played by real players between October 2006 and April 2009 (n=2,361). Where the first approach tested whether strategies based on DOI theory were successful in the DSG, this approach sought winning strategies that could not be explained by DOI theory. Therefore I focused on only games in the top quartile of possible points. In these games, I looked for strategies that had both a relatively high frequency of use and high mean points per turn. Because the data collected when these games were originally played were not sufficient to conduct APTMV analysis, I wrote a program to replay these games using the extant data. I called these "replayed" games to distinguish them from roboplayed games.

Several successful strategies that I discovered in the replayed games were the result of programming errors that awarded more points than intended based on the original board game algorithm. In a previous study (Enfield et al., 2012), we found that the *Print* activity was used more frequently in successful games. During my individual walkthrough of the DSG code, I discovered that every result for the *Print* activity

awarded more points than intended by the original designers. Savvy gamers had clearly noticed the effectiveness of this activity and exploited it.

The *Materials Workshop* activity was used less frequently in replayed games than in roboplayed games, but the mean points per turn was higher. This suggested that real players had devised a better strategy for using this activity than I had using DOI theory. In examining the data, I discovered that real players were using this activity later in the game, when more staff members were in the Interest and Trial phases and therefore more open to this hands-on approach. The roboplayed games had targeted the earlier adopters so that they reached the Interest and Trial phases much sooner than the later adopters, and there were enough of them to make this the highest scoring strategy. However, the strategy would have been more successful if the strategy-selection algorithm had given preference to moving some of the later adopters into the Interest and Trial phases before using the *Materials Workshop* activity. Rogers recommends trying to achieve a critical mass of adopters to increase momentum toward adoption, but perhaps a corollary should be not to let the more resistant members of the system fall too far behind.

Discussion of Research Questions

The first research question is whether APTMV is an effective method for verifying and improving a simulation's computational model. To answer this question, I wrote a program to play the DSG using a strategy-selection algorithm that analyzed the game state (i.e., current patterns of components) during each turn and selected the optimal strategy. Given the stochastic nature of the DSG and its educational goals (which include conveying the difficulty of being a change agent in the real world), I estimated that in order to verify that the computational model was sufficiently accurate, more than

60% of games that used these optimal strategies would need to accumulate enough points to fall in the top quartile of games based on total possible points.

I then ran the program to play a specified number of games and used basic data analysis techniques—such as calculating totals, frequencies, and means—to determine which particular strategies were not resulting in the predicted outcomes. This approach led me to find discrepancies between the DSG's computational model and my strategy-selection algorithm. Some of these discrepancies were the result of coding errors in the computational model. Other discrepancies occurred due to subtle differences in the implementation of a generalization from DOI theory. For example, Rogers says that allowing potential adopters to see the innovation being used, especially by an opinion leader, can increase their interest in the innovation. My strategy-selection algorithm required that a staff member have completed the Awareness phase in order to be ready for the *Demonstration* activity conducted by an opinion leader; the DSG additionally required at least one point in the Interest phase. Designers could (and perhaps did) argue the merits of both choices based on their knowledge, experience, and intuition. However, I was able to modify the code to test both implementations and see how the changes impacted turn results and game outcomes. Not requiring a point in the Interest phase resulted in more games in the top quartile, but the mean points per turn for the *Demonstration* activity was much higher than for other activities, an undesired imbalance that might result in players ignoring other activities that should be equally effective. This example illustrates the effectiveness of APTMV in quantifying the outcomes of design decisions with regard to the fidelity of the computational model, and in using that information to make more informed design decisions.

Analysis of summary data also indicated that the *Pilot Test* activity had a much lower mean score per turn than any other activity. Further investigation revealed inconsistencies in the groupings of staff for activity results, with a particularly desirable target—a highly connected opinion leader—being grouped with Late Majority and Laggards. The DOI strategy of targeting individuals with large social networks to spread information about the innovation resulted in this staff member being selected frequently for the *Pilot Test* activity, yet few points were scored, indicating a potential flaw in the computational model. After making a few changes to these groupings, based largely on adopter type, the *Pilot Test* activity scores were more consistent with expected results. Again, APTMV proved useful in identifying a flaw in the computational model and in providing quantified evidence that changes to the model had the desired results.

Similar results were found for other strategies, and it would be possible to continue experimenting with the DSG's parameters to fine-tune results and see how game outcomes were affected. The few modifications to the DSG that were explored here were sufficient to provide evidence that pattern analysis is an effective method for verifying a computational model.

The second research question concerns the usefulness of APTMV with respect to other verification methods. In terms of the Whitner and Balci taxonomy described above, APTMV is a dynamic, black box technique because it requires execution of the model and analysis of the output to make inferences about the model. Using multiple verification methods is recommended practice, but methods should be chosen to complement each other, balancing the strengths and weaknesses of each to increase confidence in the results. Several of the methods used in this study, including

walkthrough, desk checking, syntax checking, and execution tracing, are intended to ensure that the code for the computational model has as few errors as possible and conforms to standard coding practices, thereby reducing the possibility of unintended outcomes. However, a program may run without crashing and produce seemingly reasonable results yet still not be sufficiently accurate with respect to the conceptual model upon which it is based.

APTMV provides designers with a methodological approach for testing a computational model, obtaining quantified evidence of its performance, and evaluating the degree to which changes to the model affect outcomes. In this study, generalizations from DOI theory (the conceptual model) were mapped to strategies in the DSG that were expressed as patterns of actions and attributes. These patterns were enacted through automated gameplay using a strategy-selection algorithm that chose the optimal strategy for each turn given the current game state. I hypothesized that this approach should result in at least 60% of games falling into the fourth quartile in terms of adoption points achieved (the threshold of confidence) in order to verify that the computational model is reasonably accurate given the learning goals. The first APTMV results indicated that the computational model failed to meet the threshold of confidence. At that point I used statistical methods to analyze the APTMV data and seek the model's flaws.

According to the literature, statistical methods are often used to make sense of the large amount of data generated through the use of dynamic verification techniques such as APTMV. In this study, APTMV results were analyzed using descriptive statistics and t tests to identify and investigate problematic aspects of the computational model and to determine whether changes to the model resulted in significant differences and desired

outcomes. Combining APTMV with statistical methods proved useful for going beyond a simple yes/no answer to the verification question and instead seeking a satisficing solution.

The third research question asks what potential improvements can be made to APTMV. First and foremost, MAPSAT software for pattern analysis should greatly improve the efficiency of model verification using APTMV. Exporting data from the MySQL database and importing it into Excel for analysis was not difficult, but doing this repeatedly took time that would have been better spent on higher-level tasks. I spent many hours writing and debugging formulas in Excel to detect the desired patterns, but MAPSAT software should simplify and streamline the use of APT in general. Writing gameplay data directly to a MAPSAT database, from which patterns of interest could then be queried and probabilities calculated, would save designers valuable time. It would also accelerate the process of model verification for simulation designers who want to use APTMV for formative evaluation of computational models.

More detailed guidance is needed for simulation designers who want to use the APTMV method in their work. The current procedure includes brief descriptions of the steps with guiding questions and explanations. However, unless users are already familiar with APT, I suspect they will need more detailed assistance in, for example, translating a conceptual model into relevant classifications and categories. The current study can serve as an example, but more examples would be useful in demonstrating how different contexts may influence data requirements and choice of complementary model verification methods. One approach would be to build a simulation of the MAPSAT

software with an instructional overlay (Reigeluth & Schwartz, 1987) that would guide the user through the APTMV process.

The process of expert review and validation also needs to be improved. For the current study, experts were asked to validate the strategies I formulated based on DOI theory and mapped to events in the DSG. One expert had not played the DSG and so had to rely on descriptions of the DSG's context and activities. While he thought that the strategies seemed reasonable, he emphasized that the contextual phenomena surrounding the diffusion situation would influence the effectiveness of any particular strategy. The other expert I consulted regularly used the DSG in classes he taught on change management and the diffusion of innovations. He addressed the strategies more directly based on his knowledge of the DSG and also suggested design options such as including incentives and penalties. It seems to me that an entire study could focus on how to select experts for simulation review and what information to provide them. The next time I conduct model verification for a simulation, I may seek experts on the conceptual model who have not played the simulation and ask them to play it once or twice before validating strategies.

I struggled in defining the threshold of confidence for model verification, and I think it would be fruitful to develop and test other definitions. The metric for verification may vary depending on the type or context of the simulation, the goals of the simulation designers, the other methods of verification being used, or other situational variables. After APTMV has been used several times, heuristics for determining the threshold of confidence may begin to emerge.

Implications for Practice

When learners engage with an interactive system like a game or simulation, they attempt to make sense of the underlying model through a cycle of trial, feedback, and judgment, actively inferring rules and properties based on the results of their actions. If that model does not respond as intended by the designers, learners may draw the wrong conclusions from their experiences. Therefore it is critical that designers verify their models to ensure that they perform with sufficient accuracy.

As this study has shown, APTMV is a viable approach to model verification and improvement. In a best-case scenario, the APTMV process would begin early in a project lifecycle to ensure that automated gameplay is part of the design and that necessary data are being collected for pattern matching. APTMV could then be used during an iterative development process to provide formative evaluation of the computational model. However, if it is not possible or desirable to include automated gameplay (perhaps due to the additional time and expense, for example), APTMV could still be used as long as the appropriate data were collected for analysis. The drawback would be the time required to manually play enough games to generate sufficient data for analysis.

In this study, I conducted the DSG model verification, and I was not a member of the original design team. It seems to me necessary to have model verification led by someone who has no (or little) knowledge of the computational model design, to avoid bias toward accepting previous design decisions. However, that person should be a subject matter expert and/or should consult one or more experts who were not involved in the model design to ensure that the patterns used to test the model are valid.

As is common practice, APTMV should be used with other verification methods to increase the reliability of the verification decision. In this study, APTMV proved

compatible with static methods (desk checking, walkthrough, syntax checking), uncovering coding errors missed when those methods were used. The APTMV pattern scores and associated gameplay data were amenable to statistical analysis; in this study, descriptive statistics and t tests were used, but in a different context regression and other techniques may be applicable.

Limitations of the Study

A limitation of this study is that it examines only one type of simulation—a situational simulation designed for educational purposes (types of simulations are discussed in the review of the literature). It may be that the proposed method is applicable to other types of simulations, but exploration of that possibility is beyond the scope of the current study. Furthermore, this study examines a single simulation game, and other situational simulations may have different contexts, mechanics, and underlying models that make them unsuitable for APTMV.

The DSG gameplay strategies that I formulated were based on my understanding of DOI theory, based primarily on a close reading of Rogers (2003). Even though the strategies were distributed for expert review, only two experts responded to requests for assistance. It is possible that I misinterpreted Rogers or incorrectly formulated a gameplay strategy. In terms of model verification, a flawed strategy may result in an erroneous decision.

While I took many precautions to ensure that my code worked as desired, it is possible that the findings could be corrupted by, for example, a faulty calculation of strategy scores. As was seen during the verification process, even small changes to the code can result in significantly different game outcomes.

In using APTMV to conduct model verification, I necessarily drew upon my particular knowledge and skills. I have over a decade of experience in Web programming, including several recent years of intermittent PHP development. Furthermore, I was employed for many years as an analyst in institutional research at a university, extracting large volumes of data from information systems and poring through them to find and correct errors. Over the years, I have developed some expertise in spotting data anomalies and tracking down their causes. In at least some cases, the ability to replicate the APTMV approach will require one or more people with similar knowledge and skills.

Suggestions for Future Research

In discussing the results of this study in relation to the third research question (regarding improvements to the method), I noted several aspects of the APTMV method that could be improved through further application and research. Because this was a single-case study, there are several potential lines of research that could improve the usefulness and reliability of APTMV.

Further research is needed with different situational simulations and with other types of simulations to determine the proposed method's range of applicability. Can APTMV be used to verify the computational model of a physical system? How will it work with agent-based models? Does the method need to change to accommodate different types of models or contexts of use? This knowledge will be especially useful to practitioners trying to decide whether they might use APTMV in their simulation modeling process.

Another area of research concerns defining the threshold of confidence for model verification. In this study, I took into account the stochastic nature of turn results and the

designers' expressed goal of conveying the difficulty of change agentry when I set the threshold of 60% of games in the fourth quartile. Was that threshold strict enough? A follow-up study using the same code and strategies might explore the effects of setting different values for the threshold of confidence. Similar studies with other simulations should seek to find moderating factors and to establish guidelines for setting the threshold of confidence.

It would also be useful to establish a protocol for soliciting experts' reviews of simulations with respect to model verification. Research might explore formats for presenting the model for review and for gathering feedback. Is it helpful for experts to play a working prototype, or does that invite feedback on the face validity of the simulation rather than on the patterns to be used in testing the model?

Using APTMV in combination with several other model verification methods revealed both significant and subtle problems with the online version of the DSG. Many of these problems were the result of programming errors; some of these errors were discovered through a close examination of the code, but others were not detected until the simulation was executed hundreds of times and APTMV data revealed anomalies. More importantly, APTMV enabled detection of design errors that negatively impacted the fidelity of the computational model with respect to DOI theory. Furthermore, APTMV provided quantified evidence that design modifications improved the fidelity of the computational model, reducing the chance that learners would develop misunderstandings about DOI theory as a result of playing the DSG.

Final Thoughts

Simulations and games have tremendous potential for providing authentic, challenging, engaging, and effective learning experiences for people of all ages and

abilities. They can offer immersive virtual environments in which learners can take on roles in situated contexts and gain a deeper understanding and appreciation of, for example, what it is like to be a surgeon in an operating room, actively applying their knowledge and skills to save a dying patient, learning from their successes and failures without actually harming anyone.

Computational power continues to increase at a rapid pace, enabling us to carry sophisticated computers in our pockets and access vast libraries of information from nearly anywhere. Ray Kurzweil (1999) has predicted that within a few decades our educational systems will be radically transformed by our technologies, with readily accessible virtual realities providing life-long learning opportunities as needed. Until, and even during, that time we will need to determine whether these learning experiences have been designed accurately to convey their underlying conceptual models with the appropriate degree of fidelity. The verification of computational models will be increasingly important in ensuring that simulations and games for learning accurately represent that which is to be learning and thereby promote the knowledge, skills, and attitudes necessary for successful performance in the real world.

The present study has contributed a new method that the designers of simulations and games for learning may use in their efforts to verify the accuracy of their computational models. Analysis of patterns in time for model verification (APTMV) shows promise as a new and effective method for providing quantified evidence to measure and improve the performance of a computational model. Further use and research of APTMV should lead to improvements in the method and a greater

appreciation of the power of pattern analysis in the design of simulations and games for

learning.

REFERENCES

Aldrich, C. (2005). *Learning by doing: The essential guide to simulations, computer games, and pedagogy in e-learning and other educational experiences*. San Francisco, CA: John Wiley & Sons.

Aldrich, C. (2009). *Learning online with games, simulations, and virtual worlds: Strategies for teaching and learning series*. San Francisco, CA: Jossey-Bass.

Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction, 15*(2), 40-47.

Alessi, S. M. (2000). Simulation design for training and assessment. In H. F. O'Neil & D. H. Andrews (Eds.), *Aircrew training and assessment* (pp. 197-222). Mahway, NJ: Lawrence Erlbaum Associates.

Alessi, S. M. & Trollip, S. R. (2001). *Multimedia for learning: Methods and development* (3rd ed.). Boston, MA: Allyn and Bacon.

An, J. (2003). *Understanding mode errors in modern human-computer interfaces: Toward the design of usable software*. Unpublished doctoral dissertation. Indiana University. Bloomington, IN.

Andrews, D. H., Carroll, L. A., & Bell, H. H. (1995). The future of selective fidelity in training devices. *Educational Technology, 35*(6), 32-36.

Axelrod, R. (2007). Advancing the art of simulation in the social sciences. In J. P. Rennard (Ed.), *Handbook of research on nature inspired computing for economics and management* (pp. 90-100). Hershey, PA: Idea Group Reference.

Balci, O. (1997). Verification, validation, and accreditation of simulation models. In S. Andradottir, K. J. Healy, D. H. Withers, & B. L. Nelson (Eds), *Proceedings of the*

1997 Winter Simulation Conference (pp. 135-141). Retrieved from

http://www.informs-sim.org/wsc97papers/0135.PDF

Balci, O. (1998). Verification, validation, and testing. In J. Banks (Ed.), *The handbook of*

simulation: Principles, methodology, advances, applications, and practice (pp.

335-394). Hoboken, NJ: John Wiley & Sons, Inc.

Barab, S. A., & Squire, K. (2004). Design-based research: Putting a stake in the ground.

The Journal of the Learning Sciences, 13(1), 1-14.

Barrett, A. F., Howard, C. D., & Frick, T. W. (2011). *Product quality and conditions of*

anonymity in formative peer assessment: Pre-service teachers' asynchronous

computer-mediated communication. Paper presented at the Annual Meeting of the

American Educational Research Association, New Orleans, LA.

Becker, K. (2007). Pedagogy in commercial video games. In D. Gibson, C. Aldrich & M.

Prensky (Eds.), *Games and simulations in online learning: Research and*

development frameworks (pp. 21-47). Hershey, PA: Information Science

Publishing.

Becker, K. (2008). *The invention of good games: Understanding learning design in*

commercial video games. Unpublished doctoral dissertation. University of

Calgary.

Bragge, J., & Storgards, J. (2007). *Profiling academic research on digital games using*

text mining tools. Paper presented at the Situated Play, Digra 2007 Conference.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in

creating complex interventions in classroom settings. *The Journal of the Learning*

Sciences, 2(2), 141-178.

Chatfield, T. (2009, September). Videogames now outperform Hollywood movies. *The Observer*. Retrieved from

http://www.guardian.co.uk/technology/gamesblog/2009/sep/27/videogames-

hollywood

Clark, R. C. (2005). Multimedia learning in e-courses. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 589-616). New York: Cambridge

University Press.

Coleman, J. S. (1970). *The role of modern technology in relation to simulation and games for learning* (Report No. ED 039 704). Washington, D. C.: Academy for

Educational Development, Inc. Retrieved from

http://www.eric.ed.gov/PDFS/ED039704.pdf

Collins, A. (1992). The changing infrastructure of education research. In E. Scanlon & T.

O'Shea (Eds.), *New directions in educational technology*. Berlin: Springer.

Corliss, R. (2010, January). Box office 2009: A very good year. *Time*. Retrieved from

http://www.time.com/time/arts/article/0,8599,1951371,00.html

Cruickshank, D. R. (1988). The uses of simulations in teacher preparation: Past, present, and future. *Simulation & Gaming, 19*(6), 133-156.

David, N. (2009). Validation and verification in social simulation: Patterns and

clarification of terminology. In F. Squazzoni (Ed.) *Epistemological aspects of computer simulation in the social sciences* (pp. 117-129). Heidelberg: Springer-

Verlag Berlin.

DBRC (Design-Based Research Collective). (2003). Design-based research: An emerging

paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and

simulations within the curriculum be most effectively evaluated? *Computers &

Education, 46*(3), 249-264.

de Jong, T., & van Joolingen, W. R. (1998). Scientific discovery learning with computer

simulations of conceptual domains. *Review of Educational Research, 68*(2), 179-

201.

Dempsey, J. V., Lucassen, B., Gilley, W., & Rasmussen, K. (1993-1994). Since Malone's

theory of intrinsically motivating instruction: What's the score in the gaming

literature? *Journal of Educational Technology Systems, 22*(2), 173-183.

Dempsey, J. V., Rasmussen, K., & Lucassen, B. (1996). *Instructional gaming:

Implications for instructional technology*. Paper presented at the Annual Meeting

of the Association for Educational Communications, Nashville, TN.

Dewey, J. (1997). *Experience and education*. New York: Touchstone. (Original work

published 1938).

Diffusion Simulation Game (Online Version 1, 2002). Bloomington, IN: Indiana

University (201 North Rose Avenue, Bloomington, IN 47405-1006). Available at

https://www.indiana.edu/~simed/istdemo/

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *The

Journal of the Learning Sciences, 11*(1), 105-121.

Edelson, D. C. (2006). Balancing innovation and risk: Assessing design research

proposals. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen

(Eds.), *Educational design research* (pp. 100-106). New York: Routledge.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: Exploring the educational potential of computer games*. Unpublished doctoral dissertation. IT-University Copenhagen. Copenhagen, Denmark.

Enfield, J., Myers, R. D., Lara, M., & Frick, T. W. (2012). Innovation diffusion: Assessment of strategies within the DIFFUSION SIMULATION GAME. *Simulation & Gaming, 43*(2), 188-214. doi: 10.1177/1046878111408024

Faria, A. J. (1998). Business simulation games: Current usage levels--An update. *Simulation & Gaming, 29*(3), 295-308.

FAS (Federation of American Scientists). (2006). *R&D challenges in games for learning*. Retrieved from http://www.fas.org/gamesummit/Resources/R&D_Challenges.pdf

Feinstein, A. H., & Cannon, H. M. (2002). Constructs of simulation evaluation. *Simulation & Gaming, 33*(4), 425-440.

Foreman, J. (2004). Game-based learning: How to delight and instruct in the 21st century. *EDUCAUSE Review, 39*(5), 50-66.

Frick, T. (1983). *Nonmetric Temporal Path Analysis (NTPA): An Alternative to the Linear Models Approach for Verification of Stochastic Educational Relations*. Unpublished doctoral dissertation, Indiana University. Retrieved from http://www.indiana.edu/~tedfrick/ntpa/

Frick, T. (1990). Analysis of Patterns in Time (APT): A Method of Recording and Quantifying Temporal Relations in Education. *American Educational Research Journal, 27*(1), 180-204.

Frick, T., Barrett, A., Enfield, J., Howard, C., & Myers, R. (2009). *Alternative research methods: MAPSAT your data to prevent aggregation aggravation*. Paper

presented at the annual conference of the Association for Educational

Communications & Technology, Louisville, KY. Retrieved from

https://www.indiana.edu/~tedfrick/aect2009/MAPSAT2009AECT_Frick_Barrett_

Howard_Enfield_Myers.pdf

Frick, T., Chadha, R., Watson, C., & Zlatkovska, E. (2010). Improving course

evaluations to improve instruction and complex learning in higher education.

Educational Technology Research and Development, 58(2), 115-136.

Frick, T., Kim, K.-J., Ludwig, B. and Huang, R. (2003). *A Web simulation on educational*

change: Challenges and solutions for development. Paper presented at the

meeting of the Association for Educational Communication and Technology,

Anaheim, CA. Retrieved from

http://www.indiana.edu/~tedfrick/aect2003/frick_kim_ludwig_huang.pdf

Frick, T., Myers, R., Thompson, K. & York, S. (2008). *New ways to measure systemic*

change: Map & Analyze Patterns & Structures Across Time (MAPSAT). Featured

research paper presented at the annual conference of the Association for

Educational Communications & Technology, Orlando, FL.

Frick, T. & Thompson, K. (2008). Predicting education system outcomes: A scientific

approach. In M. Orey, V. McClendon, & R. Branch (Eds.), *Educational Media*

and Technology Yearbook (Vol. 33), 62-76.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A

research and practice model. *Simulation & Gaming, 33*(4), 441-467.

Garson, G. D. (2009). Computerized simulation in the social sciences. *Simulation &*

Gaming, 40(2), pp. 267-279.

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

Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-learning, 2*(1), 5-16.

Gibbons, A. S., McConkie, M., Seo, K. K., & Wiley, D. (2009). Simulation approach to instruction. In C. M. Reigeluth & A. A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base* (Vol. III, pp. 167-193). New York: Routledge.

Gilbert, N. (2004). Computer simulation. In A. Kuper & J. Kuper (Eds.), *The social science encyclopedia* (3rd ed., pp. 145-147). New York: Routledge.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (3rd ed.). Boston: Pearson.

Gordon, J., & Zemke, R. (2000). The attack on ISD. *Training, 37*(4), 42-52.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Gredler, M. E. (2004). Games and simulations and their relationships to learning. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 571-581). Mahwah, NJ: Lawrence Erlbaum & Associates.

Hays, R. T. (2006). *The science of learning: A systems theory approach*. Boca Raton, FL: BrownWalker Press.

Heinich, R., Molenda, M., & Russell, J. D. (1993). *Instructional media and the new technologies of instruction* (4th ed.). New York: Macmillan.

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist, 39*(4), 203-212.

Howard, C. D., Barrett, A. F., & Frick, T. W. (2010). Anonymity to promote peer feedback: Pre-service teachers' comments in asynchronous computer-mediated communication. *The Journal of Educational Computing Research, 43*(1), 89-112.

Huang, W. D., & Johnson, T. (2009). Instructional game design using cognitive load theory. In R. E. Ferdig (Ed.), *Handbook of research on effective gaming in education* (Vol. 3, pp. 1143-1165). Hershey, PA: Information Science Reference.

Hypermedia (n.d.) In *Wiktionary: A wiki-based open content dictionary*. Retrieved from http://en.wiktionary.org/w/index.php?title=hypermedia&oldid=12829349

Jenkins, H. (1992). *Textual poachers:Television fans and participatory culture*. New York: Routledge.

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York: NYU Press.

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report*. Austin, TX: The New Media Consortium. Retrieved from http://net.educause.edu/ir/library/pdf/HR2011.pdf

Jones, S. (2003). *Let the games begin: Gaming technology and entertainment among college students*. Washington, D.C.: Pew Internet & American Life Project. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2003/PIP_College_Gaming_ Reporta.pdf.pdf

Judd, T. S., & Kennedy, G. E. (2004). More sense from audit trails: Exploratory

    sequential data analysis. In R. Atkinson, C. McBeath, D. Jonas-Dwyer & R.

    Phillips (Eds), *Beyond the comfort zone: Proceedings of the 21st ASCILITE*

    Conference (pp. 476-484). Perth, 5-8 December. Retrieved from

    http://www.ascilite.org.au/conferences/perth04/procs/judd.html

Kalyuga, S., & Plass, J. L. (2009). Evaluating and managing cognitive load in games. In

    R. E. Ferdig (Ed.), *Handbook of research on effective gaming in education* (Vol.

    3, pp. 719-737). Hershey, PA: Information Science Reference.

Kelly, A. (2006). Quality criteria for design research: Evidence and commitments. In J.

    van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational*

    design research (pp. 107-118). New York: Routledge.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during

    instruction does not work: An analysis of the failure of constructivist, discovery,

    problem-based, experiential, and inquiry-based teaching. *Educational*

    Psychologist, 41(2), 75-86.

Kirriemuir, J., & McFarlane, A. (2004). *Literature review in games and learning: A*

    report for NESTA Futurelab. Retrieved from

    http://hal.archives-ouvertes.fr/docs/00/19/04/53/PDF/kirriemuir-j-2004-r8.pdf

Klabbers, J. H. G. (2009). *The magic circle: Principles of gaming & simulation* (3rd ed.).

    Rotterdam, The Netherlands: Sense Publishers.

Koh, J. (2008). *The use of scaffolding in introductory technology skills introduction for*

    preservice teachers. Unpublished doctoral dissertation. Indiana University.

    Bloomington, IN.

Korzybski, A. (1995). *Science and sanity: An introduction to non-Aristotelian systems and general semantics* (5th ed.). Lakeville, CN: International Non-Aristotelian Library Publishing Company.

Krahl, D. (2005). *Debugging simulation models*. Paper presented at the Winter Simulation Conference, Orlando, FL.

Kuppers, G., & Lenhard, J. (2005) Validation of simulation: Patterns in the social and natural sciences. *Journal of Artificial Societies and Social Simulation, 8*(4). Retrieved from http://jasss.soc.surrey.ac.uk/8/4/3.html

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* New York: Viking.

Landa, L. N. (1999). Landamatics instructional design theory and methodology for teaching general methods of thinking. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 341-369). Mahwah, NJ: Lawrence Erlbaum Associates.

Leemkuil, H., de Jong, T., & Ootes, S. (2000). *Review of educational use of games and simulations.* Twente: University of Twente. Retrieved from http://kits.edte.utwente.nl/documents/D1.pdf

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: III*. Conative and affective process analysis (pp. 255-286). Hillsdale, NJ: Erlbaum.

Lerdorf, R., Tatroe, K., & MacIntyre, P. (2006). *Programming PHP* (2nd ed.). Sebastopol, CA: O'Reilly Media Inc.

Lieberman, D. A. (2006). What can we learn from playing interactive games? In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 379-397). Mahwah, NJ: Lawrence Erlbaum Associates.

Liu, D., Blickensderfer, E. L., Macchiarella, N. D., & Vincenzi, D. A. (2009). Simulation fidelity. In D. A. Vincenzi, J. A. Wise, M. Mouloua & P. A. Hancock (Eds.), *Human factors in simulation and training* (pp. 61-73). Boca Raton, FL: CRC Press.

Loh, C. S. (2007). Designing online games assessment as "information trails". In D. Gibson, C. Aldrich & M. Prensky (Eds.), *Games and simulations in online learning: Research and development frameworks* (pp. 323-348). Hershey, PA: Information Science Publishing.

Maier, F. H., & Grossler, A. (2000). What are we talking about--A taxonomy of computer simulations to support learning. *System Dynamics Review, 6*(2), 135-148.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333-369.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: III*. Conative and affective process analysis (pp. 223-253). Hillsdale, NJ: Erlbaum.

McEneaney, J. E. (2001). Graphic and numerical methods to assess navigation in hypertext. *Int. J. Human-Computer Studies, 55*, 761-786.

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). *Report on the educational use of games*. Cambridge, UK: Teem. Retrieved from http://www.teem.org.uk/publications/teem_gamesined_full.pdf

Misanchuk, E. R., & Schwier, R. A. (1991). Interactive media audit trails: Approaches and issues. *Proceeding of Selected Research Presentations at the Annual Convention of the Association for Educational Communications and Technology*.

Misanchuk, E. R., & Schwier, R. A. (1992). Representing interactive multimedia and hypermedia audit trails. *Journal of Educational Multimedia and Hypermedia, 1*(3), 355-372.

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). *A brief introduction to evidence-centered design* (CSE Technical Report 632). Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST). Retrieved from http://www.cse.ucla.edu/products/reports/r632.pdf

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6-20.

Mitchell, A., & Savill-Smith, C. (2004). *The use of computer and video games for learning: A review of the literature*. London: Learning Skills and Development Agency, Ultralab. Retrieved from http://www.m-learning.org/archive/docs/The%20use%20of%20computer%20and%20video%20games%20for%20learning.pdf

Molenda, M. (1976). *The Diffusion Simulation Game: Monitor's Instructions*.

Molenda, M. (2008). Historical foundations. In J. M. Spector, M. D. Merrill, J. van Merrienboer & M. P. Driscoll (Eds.), *Handbook of research on educational*

communications and technology (pp. 3-20). New York: Lawrence Erlbaum Associates.

Molenda, M., & Rice, J. M. (1979). Simulation review: The Diffusion Simulation Game. *Simulation & Games, 10*(4), 459-467.

NPD Group. (2010, January). 2009 U.S. video game industry and PC game software retail sales reach $20.2 billion. Retrieved from http://www.npd.com/press/releases/press_100114.html

Oblinger, D. G. (2006). Simulations, games, and learning (Report No. ELI3004). Educause Learning Initiative: EDUCAUSE. Retrieved from http://net.educause.edu/ir/library/pdf/ELI3004.pdf

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, 263*, 641-646.

Pace, D. K. (2004). Modeling and simulation verification and validation challenges. *Johns Hopkins APL Technical Digest, 25*(2), 163-172.

Parlett, D. (1999). *The Oxford history of board games*. New York: Oxford University Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Peters, V., & Vissers, G. (2004). A simple classification model for debriefing simulation games. *Simulation & Gaming, 35*(1), 70-84.

Peters, V., Vissers, G., & Heijne, G. (1998). The validity of games. *Simulation & Gaming, 29*(1), 20-30.

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

Punch, B., Patton, A., Wight, K., Larson, B., Masscheleyn, P., & Forney, L. (1996). A biodegradability evaluation and simulation system (BESS) based on knowledge of biodegradation pathways. In W. J. G. M. Peijnenburg & J. Damborsky (Eds.), *Biodegradability Prediction* (pp. 65-73). The Netherlands: Kluwer Academic Publishers.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: A review of recent research. *Simulation & Gaming, 23*(3), 261-276.

Reeves, T. C. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational design research* (pp. 52-66). New York: Routledge.

Reigeluth, C. M. (1994). The imperative for systemic change. In C. M. Reigeluth & R. J. Garfinkle (Eds.), *Systemic change in education* (pp. 3-11). Englewood Cliffs, NJ: Educational Technology Publications.

Reigeluth, C. M., & Frick, T. W. (1999). Formative research: A methodology for creating and improving design theories. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 633-651). Hillsdale, NJ: Lawrence Erlbaum Associates.

Reigeluth, C. M., & Myers, R. D. (2012). *Serious game design report*. Manuscript prepared for the U. S. Air Force Research Laboratory.

Reigeluth, C. M., & Schwartz, E. (1989). An instructional theory for the design of computer-based simulations. *Journal of Computer-Based Instruction, 16*(1), 1-10.

Richey, R. C., & Klein, J. D. (2007). *Design and development research*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Richey, R. C., Klein, J. D., & Nelson, W. A. (2004). Developmental research: Studies in instructional design and development. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (2nd ed., pp. 1099-1130). Mahwah, NJ: Lawrence Erlbaum & Associates.

Rieber, L. P. (2005). Multimedia learning in games, simulations, and microworlds. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 549-567). New York: Cambridge University Press.

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Science, 4*, 155-169.

Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Simon & Schuster.

Romiszowski, A. (1999). The development of physical skills: Instruction in the psychomotor domain. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. II, pp. 460-481). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ruben, B. D. (1999). Simulations, games, and experience-based learning: The quest for a new paradigm for teaching and learning. *Simulation & Gaming, 30*(4), 498-505.

Rutter, J., & Bryce, J. (2006). An introduction to understanding digital games. In J. Rutter & J. Bryce (Eds.), *Understanding digital games* (pp. 1-17). London: Sage Publications Ltd.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. Cambridge, Massachusetts: The MIT Press.

Sargent, R. G. (2008). *Verification and validation of simulation models*. Paper presented at the Winter Simulation Conference, Miami, FL.

Schwier, R. A., & Misanchuk, E. R. (1990). *Analytical tools and research issues for interactive media*. Paper presented at the Annual Meeting of the Association for Media and Technology in Education in Canada, St. John's, Newfoundland.

Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). London: SAGE Publications Ltd.

Shaffer, D. W. (2009). Wag the kennel: Games, frames, and the problem of assessment. In R. E. Ferdig (Ed.), *Handbook of research on effective gaming in education* (Vol. 2, pp. 577-592). Hershey, PA: Information Science Reference.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104-111.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2008). *Monitoring and fostering learning through games and embedded assessments* (Report No. RR-08-69). Princeton, NJ: Educational Testing Service.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295-321). New York: Routledge.

Sims, R. C., & Koszalka, T. A. (2008). Competencies for the New-Age instructional designer. In J. M. Spector, M. D. Merrill, J. van Merrienboer & M. P. Driscoll

(Eds.), *Handbook of research on educational communications and technology* (pp. 569-575). New York: Lawrence Erlbaum Associates.

Smith, J. E. (1996). Computer simulation. In A. Kuper& J. Kuper (Eds.), *The social science encyclopedia* (2nd ed., pp. 118-120). New York: Routledge.

Smith, R. (2010). The long history of gaming in military training. *Simulation & Gaming, 41*(1), 6-19.

Strommen, E., & Lincoln, B. (1992). Constructivism, technology, and the future of classroom learning. *Education and Urban Society, 24*(4)*,* 466-476.

Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). *Concepts of model verification and validation* (No. LA-14167-MS). Los Alamos, NM: Los Alamos National Laboratory.

Thompson, K. R. (2005). Axiomatic theories of intentional systems: Methodology of theory construction. *Scientific Inquiry Journal, 7*(1), 13-24. Retrieved from http://www.iigss.net/Scientific-Inquiry/THOMPSON-2.pdf

Thompson, K. R. (2008a). ATIS: Axiomatic Theories of Intentional Systems. Columbus, OH: System-Predictive Technologies. Retrieved from http://www.indiana.edu/~aptac/glossary/atisTheory.html

Thompson, K. R. (2008b). ATIS graph theory. Columbus, OH: System-Predictive Technologies. Retrieved from http://www.indiana.edu/~aptfrick/reports/11ATISgraphtheory.pdf

van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational design research*. New York: Routledge.

van den Akker, J. (1999). Principles and methods of development research. In J. van den
    Akker, N. Nieveen, R. M. Branch, K. L. Gustafson & T. Plomp (Eds.), *Design
    methodology and developmental research in education and training* (pp. 1-14).
    The Netherlands: Kluwer Academic Publishers.

Van Sickle, R. (1986). A quantitative review of research on instructional simulation
    gaming: A twenty-year perspective. *Theory and Research in Social Education,
    14*(3), 245-264.

W3C (2010a). HTML & CSS. Retrieved from
    http://www.w3.org/standards/webdesign/htmlcss

W3C (2010b). XML Essentials. Retrieved from http://www.w3.org/standards/xml/core

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced
    learning environments. *Educational Technology Research and Development,
    53*(4), 5-23.

Washbush, J., & Gosen, J. (2001). An exploration of game-derived learning in total
    enterprise simulations. *Simulation & Gaming, 32*(3), 281-296.

Whitner, R. B., & Balci, O. (1989). *Guidelines for selecting and using simulation model
    verification techniques* (No. TR-89-17). Blacksburg, Virginia: Virginia
    Polytechnic Institute and State University.

Wideman, H. H., Owston, R. D., Brown, C., Kushniruk, A., Ho, F., & Pitts, K. C. (2007).
    Unpacking the potential of educational gaming: A new tool for gaming research.
    Simulation & Gaming, 38(3), 10-30.

Williams, M. D., & Dodge, B. J. (1993). *Tracking and analyzing learner-computer
    interaction*. Paper presented at the Proceedings of Selected Research and

Development Presentations at the Convention of the Association for Educational Communications and Technology.

Winn, W. (2002). Current trends in educational technology research: The study of learning environments. *Educational Psychology Review, 14*(3), 331-352.

Wittgenstein, L. (2001). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Oxford: Blackwell Publishing Ltd.

Wolfe, J., & Crookall, D. (1998). Developing a scientific knowledge of simulation/gaming. *Simulation & Gaming, 29*(1), 7-19.

Yang, Q., & Koutsopoulos, H. N. (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research.C, 4*(3), 113-129.

Yazzie-Mintz, E. (2007). *Voices of students on engagement: A report on the 2006 high school survey of student engagement*. Retrieved from http://ceep.indiana.edu/hssse/pdf/HSSSE_2006_Report.pdf

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

Youngblood, S. (Ed.). (2006). *VV&A recommended practices guide*. Washington, DC: Department of Defense. Retrieved from http://vva.msco.mil/

Zeigler, B. P. (1990). *Object-oriented simulation with hierarchical, modular models*. San Diego, CA: Academic Press.

Zemke, R., & Rossett, A. (2002). A hard look at ISD. *Training, 39*(2), 26-34.

APPENDICES

Appendix A: Materials Provided for Expert Review

Activity Titles and Descriptions

Information Activities

Title	Description
Lunchmates	Observe carefully who lunches with whom each noon.
Committees	Find out who are members of the various formal committees set up in the school.
Social Network	Observe the out-of-school social patterns to learn who plays poker together, who bowls together, etc.

Diffusion Activities

Title	Description
Get Personal Info	Choose FIVE staff members for whom you would like to obtain personal information. (Cost: 1 week)
Talk To	You make a conscious effort, over a period of about one week, to engage any ONE person in a number of one to one conversations.
Ask Help	You ask any ONE of the staff for advice or for help in one of your projects... preparing some learning materials, setting up a demonstration, running a workshop, etc.
Pilot Test	You attempt to influence ONE teacher by asking to let you conduct an informal pilot test of peer tutoring with his/her students.
Site Visit	You select any FIVE persons to visit Lighthouse School, in the next state, where an exemplary tutoring program is in progress.
Print	You circulate a brochure describing the many advantages of peer tutoring to any FIVE persons.
Presentation	You get on the agenda of a regularly scheduled staff meeting to explain about peer tutoring and encourage discussion about it.
Demonstration	You invite the staff into a particular teacher's classroom (an adopter's!) to see peer tutoring in action.
Training Workshop (Self)	You conduct an in-service workshop which trains teachers in the operational details of setting up and carrying on a peer tutoring program in their classrooms.

Training Workshop (Professor)	You arrange to have Professor Portney of Centralia Teachers college conduct an in-service workshop on "Peer Tutoring: Its Role in Student Self-Development."
Materials Workshop	You conduct an inservice workshop in which teachers team up to develop creative materials-games, flash-cards, etc. for student tutor use.
Local Mass Media	You arrange to be interviewed about peer tutoring by a reporter from the local Eyewitness News program.
Compulsion	You persuade the principal to issue a memo directing all teachers to institute some form of peer tutoring in their classrooms next year. (Use only if the principal has adopted the innovation.)
Confrontation	You work behind the scenes with a group of parents, encouraging them to protest about the students poor reading achievement. They take their protest to a school board meeting. (Use only if you have used mass media twice.)

Predicted Successful Strategies

The following strategies are predicted lead to success in the *Diffusion Simulation Game* based on Rogers' theory.

1. **Target earlier adopters and opinion leaders to work toward critical mass.**
   Achieving a critical mass of adopters can increase the rate of adoption in a system. Innovators are among the first to try innovations and tend to act as gatekeepers in introducing new ideas into a system, but they exert little influence over others. Early Adopters tend to have greater social participation and more central positions in communication networks. Early Adopters also have the highest degree of opinion leadership, and adoption by opinion leaders can have the greatest impact on the rate of adoption.
2. **Use *Personal Information* and *Talk To.***
   The change agent should establish empathy and rapport in order to understand a client's needs, sociocultural values and beliefs, and previous exposure to related ideas.
3. **Use *Local Mass Media* and *Print* to gain points in Awareness and Interest among earlier adopters.**
   Mass media communication channels are especially effective in creating awareness and interest among earlier adopters, who tend to pay more attention to external sources of innovation than later adopters. *Print* is not exactly a mass medium in the DSG since its use is targeted at five staff members at a time; however, because earlier adopters tend to have more in common with change

agents (Rogers uses the term "homophilous"), they should be receptive to this approach.

4. **Use *Presentation* to gain points in Awareness and Interest.**
Because the change agent is giving the presentation, earlier adopters are more likely to be influenced. Nevertheless, once potential adopters (including later adopters) are aware of an innovation, they tend to be receptive to information about the innovation as they seek to reduce uncertainty about the consequences of adoption.

5. **Use *Demonstration*, especially by an opinion leader, to gain points in Interest for other potential adopters.**
Allowing potential adopters to see the innovation in use provides how-to knowledge and can reduce uncertainty about the consequences of adoption, especially if the demonstration is by a respected peer.

6. **Use *Site Visit* to gain points in Interest and move into Trial.**
The strengths of this activity are similar to those of *Demonstration* except the visitors see the innovation in use on a larger scale. The five targeted visitors are likely to see the innovation used in classrooms similar to their own and be persuaded to try the innovation on a limited basis.

7. **Use *Pilot Test* to gain additional points for those with some Interest or in Trial.**
The ability to try the innovation on a limited basis can further reduce uncertainty about the consequences of adoption. If used with opinion leaders or other highly connected individuals, it may have a similar effect on others in their social networks. Rogers says that trial by peers can serve as a vicarious trial for later adopters.

8. **Target highly connected individuals to gain additional points in Interest among later adopters in their social networks and move them into Trial.**
Mass media channels are less effective at later stages of the innovation-decision process; instead, interpersonal channels can be quite effective in persuading someone to try an innovation as people tend to trust the opinions of their near peers.

9. **Use *Training Workshop (Self)* and *Materials Workshop* to gain points in Trial.**
How-to knowledge is probably most essential when someone becomes willing to try an innovation. The change agent should provide knowledge and assistance regarding procedures and principles to further reduce uncertainty and increase confidence.

Appendix B: Generalizations from Rogers' (2003) *Diffusion of Innovations*

Chapter 5: The Innovation-Decision Process		
Category/Number	**Description**	**Page(s)**
Early v. Late Knowers	5.1 – 5.7	174
5.1	Earlier knowers of an innovation have more education than do later knowers.	174
5.2	Earlier knowers of an innovation have higher social status than do late knowers.	174
5.3	Earlier knowers of an innovation have more exposure to mass media channels of communication that [sic] do later knowers.	174
5.4	Earlier knowers of an innovation have more exposure to interpersonal channels than do later knowers.	174
5.5	Earlier knowers of an innovation have more contact with change agents than do later knowers.	174
5.6	Earlier knowers of an innovation have more social participation than do later knowers.	174
5.7	Earlier knowers of an innovation are more cosmopolite than later knowers.	174
Re-invention	5.8 – 5.10	183
5.8	Re-invention occurs at the implementation stage for many innovations and for many adopters.	183
5.9	A higher degree of re-invention leads to a faster rate of adoption of an innovation.	183
5.10	A higher degree of re-invention leads to a higher degree of sustainability of an innovation.	183
Discontinuance	5.11	191
5.11	Later adopters are more likely to discontinue innovations than are earlier adopters.	191
Stages of the Innovation-Decision Process	5.12	198
5.12	Stages exist in the innovation-decision process.	198
Mass Media v. Interpersonal Channels	5.13	205
5.13	Mass media channels are relatively more important at the knowledge state, and interpersonal channels are relatively more important at the persuasion stage in the innovation-decision process.	205
Cosmopolite v. Localite Channels	5.14	207
5.14	Cosmopolite channels are relatively more important at the knowledge state, and localite channels are relatively more important at the	207

	persuasion stage in the innovation-decision process.	
Communication Channels by Adopter Categories	5.15 – 5.16	211 – 213
5.15	Mass media channels are relatively more important than interpersonal channels for earlier adopters than for later adopters.	211
5.16	Cosmopolite channels are relatively more important than localite channels for earlier adopters than for later adopters.	213
The Innovation-Decision Period	5.17	214
5.17	The rate of awareness-knowledge for an innovation is more rapid than its rate of adoption.	214
The Length of the Innovation-Decision Period by Adopter Category	5.18	214
5.18	Earlier adopters have a shorter innovation-decision period than do later adopters.	214

Chapter 6: Attributes of Innovations and Their Rate of Adoption		
Category/Number	Description	Page(s)
Relative Advantage	6.1	233
6.1	The relative advantage of an innovation, as perceived by members of a social system, is positively related to its rate of adoption.	233
Compatibility	6.2	249
6.2	The compatibility of an innovation, as perceived by members of a social system, is positively related to its rate of adoption.	249
Complexity	6.3	257
6.3	The complexity of an innovation, as perceived by members of a social system, is negatively related to its rate of adoption.	257
Trialability	6.4	258
6.4	The trialability of an innovation, as perceived by the members of a social system, is positively related to its rate of adoption.	258
Observability	6.5	258
6.5	The observability of an innovation, as perceived by members of a social system, is positively related to its rate of adoption.	258

Chapter 7: Innovativeness and Adopter Categories		
Category/Number	**Description**	**Page(s)**
The S-Shaped Curve of Adoption and Normality	7.1	275
7.1	Adopter distributions follow a bell-shaped curve over time and approach normality.	275
Socioeconomic Characteristics	7.2 – 7.7	288
7.2	Earlier adopters are no different from later adopters in age.	288
7.3	Earlier adopters have more years of formal education than do later adopters.	288
7.4	Earlier adopters are more likely to be literate than are later adopters.	288
7.5	Earlier adopters have higher social status than do later adopters.	288
7.6	Earlier adopters have a greater degree of upward social mobility than do later adopters.	288
7.7	Earlier adopters have a greater degree of upward social mobility than do later adopters.	288
Personality Variables	7.8 – 7.17	289 – 290
7.8	Earlier adopters have greater empathy than do later adopters.	289
7.9	Earlier adopters may be less dogmatic than are later adopters.	289
7.10	Earlier adopters have a greater ability to deal with abstractions than do later adopters.	289
7.11	Earlier adopters have greater rationality than do later adopters.	289
7.12	Earlier adopters have more intelligence than do later adopters.	289
7.13	Earlier adopters have a more favorable attitude toward change than do later adopters.	290
7.14	Earlier adopters are better able to cope with uncertainty and risk than are later adopters.	290
7.15	Earlier adopters have a more favorable attitude toward science than do later adopters.	290
7.16	Earlier adopters are less fatalistic than are later adopters.	290
7.17	Earlier adopters have higher aspirations (for formal education, higher status, occupations, and so on) than do later adopters.	290
Communication Behavior	7.18 – 7.26	290 – 291
7.18	Earlier adopters have more social participation than do later adopters.	290
7.19	Earlier adopters are more highly interconnected through interpersonal networks in their social system than are later adopters.	290

7.20	Earlier adopters are more cosmopolite than are later adopters.	290
7.21	Earlier adopters have more contact with change agents than do later adopters.	291
7.22	Earlier adopters have greater exposure to mass media communication channels than do later adopters.	291
7.23	Earlier adopters have greater exposure to interpersonal communication channels than do later adopters.	291
7.24	Earlier adopters seek information about innovations more actively than do later adopters.	291
7.25	Earlier adopters have greater knowledge of innovations than do later adopters.	291
7.26	Earlier adopters have a higher degree of opinion leadership than do later adopters.	291

Chapter 8: Diffusion Networks		
Category/Number	**Description**	**Page(s)**
Homophily as a Barrier to Diffusion	8.1 – 8.2	307 - 308
8.1	Interpersonal diffusion networks are mostly homophilous.	307
8.2	When interpersonal diffusion networks are heterophilous, followers seek opinion leaders of higher socioeconomic status, with more formal education, with a greater degree of mass media exposure, who are more cosmopolite, have greater contact with change agents, and are more innovative.	308
External Communication	8.3 – 8.5	316 – 317
8.3	Opinion leaders have greater exposure to mass media than their followers.	316
8.4	Opinion leaders are more cosmopolite than their followers.	317
8.5	Opinion leaders have greater contact with change agents than their followers.	317
Accessibility	8.6	317
8.6	Opinion leaders have greater social participation than their followers.	317
Socioeconomic Status	8.7	318
8.7	Opinion leaders have higher socioeconomic status than their followers.	318
Innovativeness	8.8	318

8.8	Opinion leaders are more innovative than their followers.	318
Innovativeness, Opinion Leadership, and System Norms	8.9	318
8.9	When a social system's norms favor change, opinion leaders are more innovative, but when the system's norms do not favor change, opinion leaders are not especially innovative.	318
Networks in the Diffusion of a Medical Drug	8.10	330
8.10	The network interconnectedness of an individual in a social system is positively related to the individual's innovativeness.	330
The Strength-of-Weak-Ties Theory	8.11	340
8.11	The information-exchange potential of communication network links is negatively related to their degree of (1) communication proximity and (2) homophily.	340
Who Is Linked to Whom in Networks?	8.12	341
8.12	Individuals tend to be linked to others who are close to them in physical distance and who are relatively homophilous in social characteristics.	341
Why Do Individuals Adopt Prior to the Critical Mass?	8.13	359
8.13	An individual is more likely to adopt an innovation if more of the other individuals in his or her personal network have adopted previously.	359

Chapter 9: The change Agent		
Category/Number	Description	Page(s)
Change Agent Effort	9.1	373
9.1	Change agents' success in securing the adoption of innovations by clients is positively related to the extent of change agent effort in contacting clients.	373
Client Orientation	9.2	374 – 375
9.2	Change agents' success in securing the adoption of innovations by clients is positively related to a client orientation, rather than to a change agency orientation.	374 – 375
Compatibility with Clients' Needs	9.3	375

9.3	Change agents' success in securing the adoption of innovations by clients is positively related to the degree to which a diffusion program is compatible with clients' needs.	375
Change Agent Empathy	9.4	377
9.4	Change agents' success in securing the adoption of innovations by clients is positively related to empathy with clients.	377
Homophily and Change Agent Contact	9.5 – 9.8	382
9.5	Contact with change agents is positively related to higher socioeconomic status among clients.	382
9.6	Contact with change agents is positively related to greater social participation by clients.	382
9.7	Contact with change agents is positively related to higher formal education among clients.	382
9.8	Contact with change agents is positively related to cosmopoliteness among clients.	382
Change Agents' Contact with Lower-Status Clients	9.9	384
9.9	Change agents' success in securing the adoption of innovations by clients is positively related to their homophily with clients.	384
Change Agent Credibility	9.10	385
9.10	Change agents' success in securing the adoption of innovations by clients is positively related to credibility in the clients' eyes.	385
The Use of Opinion Leaders	9.11	388
9.11	Change agents's success in securing the adoption of innovations by clients is positively related to the extent that he or she works through opinion leaders.	388
Clients' Evaluative Ability	9.12	390
9.12	Change agents' success in securing the adoption of innovations by clients is positively related to increasing clients' ability to evaluate innovations.	390

Chapter 10: Innovation in Organizations		
Category/Number	Description	Page(s)
Size and Organizational Innovativeness	10.1	409
10.1	Larger organizations are more innovative.	409
Structural Characteristics and Organizational Innovativeness	10.2	412 – 413
10.2	Each of the organizational structure variables	412 – 413

	may be related to innovation in one direction during the initiation phases of the innovation process, and in the opposite direction during the implementation phases.	
The Role of Champions	10.3	414
10.3	The presence of an innovation champion contributes to the success of an innovation within an organization.	414
Agenda-Setting	10.4	422
10.4	A performance gap can trigger the innovation process.	422
Innovation and Organization Structure	10.5	425
10.5	Both the innovation and the organization usually change in the innovation process in an organization.	425

Chapter 11: Consequences of Innovations		
Category/Number	Description	Page(s)
Desirable Versus Undesirable Consequences	11.1	445
11.1	The effects of an innovation usually cannot be managed so as to separate the desirable from the undesirable consequences.	445
Anticipated Versus Unanticipated Consequences	11.2	449
11.2	The undesirable, indirect, and unanticipated consequences of an innovation usually go together, as do the desirable, direct, and anticipated consequences.	449
Form, Function, and Meaning of an Innovation	11.3	451
11.3	Change agents more easily anticipate the form and function of an innovation for their clients than its meaning.	451
Gap-Widening Consequences of the Diffusion of Innovations	11.4 – 11.5	460 – 461
11.4	The consequences of the diffusion of innovations usually widen the socioeconomic gap between the earlier and later adopting categories in a system.	460
11.5	The consequences of the diffusion of innovation usually widen the socioeconomic gap between the audience segments previously high and low in socioeconomic status.	460 – 461

Social Structure and the Equality of Consequences	11.6	463
11.6	A system's social structure partly determines the equality versus inequality of an innovation's consequences.	463
Wider Gaps Are Not Inevitable	11.7	467
11.7	When special efforts are made by a diffusion agency, it is possible to narrow, or at least not to widen, socioeconomic gaps in a social system.	467

Appendix C: Sample Currentinfo.xml File from the *Diffusion Simulation Game*

This file contains data about the current game state, including the number of weeks elapsed, the total number of adopters, and the number of points obtained in each innovation-decision stage for each staff member.

```
<?xml version="1.0" standalone="yes"?>
<week>72</week>

<totaladopters>11</totaladopters>

<totalsteps>54</totalsteps>

<lunchmates>1</lunchmates>
<committees>1</committees>
<social>1</social>

<massmedia>0</massmedia>

<compulsion>0</compulsion>
<confrontation>0</confrontation>

<principal>
<code>A</code>
<title>Principal</title>
<info>
Very ambitious (has a 20-year plan); member of the Rotary Club and local Republican
Club (active in both); delegates authority to able administrative assistants and runs a
"tight ship." Has a "master's-plus" in administration.
</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>5</trial>
<adoption>1</adoption>
<talkto>6</talkto>
</principal>

<secretary>
<code>B</code>
<title>Secretary</title>
<info>
Has been in this school since it was built and quite indispensable to its smooth
functioning. Runs most faculty social functions.
</info>
<awareness>1</awareness>
<interest>4</interest>
```

```
<trial>0</trial>
<adoption>0</adoption>
<talkto>1</talkto>
</secretary>

<janitor>
<code>C</code>
<title>Janitor</title>
<info>
Fond of children, but stern. He tends to allow extensive use of the school building, but
has strict rules and is inflexible about infractions.
</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>0</trial>
<adoption>0</adoption>
<talkto>1</talkto>
</janitor>

<mathchairman>
<code>D</code>
<title>Math Chairman</title>
<info>
A veteran in the school, he runs the most experienced department with a minimum of
effort. Is involved much more in out-of-school activities such as the local garden club and
conservation organization. Still regrets the repeal of prohibition.
</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>2</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</mathchairman>

<mathteacher1>
<code>E</code>
<title>Math Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>6</interest>
<trial>2</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</mathteacher1>
```

```
<mathteacher2>
<code>F</code>
<title>Math Teacher</title>
<info>
Just about the most respected and liked teacher in the school. Students enjoy the
humorous examples he uses in teaching algebra. Exudes a sense of self-confidence and
has no enemies among the staff. Serves as advisor to the Student Council. Never misses a
PTA meeting.
</info>
<awareness>1</awareness>
<interest>3</interest>
<trial>3</trial>
<adoption>1</adoption>
<talkto>3</talkto>
</mathteacher2>

<sciencechairman>
<code>G</code>
<title>Science Chairman</title>
<info>
Known more for his eagerness and energy than administrative skill. He comes up with
new instructional ideas faster than they can be implemented since he is working on a
Master's and often tries out suggestions discussed in his graduate classes. Among older
staff, he's considered somewhat erratic.
</info>
<awareness>1</awareness>
<interest>2</interest>
<trial>2</trial>
<adoption>1</adoption>
<talkto>1</talkto>
</sciencechairman>

<scienceteacher1>
<code>H</code>
<title>Science Teacher</title>
<info>
Has spent years collecting specimens, preparing new instructional materials, and
organizing laboratory procedures. Open to new techniques of proven value. Insists on an
orderly classroom. His success in teaching makes his respected by alumni and fellow
teachers.
</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>3</trial>
<adoption>0</adoption>
```

```
<talkto>2</talkto>
</scienceteacher1>

<scienceteacher2>
<code>I</code>
<title>Science Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>0</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</scienceteacher2>

<socialstudieschairwoman>
<code>J</code>
<title>Social Studies Chairwoman</title>
<info>

</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>2</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</socialstudieschairwoman>

<socialstudiesteacher1>
<code>K</code>
<title>Social Studies Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>0</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</socialstudiesteacher1>

<socialstudiesteacher2>
<code>L</code>
<title>Social Studies Teacher</title>
<info>
```

```
</info>
<awareness>1</awareness>
<interest>2</interest>
<trial>2</trial>
<adoption>1</adoption>
<talkto>0</talkto>
</socialstudiesteacher2>

<languageartschairman>
<code>M</code>
<title>Language Arts Chairman</title>
<info>
```
Likes his job and well he should for he has an energetic and cohesive department.
Encourages inter-visitation among his teachers. Otherwise, he follows established routine
and urges his teachers to develop efficient classroom management procedures. Seems to
be on the inside in school "politics."
```
</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>4</trial>
<adoption>1</adoption>
<talkto>2</talkto>
</languageartschairman>

<languageartsteacher1>
<code>N</code>
<title>Language Arts Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>1</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</languageartsteacher1>

<languageartsteacher2>
<code>O</code>
<title>Language Arts Teacher</title>
<info>
```
Just awarded tenure. Considered to be a good teacher by her colleague and the chairman.
Although usually not among the first to try new methods, she keeps up with recent
developments in literature, the theater, and the arts in general. Is faculty advisor to the
Drama Club.

```
</info>
<awareness>1</awareness>
<interest>3</interest>
<trial>5</trial>
<adoption>1</adoption>
<talkto>3</talkto>
</languageartsteacher2>

<foreignlanguageteacher>
<code>P</code>
<title>Foreign Language Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>2</interest>
<trial>2</trial>
<adoption>1</adoption>
<talkto>0</talkto>
</foreignlanguageteacher>

<industrialartsteacher>
<code>Q</code>
<title>Industrial Arts Teacher</title>
<info>
A very good teacher and an acclaimed metal sculptor. Advocates hands-on education and
encouragement of students' natural curiosity. Runs the metal and woods shops, and
teaches drafting and mechanical drawing as well. Seems always to have time for students
but spends little time socializing among the staff.
</info>
<awareness>1</awareness>
<interest>3</interest>
<trial>3</trial>
<adoption>1</adoption>
<talkto>1</talkto>
</industrialartsteacher>

<artteacher>
<code>R</code>
<title>Art Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>7</interest>
<trial>1</trial>
```

```
<adoption>0</adoption>
<talkto>0</talkto>
</artteacher>

<musicteacher>
<code>S</code>
<title>Music Teacher</title>
<info>
Carries a heavy load of extracurricular activities: the school band, orchestra, spring
festival, winter musicale and, of course, appreciation classes. Loves music, enjoys his
colleagues and likes his students although he wishes they had more talent and
enthusiasm.
</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>4</trial>
<adoption>1</adoption>
<talkto>1</talkto>
</musicteacher>

<boyspeteacher>
<code>T</code>
<title>Boys' P.E. Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>6</interest>
<trial>3</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</boyspeteacher>

<girlspeteacher>
<code>U</code>
<title>Girls P.E. Teacher</title>
<info>

</info>
<awareness>1</awareness>
<interest>5</interest>
<trial>3</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</girlspeteacher>
```

```
<homeeconomicsteacher>
<code>V</code>
<title>Home Economics Teacher</title>
<info>
Popular with the students whom she sees in class and at after-school functions. The
faculty, too feel comfortable around her and rely on her predictable domesticity.
</info>
<awareness>1</awareness>
<interest>4</interest>
<trial>4</trial>
<adoption>1</adoption>
<talkto>1</talkto>
</homeeconomicsteacher>

<guidancecounselor>
<code>W</code>
<title>Guidance Counselor</title>
<info>
Formerly in personnel in a small advertising firm in the city. Is happy with this position,
but aware of its limitations. Engaged in developing alternative procedures for dealing
with discipline problems. The object of some resentment from teachers who feel he is
more supportive of difficult students than is necessary.
</info>
<awareness>1</awareness>
<interest>3</interest>
<trial>3</trial>
<adoption>1</adoption>
<talkto>1</talkto>
</guidancecounselor>

<mediaspecialist>
<code>X</code>
<title>Library/AV Coordinator</title>
<info>

</info>
<awareness>1</awareness>
<interest>3</interest>
<trial>1</trial>
<adoption>0</adoption>
<talkto>0</talkto>
</mediaspecialist>
```

Appendix D: Sample History.xml File from the *Diffusion Simulation Game*

This file stores information about each turn in the game session, including the activity selected, staff members selected, and the text of the feedback that the player received.

```
<?xml version="1.0" standalone="yes"?>
<step1>
<strategy>
Personal Information
</strategy>
<staff>
A (Principal)<br>B (Secretary)<br>C (Janitor)<br>F (Math Teacher)<br>W (Guidance
Counselor)<br>
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step1>

<step2>
<strategy>
Talk To
</strategy>
<staff>
B (Secretary)
</staff>
<feedback>
She now understand that your ideas (and you!) are worth bothering the
Principal with.
</feedback>
<adopters>
0
</adopters>
</step2>

<step3>
<strategy>
Talk To
</strategy>
<staff>
F (Math Teacher)
</staff>
<feedback>
```

<p>A potentially useful contact; if he/she adopts, a number of others will be favorably disposed.</p>
<p>Unfortunately, this is the week his/her family was moving into a new home... no time for serious talk. May be worth trying again later.</p>
</feedback>
<adopters>
0
</adopters>
</step3>

<step4>
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>
<feedback>
Unfortunately, you were hired by the district rather than by the school, so it's not going to be easy to get the principal on your side. However, if you do, your task will be much easier. He will not give you an appointment at this time, but your initiative is beginning to impress him.
</feedback>
<adopters>
0
</adopters>
</step4>

<step5>
<strategy>
Committees
</strategy>
<staff>
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step5>

<step6>
<strategy>
Lunchmates
</strategy>
<staff>

```
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step6>

<step7>
<strategy>
Social Contacts
</strategy>
<staff>
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step7>

<step8>
<strategy>
Talk To
</strategy>
<staff>
C (Janitor)
</staff>
<feedback>
He now sees that your efforts are for the good of the kids. He'll be
cooperative.
</feedback>
<adopters>
0
</adopters>
</step8>

<step9>
<strategy>
Talk To
</strategy>
<staff>
F (Math Teacher)
</staff>
<feedback>
```

<p>A potentially useful contact; if he/she adopts, a number of others will be favorably disposed.</p>
<p>Unfortunately, this is the week his/her family was moving into a new home... no time for serious talk. May be worth trying again later.</p>
</feedback>
<adopters>
0
</adopters>
</step9>

<step10>
<strategy>
Presentation
</strategy>
<staff>
</staff>
<feedback>
Your presentation appeals to the teachers' sense of professionalism.
</feedback>
<adopters>
0
</adopters>
</step10>

<step11>
<strategy>
Talk To
</strategy>
<staff>
G (Science Chairman)
</staff>
<feedback>
If you gather some information about this person before you talk to him/her you can probably be much more effective in reaching him/her.
</feedback>
<adopters>
0
</adopters>
</step11>

<step12>
<strategy>
Personal Information
</strategy>
<staff>
G (Science Chairman)

```
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step12>

<step13>
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>
<feedback>
<p>Oh-oh… an irate parent just appeared at the Principal's door.</p>
<p>This is not going to be a good time for discussion of your concerns.</p>
<p>But getting through to the Principal is worth the effort of trying again.</p>
</feedback>
<adopters>
0
</adopters>
</step13>

<step14>
<strategy>
Presentation
</strategy>
<staff>
</staff>
<feedback>
Many of the staff are hearing about peer tutoring for the first time; the consensus is that it
sounds interesting.
</feedback>
<adopters>
0
</adopters>
</step14>

<step15>
<strategy>
Print
</strategy>
<staff>
A (Principal)<br>D (Math Chairman)<br>F (Math Teacher)<br>
```

\</staff\>
\<feedback\>
Printed blurbs strike most people as impersonal, but your material is circulated and it arouses some mild interest.
\</feedback\>
\<adopters\>
0
\</adopters\>
\</step15\>

\<step16\>
\<strategy\>
Presentation
\</strategy\>
\<staff\>
\</staff\>
\<feedback\>
The prior item on the meeting's agenda is a severe cutback on all instructional supplies and materials. This dampens interest in new ventures.
\</feedback\>
\<adopters\>
0
\</adopters\>
\</step16\>

\<step17\>
\<strategy\>
Talk To
\</strategy\>
\<staff\>
O (Language Arts Teacher)
\</staff\>
\<feedback\>
If you gather some information about this person before you talk to him/her you can probably be much more effective in reaching him/her.
\</feedback\>
\<adopters\>
0
\</adopters\>
\</step17\>

\<step18\>
\<strategy\>
Personal Information
\</strategy\>
\<staff\>

O (Language Arts Teacher)
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step18>

<step19>
<strategy>
Talk To
</strategy>
<staff>
O (Language Arts Teacher)
</staff>
<feedback>
This person will want to consider your idea thoughtfully and observe it in action before deciding to adopt it.
</feedback>
<adopters>
0
</adopters>
</step19>

<step20>
<strategy>
Personal Information
</strategy>
<staff>
M (Language Arts Chairman)
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step20>

<step21>
<strategy>
Talk To
</strategy>
<staff>
M (Language Arts Chairman)
</staff>

<feedback>
He/she listens attentively to your ideas and shares them with his/her out-of-school
compatriots.
</feedback>
<adopters>
0
</adopters>
</step21>

<step22>
<strategy>
Personal Information
</strategy>
<staff>
H (Science Teacher)
</staff>
<feedback>
</feedback>
<adopters>
0
</adopters>
</step22>

<step23>
<strategy>
Talk To
</strategy>
<staff>
H (Science Teacher)
</staff>
<feedback>
He/she listens attentively to your ideas and shares them with his/her out-of-school
compatriots.
</feedback>
<adopters>
0
</adopters>
</step23>

<step24>
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>

<feedback>
Unfortunately, the Principal really is too busy to see you today, but
the Secretary promises to get you in next time.
</feedback>
<adopters>
0
</adopters>
</step24>

<step25>
<strategy>
Pilot Test
</strategy>
<staff>
M (Language Arts Chairman)
</staff>
<feedback>
This seems to have some practical applications in his/her teaching
area and he/she is willing to go along at least as far as a tryout.
</feedback>
<adopters>
0
</adopters>
</step25>

<step26>
<strategy>
Pilot Test
</strategy>
<staff>
F (Math Teacher)
</staff>
<feedback>
This person is eager to see peer tutoring in action; sees possible applications to his/her
own classroom.
</feedback>
<adopters>
0
</adopters>
</step26>

<step27>
<strategy>
Pilot Test
</strategy>
<staff>

H (Science Teacher)
</staff>
<feedback>
He/she just doesn't see how peer tutoring applies to his/her subject.
</feedback>
<adopters>
0
</adopters>
</step27>

<step28>
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>
<feedback>
The Principal would like to help, but not in an obvious way. After
all, your very presence makes the teachers feel that "somebody out there" thinks
they have been doing an inadequate job academically. That hurts!
</feedback>
<adopters>
0
</adopters>
</step28>

<step29>
<strategy>
Presentation
</strategy>
<staff>
</staff>
<feedback>
Your presentation is a welcome change from having to figure out who gets lunch duty
this week.
</feedback>
<adopters>
0
</adopters>
</step29>

<step30>
<strategy>
Talk To
</strategy>

```
<staff>
F (Math Teacher)
</staff>
<feedback>
This key person approves of your idea in principle and discusses it at lunch. Others listen.
</feedback>
<adopters>
0
</adopters>
</step30>

<step31>
<strategy>
Site Visit
</strategy>
<staff>
A (Principal)<br>F (Math Teacher)<br>H (Science Teacher)<br>J (Social Studies
Chairwoman)<br>M (Language Arts Chairman)<br>
</staff>
<feedback>
Lighthouse School shines brightly. The team is illuminated.
</feedback>
<adopters>
1
</adopters>
</step31>

<step32>
<strategy>
Talk To
</strategy>
<staff>
M (Language Arts Chairman)
</staff>
<feedback>
He/she listens attentively to your ideas and shares them with his/her out-of-school
compatriots.
</feedback>
<adopters>
2
</adopters>
</step32>

<step33>
<strategy>
Talk To
```

```
</strategy>
<staff>
H (Science Teacher)
</staff>
<feedback>
He/she listens attentively to your ideas and shares them with his/her out-of-school
compatriots.
</feedback>
<adopters>
2
</adopters>
</step33>

<step34>
<strategy>
Demonstration
</strategy>
<staff>
F (Math Teacher)
</staff>
<feedback>
A demonstration with high impact! If this demo. teacher favors peer tutoring, lots of
others will go along with it.
</feedback>
<adopters>
4
</adopters>
</step34>

<step35>
<strategy>
Talk To
</strategy>
<staff>
G (Science Chairman)
</staff>
<feedback>
Immediate rapport! He/she not only supports your ideas but offers to help spread the
word to other staff members.
</feedback>
<adopters>
5
</adopters>
</step35>

<step36>
```

<strategy>
Talk To
</strategy>
<staff>
Q (Industrial Arts Teacher)
</staff>
<feedback>
If you gather some information about this person before you talk to him/her you can probably be much more effective in reaching him/her.
</feedback>
<adopters>
5
</adopters>
</step36>

<step37>
<strategy>
Personal Information
</strategy>
<staff>
Q (Industrial Arts Teacher)
</staff>
<feedback>
</feedback>
<adopters>
5
</adopters>
</step37>

<step38>
<strategy>
Talk To
</strategy>
<staff>
Q (Industrial Arts Teacher)
</staff>
<feedback>
Although initially skeptical, this person is warming to the idea. He/she is willing to go along with it if others will.
</feedback>
<adopters>
6
</adopters>
</step38>

<step39>

```
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>
<feedback>
The Principal would like to help, but not in an obvious way. After
all, your very presence makes the teachers feel that "somebody out there" thinks
they have been doing an inadequate job academically. That hurts!
</feedback>
<adopters>
6
</adopters>
</step39>

<step40>
<strategy>
Pilot Test
</strategy>
<staff>
D (Math Chairman)
</staff>
<feedback>
If you gather some information about this person before you talk to him/her you can
probably be much more effective in reaching him/her.
</feedback>
<adopters>
6
</adopters>
</step40>

<step41>
<strategy>
Personal Information
</strategy>
<staff>
D (Math Chairman)<br>
</staff>
<feedback>
</feedback>
<adopters>
6
</adopters>
</step41>
```

<step42>
<strategy>
Presentation
</strategy>
<staff>
</staff>
<feedback>
Your presentation is a welcome change from having to figure out who gets lunch duty this week.
</feedback>
<adopters>
6
</adopters>
</step42>

<step43>
<strategy>
Talk To
</strategy>
<staff>
O (Language Arts Teacher)
</staff>
<feedback>
This person will want to consider your idea thoughtfully and observe it in action before deciding to adopt it.
</feedback>
<adopters>
6
</adopters>
</step43>

<step44>
<strategy>
Presentation
</strategy>
<staff>
</staff>
<feedback>
Earlier business on the agenda drags on and on, so many staff members have drifted away before your presentation starts.
</feedback>
<adopters>
6
</adopters>
</step44>

<step45>
<strategy>
Demonstration
</strategy>
<staff>
M (Language Arts Chairman)
</staff>
<feedback>
Good choice of a demo. teacher! All the staff who had interest in peer tutoring showed up and were impressed that this teacher favors the idea.
</feedback>
<adopters>
6
</adopters>
</step45>

<step46>
<strategy>
Personal Information
</strategy>
<staff>
V (Home Economics Teacher)
</staff>
<feedback>
</feedback>
<adopters>
6
</adopters>
</step46>

<step47>
<strategy>
Talk To
</strategy>
<staff>
V (Home Economics Teacher)
</staff>
<feedback>
This person will carefully weigh the advantages and disadvantages before making a decision.
</feedback>
<adopters>
7
</adopters>
</step47>

<step48>
<strategy>
Talk To
</strategy>
<staff>
O (Language Arts Teacher)
</staff>
<feedback>
Although initially skeptical, this person is warming to the idea. He/she is willing to go along with it if others will.
</feedback>
<adopters>
8
</adopters>
</step48>

<step49>
<strategy>
Talk To
</strategy>
<staff>
S (Music Teacher)
</staff>
<feedback>
If you gather some information about this person before you talk to him/her you can probably be much more effective in reaching him/her.
</feedback>
<adopters>
8
</adopters>
</step49>

<step50>
<strategy>
Personal Information
</strategy>
<staff>
S (Music Teacher)
</staff>
<feedback>
</feedback>
<adopters>
8
</adopters>
</step50>

<step51>
<strategy>
Talk To
</strategy>
<staff>
S (Music Teacher)
</staff>
<feedback>
This person will want to consider your idea thoughtfully and observe it in action before deciding to adopt it.
</feedback>
<adopters>
9
</adopters>
</step51>

<step52>
<strategy>
Talk To
</strategy>
<staff>
A (Principal)
</staff>
<feedback>
Although your presence was not requested by the school, the federal
funds you bring are most welcome. The Principal wants to see that the money is
used as effectively as possible.
</feedback>
<adopters>
10
</adopters>
</step52>

<step53>
<strategy>
Talk To
</strategy>
<staff>
W (Guidance Counselor)
</staff>
<feedback>
This person will want to consider your idea thoughtfully and observe it in action before deciding to adopt it.
</feedback>
<adopters>
11

</adopters>
</step53>

<step54>
<strategy>
Print
</strategy>
<staff>
D (Math Chairman)
H (Science Teacher)
J (Social Studies Chairwoman)
U (Girls P.E. Teacher)
X (Library/AV Coordinator)
</staff>
<feedback>
Printed blurbs strike most people as impersonal, but your material is circulated and it arouses some mild interest.
</feedback>
<adopters>
11
</adopters>
</step54>

Appendix E: Text of E-mail Sent to Experts

Dear _____

You have been identified as someone with expertise on Everett Rogers' diffusion of innovations theory, either because you have published studies or critiques of the theory or you have taught classes that covered the theory.

The title of my dissertation is *Analyzing Interaction Patterns to Verify a Simulation/Game Model*. The case selected for analysis is the *Diffusion Simulation Game* (DSG: https://www.indiana.edu/~simed/istdemo/), which has as its primary conceptual model Rogers' diffusion of innovations theory. The DSG is a simulation game in which the player takes on the role of a change agent in a junior high school. The player's objective is to persuade as many of the 22 staff members as possible to adopt an innovation—peer tutoring. To be effective, players must learn appropriate application and sequencing of available diffusion activities given adopters of various types at different points in the innovation-decision process. I am examining the fidelity of the DSG to Rogers' theory by collecting and analyzing the results of players' choices in the game and comparing those results with what is predicted by Rogers' theory to be effective in the real world.

Attached you will find a document containing descriptions of the diffusion activities available in the game, followed by nine strategies for using those activities that I predict should be successful in the game based on my understanding of Rogers' theory. For each strategy, I provide my rationale and my approach to scoring each turn in every game.

Will you do me the favor of reviewing these strategies? For each strategy, simply indicate whether you agree that it should be effective in moving people toward adoption *based on your understanding of Rogers' theory*. If you disagree, please provide an explanation of your reasoning. Also feel free to offer alternative strategies that I should consider. You may add your comments directly to the attached document or send them to me in a separate document. My email address is rodmyers@indiana.edu. I would appreciate receiving your response by [date TBD]. Please feel free to forward this to anyone you know who might be considered highly knowledgeable about Rogers' theory and willing to assist me in this endeavor.

If you have any questions, feel free to contact me. If you would like to receive a copy of my finished dissertation, please let me know and I'll gladly send it once I have completed it.

Thank you for your time and assistance.

Rodney D. Myers
Ph.D. Candidate in Instructional Systems Technology
Indiana University – Bloomington
rodmyers@indiana.edu

Appendix F: Coding Errors in the Online *Diffusion Simulation Game*

Presentation (25b-f): Supposed to award no points if June or December, but the code to check weeks should use ANDs instead of ORs (e.g., $week != 10 && $week !=11, etc.). NOT FIXED FOR ROBOPLAY

Presentation (25f): Supposed to award no points to B, C, E, K, R, T, but the code should use ANDs instead of ORs.

Print (all feedback): Supposed to award no points if June or December, but the code to check weeks should use ANDs instead of ORs (e.g., $week != 10 && $week !=11, etc.). NOT FIXED FOR ROBOPLAY

Print (22b): Supposed to award points to 3 of 5 selected staff instead of all 5. FIXED

Print (22c): Supposed to award points in Awareness only instead of Awareness or Interest. FIXED

Print (22d): Supposed to award points in Awareness only to 2 of 5 selected staff. FIXED

Print (22e): Supposed to award points to 2 of 5 selected staff instead of all 5. FIXED

Talk To (4g): Missing "[$i]" to check staff member's Interest ("$interest==0"). FIXED

Talk To (11a): Typo—"$tiral" should be $trial. FIXED

Talk To (11d): Are the boys' and girls' PE teachers in the same department? If so, U should be awarded points when T is selected. NOT FIXED FOR ROBOPLAY

Talk To (11f): Supposed to award a point to Q instead of P. FIXED

Talk To (7e, 7f): Never selected because code says "rand(1,4)" instead of "rand(1,6)" DISCOVERED AFTER ROBOPLAY—NOT FIXED

Demonstration (31b): Unclear feedback was resolved by awarding no points. Possible selected staff are B, C, E, K, N, R, W. Feedback is "Gain 1 point for any two who have completed the interest phase." NOT FIXED IN ROBOPLAY

Local Mass Media (37b): Supposed to award points in Awareness or Interest only, but possibly awards point in Trial if staff member is 1 point away from Trial (37d does this correctly). FIXED

Site Visit (21d): Supposed to award points to only 3 of 5 selected staff instead of all 5. FIXED

Site Visit (24): Never selected? This is the feedback when one of the selected staff members has not previously been selected for Personal Information. The cost is supposed to be one week.

Materials Workshop (36e): Never selected because code says "rand(1,4)" instead of "rand(1,5)" DISCOVERED AFTER ROBOPLAY—NOT FIXED

NOT USED IN ROBOPLAY

Ask Help (13c): Typo—"totoalpoints" should be "totalpoints". NOT FIXED FOR ROBOPLAY

Compulsion (38b): Never selected? There is a 1-in-6 chance of selecting 38a or 38b and a 2-in-6 chance of selecting 38c or 38d. However, 38a was selected 36.6% of the time, 38c was 32.8%, 38d was 30.5%, and 38b was never selected.

Rodney D. Myers

1726 E Thornton Dr. Bloomington, IN 47401
Phone: 408.607.0714
Email: rodmyers@indiana.edu

Education

B.A., Ball State University
    Major: English — *July 1982*

M.A., Ball State University
    Major: English — *August 1983*

De Anza Community College
    Courses in screenwriting, production, film history and theory — *1988 - 1992*

M.A., San Jose State University
    Major: Instructional Technology — *May 2006*

Ph.D., Indiana University (expected in May, 2012) — *2007 - present*
    Major: Instructional Systems Technology
    Minor: Human Computer Interaction - Design

Honors and Awards

School of Education Fellowship (Indiana University) — *2007 - 2011*

First prize for poster presentation (Games+Learning+Society Conference) — *2008*

"Richard B. Lewis Outstanding Graduate Student" award in Instructional Technology (San Jose State University) — *2006*

President's Award for "Outstanding Service and Achievement" (Santa Clara University) — *1995*

School of Education Graduate Assistantship (Ball State University) — *1982-83*

Teaching and Research Interests

Design, development, and use of games and simulations for learning
Design and development of instruction for emerging technologies
Design and development of ubiquitous learning
Educational design research

Teaching Experience

Adjunct Lecturer **,** Indiana University**,** IST Department — *2009*
    Taught an online graduate-level class in computer-mediated learning (R547).

Associate Instructor, Indiana University, IST Department — *2008 - 2009*
    Taught classes in technology integration (W301) for pre-service teachers.

Teaching Assistant, Indiana University, IST Department — *2008*
    Provided technical support and feedback on deliverables to graduate students in R547 (taught by Dr. Theodore Frick).

Adjunct Lecturer , Santa Clara University, English Department   *1985 - 1986*
    Taught undergraduate classes in composition and rhetoric.

Adjunct Lecturer , IUPU – Fort Wayne, English Department   *1983 - 1984*
    Taught undergraduate classes in composition and rhetoric.

Associate Instructor , Ball State University, English Department   *1982 - 1983*
    Taught undergraduate classes in composition and rhetoric.

Professional Experience

Invited Participant , NSF/AECT Early Career Symposium   *2010*
    One of nine doctoral students selected to participate in a 1 ½ day
    symposium for scholars who are early in their careers.

Graduate Assistant: Distance Education Research, Indiana University, IST   *2009 - 2010*
Department
    Assisted Prof. Elizabeth Boling in recommending to the Dean a strategy
    for supporting and enhancing distance education in the School of
    Education. We conducted a needs analysis of current activities and
    future plans, researched best practices at other institutions, and wrote a
    white paper describing our findings and recommendations.

Graduate Assistant: Distance Education Webmaster, Indiana University, IST   *2009 - 2010*
Department
    Maintained the IST department's website and supported faculty in their
    use of technology for online classes.

Research Assistant, Indiana University, IST Department   *2007 - 2008*
    Developed a Web-database application as part of a software project
    (MAPSAT) for Dr. Ted Frick using PHP, MySQL, HTML, CSS, XML,
    Javascript, and Actionscript (Flash).

Research Analyst, Santa Clara University, Institutional Research   *2001 - 2006*
    Developed and managed a data mart to support strategic planning for   *1988 - 1997*
    the university using Cognos DecisionStream for ETL (extract,
    transform, load) of data from PeopleSoft Student Administration
    (Informix database) into a conformed data mart (MS SQL Server) for
    reporting student, faculty, and course data.
    Developed and managed Web and desktop applications to support data
    analysis/reporting and work processes using Cold Fusion, MS SQL
    Server, MS Access, and MS Excel.
    Designed and developed reports of institutional data to meet federal
    reporting requirements (e.g. IPEDS) and accreditation requirements
    (WASC) and to support strategic planning and reporting to external
    agencies and college guide publishers.

Senior Design Technologist, Atomic Tangerine   *2000*
    Conducted needs analysis and designed information architecture for
    clients' websites.
    Worked as intermediary between designers and programmers.
    Wrote HTML, XML, CSS, Javascript, and Actionscript (Flash) for
    websites.

Animator, Blue Mountain Arts   *1999 - 2000*
    Designed original animated greeting cards (GIFs and Flash).

University Webmaster, Santa Clara University, Media Services   *1997 - 1999*
    Led the redesign of the University's website, which included assessing
    the needs of key stakeholders, designing the site's information
    architecture, and managing a small staff of Web designers/developers.
    Developed and managed Web-database applications using Cold Fusion
    and MS SQL Server.
    Founded and managed *Codesign*, a group of student
    designers/developers who provided Web services for campus
    organizations.
    Trained over 100 faculty, staff, and students to design, develop, and
    maintain websites.

Writer, Interactive Network   *1993 - 1994*
    Wrote and edited a variety of content, including film reviews and "just-
    in-time" information for simulcast with the 1994 Winter Olympics.

Independent Filmmaker, Cine22 Productions   *1988 - 1993*
    Wrote, co-produced, and co-directed several short films, including game
    (18 min., 16mm color) which won several awards and played in a dozen
    film festivals around the world. In addition, two of my screenplays were
    developed into award-winning short films by other filmmakers.

Publications

Reigeluth, C. M., & Myers, R. D. (2012). *Serious game design report.* Manuscript
prepared for the U. S. Air Force Research Laboratory.

Enfield, J., Myers, R. D., Lara, M., & Frick, T. W. (2012). Innovation
diffusion: Assessment of strategies within the DIFFUSION
SIMULATION GAME. *Simulation & Gaming, 43*(2), 188-214. doi:
10.1177/1046878111408024

Howard, C. D., & Myers, R. (2010). Creating video annotated discussions:
An asynchronous alternative. *International Journal of Designs for Learning, 1*(1).
http://scholarworks.iu.edu/journals/index.php/ijdl/index

Lara, M., Myers, R., Frick, T., Karabacak, S., & Michaelidou, T. (2010). A
design case: Creating an enhanced version of the *Diffusion Simulation Game.*
International Journal of Designs for Learning, 1(1).
http://scholarworks.iu.edu/journals/index.php/ijdl/index

Lara, M., Enfield, J., Myers, R., & Frick, T. (2010). The *Diffusion Simulation
Game:* Does learning occur across multiple game plays? Manuscript
submitted for publication.

Myers, R., & Frick, T. (2009). *SimTIE-Math*: A simulation-game on
technology integration for mathematics learning. *Proceedings of the 32nd annual
convention of the Association for Educational Communications & Technology.*

Lara, M., Myers, R., Frick, T., Karabacak, S., & Michaelidou, T. (2009). A design case: Creating an enhanced version of the *Diffusion Simulation Game*. *Proceedings of the 32ⁿᵈ annual convention of the Association for Educational Communications & Technology.*

Frick, T., Howard, C., Barrett, A., Enfield, J., & Myers, R. (2009). Alternative research methods: MAPSAT your data to prevent aggregation aggravation. *Proceedings of the 32ⁿᵈ annual convention of the Association for Educational Communications & Technology.*

Presentations at Conferences

Myers, R. D. (2012). *Analyzing patterns of gameplay data to improve the design of a simulation game.* Paper presented at the annual meeting of the American Educational Research Association, Vancouver, BC, Canada.

Myers, R. (2011). *Analyzing interaction patterns to verify a simulation/game model.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Jacksonville, FL.

Myers, R., & Modell, M. G. (2011). *Gamifying the conference experience: A design case.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Jacksonville, FL.

Frick, T., Barrett, A., Howard, C., & Myers, R. (2011). *Applications of MAPSAT in educational research.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Jacksonville, FL.

Howard, C. D., & Myers, R. (2011). *Creating video-annotated discussions: An asynchronous alternative.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Lara, M., Enfield, J., Myers, R., & Frick, T. (2011). *Learning assessment in the Diffusion Simulation Game.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Myers, R., & Modell, M. G. (2011). *Gamifying the conference experience: A design case.* Paper presented at the annual IST Conference, Bloomington, IN.

Myers, R., Enfield, J., Lara, M., & Frick, T. (2010). *Analyzing gameplay to improve the design of a simulation game.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Anaheim, CA.

Lara, M., Enfield, J., Myers, R., & Frick, T. (2010). *Assessing learning in online simulations: A case study of the Diffusion Simulation Game.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Anaheim, CA.

Frick, T., & Myers, R. (2010). *Map and analyze patterns and structures across time (MAPSAT): Theoretical foundations for observing systemic change.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Anaheim, CA.

Modell, M. G., & Myers, R. (2010). *Exploring the role of identity in the design of a serious game.* Poster presented at the bi-annual Meaningful Play 2010 conference, East Lansing, MI.

Enfield, J., & Myers, R. (2010). *Assessment of strategies within the Diffusion Simulation Game.* Paper presented at the annual Games+Learning+Society conference, Madison, WI.

Kwon, S., Lara, M., Modell, M., & Myers, R. (February, 2010). *Using digital games and simulations for learning.* Panel discussion conducted at the annual IST Conference, Bloomington, IN.

Myers, R., & Frick, T. (2009). *SimTIE-Math: A simulation-game on technology integration for mathematics learning.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Louisville, KY.

Frick, T., Howard, C., Barrett, A., Enfield, J., & Myers, R. (2009). *Alternative research methods: MAPSAT your data to prevent aggregation aggravation.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Louisville, KY.

Lara, M., Myers, R., Frick, T., Karabacak, S., & Michaelidou, T. (2009). *A design case: Creating an enhanced version of the Diffusion Simulation Game.* Paper presented at the annual convention of the Association for Educational Communications & Technology, Louisville, KY.

Frick, T., Lara, M., Myers, R., Michaelidou, T., Karabacak, S., Enfield, J., & Kwon, S. (2009). *The Diffusion Simulation Game: Development of an enhanced version.* Paper presented at the annual IST Colloquium, Bloomington, IN.

Frick, T., Lara, M., Myers, R., Michaelidou, T., & Karabacak, S. (2009). *The Diffusion Simulation Game: Design and development of an enhanced online version.* Paper presented at the annual IST Conference, Bloomington, IN.

Frick, T., Howard, C., Barrett, A., Enfield, J., & Myers, R. (2009). *MAPSAT your research data to prevent aggregation aggravation.* Paper presented at the annual IST Conference, Bloomington, IN.

Frick, T., Myers, R., Thompson, K. & York, S. (2008). *New ways to measure systemic change: Map & Analyze Patterns & Structures Across Time (MAPSAT).* Featured research paper presented at the annual conference of the Association for Educational Communications & Technology, Orlando, FL.

Myers, R. (2008). *Simulating Personalized Learning in STEM.* Poster presentation that was awarded first prize for the Game+Learning+Society Conference, Madison, WI.

Frick, T., Myers, R., Thompson, K. & York, S. (2008). *New ways to measure systemic change: Map & Analyze Patterns & Structures Across Time (MAPSAT).* Paper presented at the Second Annual Summer Research Symposium, Association for Educational Communications & Technology, Bloomington, IN.

Myers, R. (2008). *SimTIE: A Simulation Game on Technology Integration in Education.* Paper presented at the annual IST Conference, Bloomington, IN.

Myers, R. (2005). *Building a reporting metadata explorer.* Technical presentation at the PeopleSoft Higher Education User Group Conference, Las Vegas, NV.

Service Activities

Game Host, IST Conference 2011 — *Feb. 2011*
   Designed, developed, and implemented The Dossier Game, a face-to-face multiplayer game which had as its primary goal the increase of participation and interaction of conference attendees.

Editorial Assistant, International Journal of Designs for Learning — *January 2010 – Present*
   Worked closely with the founding editor and editorial board to launch an online, design-oriented education journal; analyzed needs and explored digital publishing options; designed/created prototype journal.

Session Facilitator, AECT Conference 2010 — *Oct. 2010*
   Hosted conference sessions, which included greeting participants, introducing presenters, moderating discussion, and distributing and collecting session surveys.

Game Host, IST Conference 2010 — *Feb. 2010*
   Designed, developed, and implemented The Dossier Game, a face-to-face multiplayer game which had as its primary goal the increase of participation and interaction of conference attendees.

Volunteer, IST Colloquium 2009 — *July 2009*
   Recorded video of presentations to be made into podcasts.

Treasurer, Graduates in Instructional Systems Technology — *Jan. 2008 – June 2010*
   Managed the financial affairs of GIST.

Organizer, IST Conference 2009 — *Sep. 2008 – Mar. 2009*
   Planned the IST department's annual conference, which included managing over a dozen volunteers; recruiting a keynote speaker; arranging the catering of breakfast, lunch, and an evening reception; designing and developing an online submissions review system; managing funds; and documenting the entire process for future organizers.

Portfolio Reviewer, San Jose State University
    Served as an external reviewer of portfolios created by graduate students
    in the Instructional Technology program.
<div align="right"><i>Oct. 2008</i></div>

Session Facilitator, AECT Conference 2008
    Hosted conference sessions, which included greeting participants,
    introducing presenters, moderating discussion, and distributing and
    collecting session surveys.
<div align="right"><i>Nov. 2008</i></div>

Volunteer, IST Conference 2008
    Served at the conference registration desk, hosted sessions, and
    performed other tasks as needed.
<div align="right"><i>Feb. 2008</i></div>

Professional Affiliations

Member, Association for Educational Communications & Technology   *2006 - present*

Member, American Educational Research Association   *2010 - present*

Member, North American Simulation & Gaming Association   *2010 - present*