Harold F. O'Neil • Ray S. Perez

Computer Games

and Team and Individual Learning

COMPUTER GAMES AND TEAM AND INDIVIDUAL LEARNING

This page intentionally left blank

COMPUTER GAMES AND TEAM AND INDIVIDUAL LEARNING

Edited by

HAROLD F. O'NEIL
University of Southern California/CRESST

and

RAY S. PEREZ
Office of Naval Research

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo

Elsevier
Linacre House, Jordan Hill, Oxford OX2 8DP, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

First edition 2008

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN: 978-0-08-045343-9

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands
08 09 10 11 12 10 9 8 7 6 5 4 3 2 1

To our wives
Eva L. Baker and Dorothy A. Lange
Who by their support have made this book a reality

This page intentionally left blank

CONTENTS

PREFACE[1]

The goal of *Computer Games and Team and Individual Learning* is to report progress on the development of theoretical frameworks for games and to report on the empirical results of qualitative or quantitative approaches to measuring the impact of computer games on the learning of adults. Learning in both individual and team settings is the context of the book. Both civilian sector and military applications are also presented.

Educators and trainers began to take notice of the power and potential of computer games for education and training back in the 1970s and 1980s. Computer games were hypothesized to be potentially useful for instructional purposes and were also hypothesized to provide multiple benefits: (a) complex and diverse approaches to learning processes and outcomes; (b) interactivity; (c) ability to address cognitive as well as affective learning issues; and (d) perhaps most importantly, motivation for learning.

While effectiveness of game environments can be documented in terms of intensity and longevity of engagement (participants voting with their quarters or time), as well as the commercial success of the games, there is much less solid empirical information about what learning outcomes are systematically achieved by the use of individual and multiplayer games to train adult participants in acquiring knowledge and skills. Further, there is almost no guidance for game designers and developers on how to design games that facilitate learning. This book addresses these issues.

There are multiple ways of assessing learning. For example, one could specify the assessment of the training effects of a game by examining trainees' ability to solve criterion problems, their application of declarative and procedural knowledge, their willingness to raise or lower game challenge conditionally, their self-reports, and records of their play. Evaluation questions to be answered about the cognitive and affective effects of games should concern the four levels of Kirkpatrick's (1994) framework. The evaluation system should also include measures related to the attainment at different levels of expertise of the specific content and skill acquisition being trained. These may include skills to be learned along the way, as well as those of an intentional or unintentional outcome nature. There

[1] The work reported herein was supported in part by the Office of Naval Research, under Award No. N00014-04-1-0209, Award No. N00014-02-1-0179, and Award No. N00014-06-1-0711, and in part under the Educational Research and Development Centers Program, PR/Award Number R305A050004, as administered by the Institute of Education Sciences, US Department of Education. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Office of Naval Research, or the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the US Department of Education.

is also a skill related to personal development – the "learning to learn" or self-regulation outcome. To what degree do participants develop the strategic knowledge necessary to apply to the specific training topics? Do players develop more general predispositions and behaviors that support transfer of knowledge across different contexts or problem variations? What is the role of formative evaluation? These issues are addressed in this book.

Our position is that games themselves are not sufficient for learning, but there are elements in games that can be activated within an instructional context that may enhance the learning process. In other words, outcomes are affected by the instructional strategies employed. There is general consensus in our community that learning with interactive environments such as games and simulations is not effective when no effective instructional measure or support is added. For example, Kirschner, Sweller, and Clark (2006) provide a convincing argument that discovery, problem-based, experiential, and inquiry-based techniques do not work instructionally. Since such techniques are the sole instructional strategy in the vast majority of games, it would be expected that most games would not lead to learning in adults.

In summary, the book deals with (a) the primacy of learning as a focus for game technology, sometimes called serious games; (b) the need to integrate such game technology with instructional and assessment strategies; (c) the paucity of and need to support the development of game-based curriculum and tools; (d) the need to integrate assessment in game technology; (e) the need for theory-driven and evaluation studies to increase our knowledge and efficacy of games as tests; (f) the need for a psychometric approach to the use of game assessments; and (g) the need to match the skills that the game purports to develop to the skills, knowledge, and abilities that are needed to perform the criterion task. The authors are from government (both DoD and civilian sectors), industry, and academia. Game technology is characterized from a series of different theoretical and empirical viewpoints. The book is designed for professionals in the gaming, simulation, assessment and evaluation, educational technology, and educational psychology communities. It explores the state of the art in the use of computer game technology for teaching and measurement of learning in adults. The unique focus of this book is on the empirical impact, both qualitative and quantitative, of computer games on learning with adults.

This edited book is divided into four major parts. The first is the framework section that provides the reader background in a general conceptual framework for games, assessment, formative evaluation, and software support. This is followed by two sections on learning in teams and learning in individuals, in games. Finally summary and discussion of the chapters is given. The reader could also use the last chapter as a map through this edited book.

This book could not have come into existence without the help and encouragement of many people. Our thanks to our Elsevier editor, Mr Ben Davie, for his support and guidance in the publication process. We thank Ms Joanne Michiuye and Ms Katharine Fry for their excellent assistance in preparing the manuscript.

—Harold F. O'Neil
Los Angeles

—Ray S. Perez
Arlington, Virginia

References

Kirkpatrick, D. L. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco, CA: Berrett-Koehler Publishers, Inc.

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*, 75-86.

This page intentionally left blank

CONTRIBUTORS

Amy L. Alexander
Human Performance Division
Aptima, Inc.
Woburn, MA
USA

Winfred Arthur, Jr.
Department of Psychology
Texas A&M University
College Station, TX
USA

Eva L. Baker
CRESST/UCLA
Los Angeles, CA
USA

William M. Bart
Department of Educational Psychology
University of Minnesota
Minneapolis, MN
USA

Daphne Bavelier
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY
USA

James Belanich
Research & Advanced Concepts Office
U.S. Army Research Institute for the Behavioral and Social Sciences
Arlington, VA
USA

Tad T. Brunyé
Advanced Training Solutions Division
Aptima, Inc.
Woburn, MA
USA

Hsin-Hui Claire Chen
University of Southern California
Los Angeles, CA
USA

Julia E. Cohen
Institute of Child Development
University of Minnesota
Minneapolis, MN
USA

Michael T. Curtis
Team Performance Laboratory
Department of Psychology
Institute for Simulation & Training (IST)
University of Central Florida
Orlando, FL
USA

Girlie C. Delacruz
CRESST/UCLA
Los Angeles, CA
USA

Deborah DiazGranados
Department of Psychology
Institute for Simulation & Training (IST)
University of Central Florida
Orlando, FL
USA

William Ferguson
BBN Technologies
Cambridge, MA
USA

C. Shawn Green
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY
USA

Michelle Harper-Sciarini
Team Performance Laboratory
Department of Psychology
Institute for Simulation & Training (IST)
University of Central Florida
Orlando, FL
USA

Saahoon Hong
University of Minnesota
Minneapolis, MN
USA

Talib S. Hussain
BBN Technologies
Cambridge, MA
USA

Florian Jentsch
Team Performance Laboratory
Department of Psychology
Institute for Simulation & Training (IST)
University of Central Florida
Orlando, FL
USA

Robert L. Jorczak
Department of Educational Psychology
University of Minnesota
Minneapolis, MN
USA

Adrienne Y. Lee
Department of Psychology
New Mexico State University
Las Cruces, NM
USA

Jacquelyn Ford Morie
The Institute for Creative Technologies
University of Southern California
Marina Del Rey, CA
USA

Laura N. Mullin
Nuance Communications, Inc.
Burlington, MA
USA

Allen Munro
Center for Cognitive Technology
Rossier School of Education
University of Southern California
Redondo Beach, CA
USA

Harold F. O'Neil
University of Southern California
Los Angeles, CA
USA

Karin A. Orvis
Department of Psychology
Old Dominion University
Fairfax, VA
USA

Kara L. Orvis
Organizational Effectiveness Division
Aptima, Inc.
Woburn, MA
USA

Aaron M. Perez
Department of Psychology
New Mexico State University
Las Cruces, NM
USA

Ray S. Perez
Office of Naval Research
Arlington, VA
USA

Eduardo Salas
Department of Psychology
Institute for Simulation & Training (IST)

University of Central Florida
Orlando, FL
USA

Chun-Yi Shen
Tamkang University
Taipei County, Taiwan

Tae Seob Shin
Michigan State University
East Lansing, MI
USA

Jason Sidman
Advanced Training Solutions Division
Aptima, Inc.
Woburn, MA
USA

Robert D. Tennyson
Department of Educational Psychology
University of Minnesota
Minneapolis, MN
USA

Rebecca Tortell
The Institute for Creative Technologies
University of Southern California
Marina Del Rey, CA
USA

Richard A. Volz
Department of Computer Science
Texas A&M University
College Station, TX
USA

Richard Wainess
CRESST/UCLA
Los Angeles, CA
USA

Shawn A. Weil
Advanced Training Solutions Division
Aptima, Inc.
Woburn, MA
USA

Jonathan H. Whetzel
Sandia National Laboratories
Albuquerque, NM
USA

Josh Williams
The Institute for Creative Technologies
University of Southern California
Marina Del Rey, CA
USA

PART 1

FRAMEWORK

This page intentionally left blank

A CONCEPTUAL FRAMEWORK FOR THE EMPIRICAL STUDY OF INSTRUCTIONAL GAMES

Robert D. Tennyson and Robert L. Jorczak
University of Minnesota

Abstract

Educational researchers want to identify the characteristics of games that influence player motivation and learning to design effective instructional games. Simulation games offer the highest potential for learning and research, as such games use realistic contexts, carefully structured problem difficulty, and learning aids to motivate players and support problem solving. Game characteristics should be investigated in reference to models of cognition and learning, such as the Interactive Cognitive Complexity model. We propose a conceptual framework that includes virtual context, problem specification, interaction and control characteristics, learning support, and social interaction as key categories of independent variables for instructional game studies.

1. Why Study Games for Educational Purposes?

Video games are played using a keyboard or controller and present visual output on a video screen. Playing and mastering entertaining video games involves much learning including enhanced motor skills, increased declarative and procedural knowledge, and improved problem solving abilities. When students address similar learning goals by doing traditional learning activities, they rarely demonstrate the high level of effort and motivation observed of entertainment game players. Video games used for entertainment seem inherently motivating to game players. In contrast, traditional classroom activities and assignments often require additional motivators, such as grades, to maintain student interest and effort.

The high level of player effort displayed in learning and mastering video games has led educators and learning researchers to hypothesize that the techniques of such games can

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

be used to increase student motivation in learning traditional academic outcomes (Garris, Alhers, & Driskell, 2002; Gee, 2005; Habgood, Ainsworth, & Benford, 2005; Prensky, 2001). Educators are also interested in learning via games because of the instructional trend toward learner-centered education wherein learners have increased involvement in learning activities, and are given more control over learning goals and resources (Garris et al., 2002; Kafai, 2001).

Designers of learning activities hope to improve academic learning by merging the content of education (or training) with the format and techniques of entertainment games, thereby designing instructional games. To help instructional designers create games that are motivating and enable players to achieve desired learning outcomes, researchers must identify the characteristics of entertainment games that influence player motivation and promote learning. In addition, researchers are interested in how instructional games can be used to promote higher order learning outcomes such as problem solving (Kafai, 2001).

2. Research Needs for Instructional Game Design and Use

While the fact that entertainment games are highly motivating and are often successful at helping players achieve game mastery is clear, just what makes them so, and whether or not those factors can be applied to academic learning, is not clear. Educators would like to see students apply a similar amount of effort to learning school subjects as they see applied to learning entertainment games, but they are not sure how to use games to achieve high learner motivation and also achieve meaningful learning. The current situation is summarized by Garris et al. (2002, p. 442): "Unfortunately, there is little consensus on game features that support learning, the process by which games engage learners, or the types of learning outcomes that can be achieved through game play." Educators and instructional game designers need the input and direction provided by empirical studies to make good choices in applying the lessons of video games to learning activities.

Instructional games are not new, but research into the effectiveness of gaming has been limited. Some empirical evidence suggests that games can efficiently promote learning (Cordova & Lepper, 1996; Henderson, Klemes, & Eshet, 2000; Moreno & Mayer, 2005; Ricci, Salas, & Cannon-Bowers, 1996), but research has been unfocused in regard to how games can promote learning (Dempsey, Lucassen, Haynes, & Casey, 1996; Habgood et al., 2005; Kafai, 2001; Moreno & Mayer, 2005). O'Neil and Fisher (2004) and O'Neil, Wainess, and Baker (2005) searched the research literature for empirical evidence of the training effectiveness of games. They found studies providing empirical evidence of increased motivation, enhancement of cognitive processes, and improved learning outcomes. The number of empirical studies found, however, was very small. More studies and more focused studies are clearly needed.

3. Instructional Games as Simulations

Many of the concepts associated with games do not have precise or even generally accepted definitions. The concept of a game itself is not well-specified and several competing definitions are found in the literature of instructional games (Dempsey et al., 1996; Garris et al., 2002; Mitchel, & Savill-Smith, 2004). Games are often associated with play, which can be defined as a pleasing activity disassociated from direct consequences in the real world (Fabricatore, 2000).

Instructional games have specific learning outcomes as primary goals. Garris et al. (2002) proposed an instructional game model that has learning outcomes as the goal of an interaction cycle between player and game. Their model includes specification of a game context that includes both game characteristics and instructional content. How well the game characteristics support instructional objectives is a key variable of instructional game effectiveness, though hard to quantify.

Garris et al. (2002) distinguish games from simulations in that simulations represent real-world systems and games do not. A more specific distinction is that games often have artificial consequences (such as amassing points) that are not realistic if used to simulate real systems. We find distinctions between simulation-like games and simulations (e.g., O'Neil et al., 2005) of little value in determining how games promote learning and prefer to view games and simulations as ends of continuum of contextual realism. Both present a dynamic virtual system, the state of which is continuously altered toward a desired goal state by the user. Instructional simulation games include targeted learning outcomes a goal. We propose that simulation games offer the highest potential as instructional games. Simulation games also provide the most comprehensive and flexible means to study instructional game characteristics. Our framework of instructional game variables, therefore, is intended primarily for simulation-type instructional games.

4. The Need for a Conceptual Framework in Studying Instructional Games

A conceptual framework that identifies the key variables of instructional games affecting player motivation and learning, and also relates those variables to learning theory, can help focus research on factors leading to the effective design and use of instructional games. The value of such a conceptual framework lies in how well it suggests studies and raises questions of interest to theorists and researchers. A framework can best suggest direction to researchers if it identifies potential independent variables for study and places such variables in a theoretical context understood by theorists and researchers. A useful conceptual framework identifies, categorizes, and links game variables likely to affect learning to a learning model specified by a learning theory.

The Interactive Cognitive Complexity (ICC) learning model (Tennyson & Breuer, 1997, 2002) is an integrative information-processing learning model that views learning as the result of complex and non-linear interactions of variables internal and external to the cognitive system of a learner. According to the ICC model, the components of

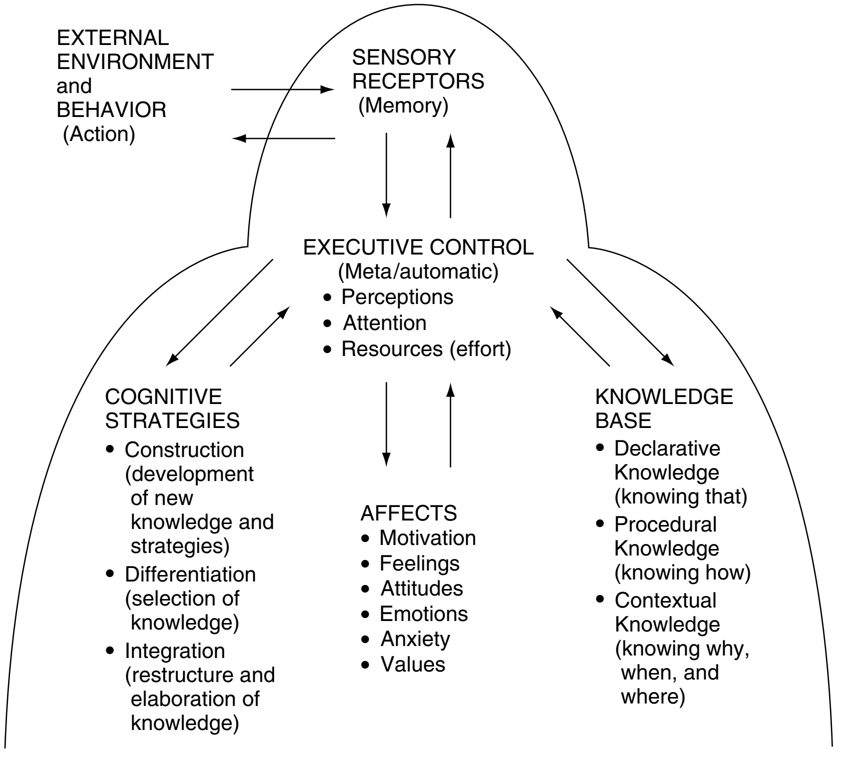

Figure 1. ICC model.

a learner's cognitive system include learner affect, cognitive strategies, and a knowledge base; in addition to executive control and internal processing components (see Figure 1).

The knowledge base of the ICC model is the repository for all previously acquired declarative, procedural, and contextual knowledge. Declarative knowledge includes concepts, rules, and principles, i.e., knowing that. Procedural knowledge is the use of declarative knowledge to accomplish tasks and solve problems, i.e., knowing how. Contextual knowledge is knowing why, when, and where to use declarative and procedural knowledge. Tennyson and Breuer (1997) see the structure of the knowledge base as complex networks of concepts and propositions (domains) that are organized into meaningful associations (schemata).

Cognitive strategies are processes that learners apply in elaborating and altering their knowledge base and include differentiation, integration, and construction (Tennyson & Breuer, 1997). Differentiation is the ability to apply appropriate contextual criteria to selectively retrieve knowledge from the knowledge base. Integration is the ability to elaborate or restructure existing knowledge. Construction is the ability to create new knowledge in novel situations. Internal processing components use these cognitive strategies on variables input from the environment and also on internal variables such as current emotional states and existing knowledge to create new (or altered) knowledge.

Garris et al. (2002) include skill-based (i.e., motor skills) and affective outcomes as potential learning outcomes in addition to the declarative, procedural, and contextual ("strategic") outcomes. The ICC model finds a separation of affect and cognitive domains "somewhat arbitrary" and includes affect as directly influencing cognitive processing (Tennyson & Breuer, 1997). The affect component of the ICC model includes motivational variables in addition to attitudes, emotions, and self-efficacy.

All components of the ICC model interact with each other and with sensory information from an external source, such as an instructional game, to elaborate the individual's knowledge base (Tennyson & Breuer, 1997). The process is iterative as information enters the cognitive system from the senses continually, affecting the system components in combination with internally stored information (from the knowledge base) and various internal state variables (of, for example, the affective or executive control systems). New states and knowledge result from the operation of cognitive strategies on new and stored information. The cognitive system continually cycles through this process of updating internal states and information based on external and internal information. Because the ICC model posits that affective and cognitive variables are essential parts of the cognitive system, input that includes both affective and cognitive information may be more successful in increasing motivation and learning. Simulation games excel at engaging both affective and cognitive systems of players.

The interaction cycle of a simulation complements the internal cycle of the cognitive system (i.e., iterative update of stored information based on environmental input and the state of internal variables). The output of a simulation is the input to the cognitive system and the output of the cognitive system (behavior) provides input to the simulation. The state of the simulation is altered toward a desired goal, and the internal state of the user (e.g., learning) is similarly altered by interacting with the simulation. This compatibility is not surprising as simulations are meant to mimic real environments that the human cognitive system has evolved to deal with. The high motivation associated with simulation-type games may, in part, be due to the fact that the human cognitive systems is most engaged when dealing with complex processes that mimic real-world situations.

Consistent with the ICC model is the cognitive model of multimedia learning of Moreno and Mayer (2005), which posits that meaningful learning occurs when learners construct a coherent knowledge representation based on both newly acquired information and existing knowledge (internal and external variables in the ICC model). Learners use cognitive processes for selecting relevant information in addition to organizing and integrating that information into a structure coherent to the learner. Moreno and Mayer suggest that providing guidance (e.g., help in the selection of new information or in activating relevant existing information) can enhance learning. In addition, they suggest that the instructional treatments of interaction and reflection can aid the cognitive processes of organizing and integrating (Moreno & Mayer, 2005).

Simulation games promote learning in regard to the Moreno and Mayer (2005) model by posing problems and tasks to which the user must select relevant information to succeed. Games also aid in the organization and integration of information by using highly interactive environments. In addition, games have clear consequences for failure (feedback) that help the player in subsequent attempts to solve a problem.

Tennyson and Breuer (2002) present complex-dynamic simulations (simulations employing a complex virtual reality) as an instructional method that can significantly improve higher-order thinking strategies by exercising the cognitive strategies defined by the ICC model. Complex-dynamic simulations are beneficial to development and improvement of higher-order thinking strategies (problem solving and creativity) because they require students to use their integration and construction cognitive strategies to solve problems. Tennyson and Breuer (2002) list various examples of complex-dynamic simulations used for learning that can also be described as instructional simulation games (e.g., a microeconomic system of a company that is to be run by a manager – the game player).

Our focus is on the creation of a framework that suggests variables of simulation games that can potentially affect player motivation and higher-order learning. The framework variables are variables of the games themselves (i.e., external variables); but these variables affect the function and content of the cognitive system of the game player. Individual player differences are found within the variables of the cognitive system, such as different information in the knowledge base, differences in self-esteem, etc. The value of those variables can and does affect motivation and learning – these are the internal variables of the ICC model.

5. Characteristics of Entertainment Games Related to Learning

Simulation games used for entertainment seem to be motivating because they let humans do in a virtual world what they are designed to do in the real world – interact with a rich environment to achieve a goal. This characterization suggests that "rich environment" and "interaction" are concepts that need further examination. A rich environment can be achieved without interaction when readers become cognitively and emotionally immersed in an imaginative and detailed story. Activity and interaction without a rich immersive environment can also be achieved as when students drill and practice activities out of a meaningful context. A rich environment and interaction can each be motivating independently, but the motivation caused by their combination may be much more than the sum of their independent affects on motivation.

Another characteristic of instructional interest is the way entertainment games enable players to improve their performance and move toward goal achievement in small steps starting from low competence and knowledge. Instructional strategies also seek to gradually move novices to mastery. How do entertainment games achieve this gradual progression toward mastery? The structure of the progression of game play from easy to harder tasks is one factor, but games also include components specifically designed to aid and guide players (i.e., explicit learning support components).

A rich virtual environment (complex and elaborated context), high interactivity within that environment, a carefully constructed progression of challenges in obtaining a goal, and learning support components are factors of interest in producing a framework of instructional game characteristics that affect learning. Questions about the motivational and instructional characteristics of games can best be answered by game research that uses a conceptual framework that links game variables to cognitive components of learners and categorizes game variables based on the characteristics of entertainment games. In creating such a framework, we propose five conceptual categories of instructionally relevant game variables: (1) virtual context, (2) problem specification, (3) interaction and control, (4) learning support, and (5) social interaction.

5.1. Virtual Context

Most entertainment games are played in highly contextualized virtual environments comprising the basic story and virtual setting of the game. While the virtual contexts of entertainment games are not always realistic, in the sense of simulating real world as opposed to imaginary situations, game contexts are internally consistent and logically elaborated to make the contexts seem authentic, even for imaginary settings. Game players acquire skills that seem very authentic because representations and events of the game, plus the skills, tasks, and consequences required to achieve the game goals are consistent with the context of the game (i.e., game play is meaningful). Such consistency makes game play seem authentic. Authentic tasks are thought to positively affect learning (Jonassen, Howland, Moore, & Marra, 1999; Ormrod, 2004). Learning theorists recognize the importance of context in learning, a recognition associated with theoretical perspectives of situated learning and supported by the ICC model in that learners respond to authentic tasks by using learner-specific information in the knowledge base and engaging cognitive, affective, and volitional processes.

Context type. The specification of a real world (e.g., WWII combat) or imaginary (e.g., interstellar combat) context is an aspect of game design that can affect player interest in the game. Instructional games are successful in motivating players by offering virtual contexts that are motivating to the target audience (Kirriemuir, 2003). The Becta (2001, p. 7) study of computer games in education makes an explicit connection between context and learner interest: "The over-riding appeal of the games in lessons was the way in which learning opportunities and skills were presented in the context of a situation attractive to young people."

So what contexts are attractive to students? It is likely that some contexts are inherently interesting to most students in a specific age group but the appeal of a context may be related to learner-specific preferences and needs. Gender differences have also been found regarding game context (e.g., Carr, 2005; Malone & Lepper, 1987). The ICC model suggests that aspects of learner affect (e.g., motivation) are influenced by how game content matches the state of the learner's cognitive system, for example, by matching information already present in a learner's knowledgebase. Learner-specific variables, such as existing declarative knowledge, may influence a learner's interest in specific contexts. For example, a learner may respond positively to a game situated in ancient Rome if that learner knows much about ancient Roman history. Individual differences in affective variables such as self-efficacy, self-image, etc., also influence the appeal of game contexts according to the ICC model.

Intrinsic and Extrinsic Fantasy. Game context has previously been investigated in terms of intrinsic and extrinsic fantasy. Malone (1981) studied the connection between game fantasy (similar to game context, but constructed in the mind of the player) and motivation, based on whether the fantasy suggested by the game is intrinsic (connected to instructional goals of the game). Malone and Lepper (1987) studied student use of simple computer games in which the fantasy component was altered or eliminated and found that presence of fantasy can affect motivation. Habgood et al. (2005) suggest that game fantasy is important in initially engaging a player's interest, but that other game characteristics, such as game operations and interactions, are more important for continuing motivation during game play.

Point of View. Players interact with entertainment simulation games from various points of view. In first-person games the player literally sees through the eyes of the character in the game. With other points of view, the player controls a character they can see from various perspectives in the virtual space (e.g., from behind the game character). In addition to physical points of view, role-playing games have functional and attitudinal perspectives that are assumed by players. In assuming attitudes different from the players' own attitudes, new perspectives are opened to the player, enabling learning that might be otherwise difficult or impossible (Gee, 2005). The influence of point of view on learner motivation and learning is another potential area of investigation by instructional game researchers.

Media Modality, Fidelity, and Mood. The use of multiple media modalities (e.g., audio and video) to present a virtual context seems to increase both interest and attention of game players. Another variable suspected of increasing positive feelings about a game is the fidelity of the medium used. Entertainment games impress players with visual and aural representations of the virtual context that are detailed and realistic. Such games use increasingly realistic graphics and animations to add to the realism and interest of the virtual context. The degree to which game media seem real depends on the subjective perception of the player, but quantitative parameters such as maximum resolution of graphics and animation frame rates (i.e., media fidelity) are assumed by game designers to be related positive attitudes about a game context.

Qualitative factors, such as the artistic style of graphics, the graphic design of scenes, and genre of music, may contribute to increased affective qualities that players associate with the game. These qualitative characteristics, can, for example, establish a mood that promotes the targeted emotions of the game (e.g., creepy) which likely heightens player involvement.

Dynamic Media. Entertainment games use multiple media to great effect but emphasize animated visual stimuli, often by presenting a graphically realistic and highly dynamic three-dimensional virtual environment. Thus, the entertainment gaming medium is very different than the static and two-dimensional graphics of text-based media used in many educational activities.

Habgood et al. (2005) suggest that research on how to represent information for learning identifies two key ways by which representations can support learning: (a) by using representations that make key domain features explicit and (b) by using dynamic and interactive features. Game contexts which include representations that are related to the targeted learning domain, particularly those that serve to clarify or provide metaphors, may be beneficial to learning. Animated graphics increase player motivation and are preferred by game players (Rieber, 1991). Highly dynamic three-dimensional visual representations and sound effects may be inherently more interesting because they increase the realism which engages more cognitive resources of the player making games more experiential. Increased use of multiple dynamic media is a potential variable for examination by researchers studying the effect of instructional games on learner affect and cognition. Malone and Lepper (1987), for example, found that students preferred games with audio effects. Increasing media modalities and dynamism also engage more cognitive systems and resources which may influence the amount and depth of learning.

How context variables affect the cognitive system components of players are key questions of the study of instructional games. The content of game narrative, the point of view of the player, the fidelity of context representations, and the dynamism of representations are all potential variables of context that can contribute to motivation and are therefore targets of investigation by researchers in relation learning.

5.2. Problem Specification

Popular entertainment games do not depend only on interesting and realistic contexts for their appeal. Games pose challenging problems for players to solve via their virtual actions. Many entertainment games are essentially a series of problem-solving challenges in which problems are solved with knowledge and skills acquired by playing the game and manipulating the virtual environment. A problem, in this context, is loosely defined ranges from virtual physical situations that require motor skills (e.g., avoiding being killed by a game agent) to complex puzzles requiring a series of actions or responses by the player.

Learning theorists view problem solving as relatively high-level learning that has the potential to engage learners in the construction of solutions using their cognitive strategies and knowledge base (Tennyson & Breuer, 1997). Problem-based learning activities let learners direct and control actions in dealing with a presented problem in a specific context. The characteristics of real-world problem-solving activities are the same factors observed in the virtual contexts and interactions of entertainment games. Problem solving may be inherently motivating to learners because it provides an environment in which learners can apply some or all of their cognitive strategies in achieving meaningful goals. In attempting to solve problems, a learner often must choose from their current knowledge, integrate new knowledge acquired about the problem, and construct a solution. Potentially, problem solving can employ all cognitive strategies noted in the ICC model.

From a learning theory perspective, the components of problems are known to include givens, goals, and operations (Ormrod, 2004). These components are also of primary concern to entertainment designers in specifying game problems. *Givens* include the context of the game in addition to the many virtual objects and situations of the game. Game *goals* are often the solutions to problems presented by the game or the achievement of a targeted end state. In many simulation-type entertainment games, problems are presented in hierarchical levels from the overall problem (the solution of which is the major goal of the game) to solutions of smaller-scale problems that prevent the player from proceeding toward the overall goal (and be forced to repeat the subordinate problem). Game *operations* include all the actions that a player can take, the objects or abilities players can acquire, and the results of actions taken. Game operations, therefore, include rules governing game interactions including what options are available for solving problems (i.e., the potential solutions to a problem and the consequences of correct or incorrect solutions).

All three components of problem specification (givens, goals, and operations) offer variables for study based on traditional research questions related to problem solving. For example, Malone and Lepper (1987) found that computer games with goals had the highest correlation with student game preference compared to other game characteristics. Malone and Lepper (1987) also distinguish games with fixed goals from those with emergent goals, open-ended games in which players define their own goals (e.g., The Sims).

Gee (2005) speculates that the motivation generated by a game is related to the ability of the game to provide pleasantly frustrating challenges (i.e., doable tasks with gradually increasing difficulty). Related to this variable is Fabricatore's (2000) contention that videogames necessarily include some type of opposition the player must struggle against in attaining game goals. He finds that "most of the products that crowd the educational games market cannot be defined as videogames at all since they lack the element of struggle, which deeply compromises the challenge that the player faces during the game-playing" (Fabricatore, 2000, p. 13). The assertion is that instructional game designers are reluctant to hinder learning tasks with active game opposition and therefore design games that are less challenging and fun. Lepper and Malone (1987) include challenge as one category affecting motivation in instructional games. They see challenge as the presentation of performance goals whose attainment is uncertain but likely to contribute to enhanced self-esteem of the players. Lepper and Malone (1987) also suggest that uncertainty in goal achievement is based on difficulty levels and the inclusion of a random element. These variables of challenge are used extensively in entertainment games and their affect upon learning games is a potential key area of study.

Problem *organization* is another variable of potential study. Gee (2005) recommends that learners should not be "set adrift" to solve complex problems, but rather presented with problems they are currently equipped to solve (with the domain knowledge in their knowledge base). Malone and Lepper (1987) note the importance of an optimal level of problem challenge in motivating game players and suggest that the level at which a player succeeds 50% of the time (the highest level of uncertainty) is the most motivating. Successful games are adept at presenting a sequence of increasingly challenging problems that are within the ability of the progressing player to solve. By this means, players are challenged to gradually move toward competence.

Other variables associated with problem solving include: the complexity of required problem-solving algorithms or heuristics, well-defined versus ill-defined problems, type of problem solving strategies (e.g., trial and error) required to solve game problems, and the type of cognitive resources involving problem solving (Ormrod, 2004). For example, a study could attempt to correlate motivation with the types of problems presented by an instructional game, such as those with solutions requiring motor skills, mental puzzles, or those requiring a specific sequence of actions. Games requiring a specific type of problem-solving strategy probably are best for teaching that type of problem solving, but other learning outcomes, such as acquisition of declarative domain knowledge, may be linked with the type of problem solving strategy required in a game.

Malone and Lepper (1987) suggest that the human cognitive system includes curiosity (similar to challenge but not involving self-esteem). One type of curiosity, sensory curiosity, is an attraction to unexpected sensory stimuli and the other, cognitive curiosity, attempts to achieve completeness, consistency, and parsimony in cognitive structures. Designers of games must consider the potential of game problems to stir the curiosity of the targeted audience of learners.

Learner affect, measured by reports of satisfaction, frustration, and perceived arousal and enjoyment, can be measured in regard to challenge, opposition, goal orientation, and difficulty levels of game problems. These variables are related to differences in affect and can be presumed to influence learning by the ICC model. Also, the retention and comprehension of declarative, procedural, and contextual information may differ with variations of instructional game problem specification.

5.3. Interaction and Control

Learner control is another category of variables that may affect players' motivation, cognition, and learning when using instructional games. Learner control variables include interactivity (e.g., choice of action) and personalization (e.g., choice of representation) and can extend to the selection or specification of game goals.

Interactivity. Interaction involves the transfer of information between the game and the cognitive system of the player. Players observe the virtual environment and input information or manipulate that environment to solve problems or meet challenges. Players receive consistent, realistic feedback for their actions. Designers of computer-delivered instruction have long believed that high levels of interaction in computer-assisted learning correlate with more and better learning (e.g., Northup, 2002). Some theorists believe that learning requires (or is equated with) activity itself (e.g., Jonassen, 2002). So, it is reasonable to expect that the highly interactive environment of simulation-type instructional games can have positive affects on learning. High levels of interaction involve players in the game (e.g., demand high levels of resource allocation, such as attention) and give players a sense of empowerment and control because the players' actions are seen to solve problems and resolve situations.

Learner control in simulation games often manifests as the ability of players to choose or control virtual activity (i.e., operations). In entertainment games, players are given much discretion in choosing virtual tools (e.g., weapons), initiating actions of player avatars (e.g., jumping climbing, picking up objects), selecting directions in which to move and explore, and other volitional actions. Game designers specify problems and potential solutions, but players determine how problems are solved. High interactivity combined with many choices results in players feeling empowered. Such empowerment can affect the learner's self-efficacy and self-esteem, factors known to affect learning (Bandura, 1993).

In their taxonomy of motivation in instructional games, Malone and Lepper (1987) state that the amount of control a learner has depends on: (a) the range of outcomes possible in an environment and (b) the extent to which an outcome is contingent on responses available to the learner. Control can be viewed as interaction in which reaction of the simulation (feedback) is contingent on learner actions (input). Malone and Lepper (1987) point out that the range of choice (e.g., number of choices) and the power of choice (the strength or extent of a choice) are two parameters of choice and therefore are potential variables of game interaction that may affect learning.

Control extends beyond a player's ability to choose actions within the game. *Personalization* involves players choosing non-operational attributes, such as the appearance of their avatar or a name used in the game. *Customization* involves the player's ability to change aspects of game contexts or operations. The ability to choose how a game looks or operates can give players a strong perception of control which may relate to player motivation. Some research indicates that increased personalization and customization results in better learning (e.g., Cordova & Lepper, 1996; Malone & Lepper, 1987).

Higher levels of learner control could include having the players set their own game goals. Constructivist learning theorists posit that learners learn more deeply when they can choose their own learning goals and learning activities (Jonassen et al., 1999). Gee (2005) calls this setting of goals and rules by game players as "co-design." Kafai (2001) identifies "games-to-learn" as games that are created or designed by students (as opposed to "games-that-teach" designed with fixed goals by others). Kafai sees game creation as more compatible with constructivist perspectives in letting students choose game goals and construct a means to address those goals via play. Letting the learner choose learning goals and activities enables that student to elaborate from his/her existing knowledge base or construct new knowledge by employing the integration and construction cognitive strategies.

Interactivity and learner control are game variables that do not lend themselves to easy quantification. But such variables play an important role in how instructional games can affect the cognitive strategies and the affect of the learner. Researchers need to examine how variations in perceived player control influences the affective component (e.g., challenge and frustration) of the player's cognitive system, and also influence the use of cognitive strategies (e.g., the types of interaction that better support integration and construction). Exposing students to variations of games similar in all respects except variables of interactivity and learner control is one way that instructional games can prove to be a valuable tool for learning research.

5.4. Learning Support

Implicit Support. Ideally, entertainment games support the learning of game operations implicitly by having players explore game context via the game's problem-solving interactions. The interactions can be designed to teach the game simply by playing it. Game tasks reveal how to play the game. Some entertainment games employ implicit learning by a sequenced and gradual increase in problem difficulty and by breaking down complex tasks into smaller component tasks and presenting those tasks in a logical sequence. An instructional game can also impart its instructional content implicitly through the problem-solving interactions of game play.

Explicit Support. Rarely, however, is a game so well crafted that satisfactory learning occurs only by playing the game. Most entertainment games also provide explicit learning support by means of components designed only to aid in learning the game. Such components include tutorials and reference information. Explicit learning support of computer-based games is accomplished in the same ways traditionally used to support software applications including user manuals, online help, and information presented as needed ("just in time"). With context-sensitive information, players are presented (or given access to) information when it is required within the sequence of game play. Game players need not read a user manual in an attempt to store all relevant information about game context, operations, and game scenarios if that information is available on demand in context. In using context-sensitive support, entertainment games avoid problems of cognitive overload in learning the rules of complex games. Also, presentation of information in a context helps make the information meaningful. "Human beings are quite poor at using verbal information (i.e., words) when given lots of it out of context and before they can see how it applies in actual situations. People use verbal information best when it is given 'just in time' (when they can put it to use) and 'on demand' (when they feel they need it)." (Gee, 2005, p. 7).

The use of simulation games for learning can be criticized as lacking in teaching function and teaching presence. These concerns may reflect a feeling that students cannot learn domain-specific declarative knowledge from games that do not include traditional didactic instructional strategies. Some educational researchers have raised questions about learning strategies that do not provide high levels of guidance (e.g., Kirschner, Sweller, & Clark, 2006, O'Neil et al., 2005). Explicit learning support components provide information in a format more like traditional direct instruction. Components that explicitly support learning goals via guidance and directive feedback may be essential for efficient learning via games.

The learning of game players is also supported explicitly by information and advice from the community of other players. Discussion forums and chat rooms are used to exchange information and relate game experiences. Often game players or makers publish "cheats" on the Internet that allow players to skip levels or have additional resources for playing the game. Such information enables players and encourages them to keep playing the game.

Other ways in which entertainment games explicitly support the learning of game operations is with game scenarios (sub contexts) that are specifically designed for learning game basics or practicing game skills. Gee (2005) labels these functions "fish tanks" (simplified game scenarios) and "sandboxes" (safer game scenarios). These practice scenarios help players learn game operations and practice the game with less complex problems and interactions.

Feedback. Feedback, in its basic form, is the change in state of a simulation that can be observed by the person interacting with that simulation. Simulation game designers choose what information is displayed to players. The type and amount of information displayed affects the player's understanding of how the simulation is responding to input. Feedback information that better enables a user to understand the operation or state of the simulation is said to make the simulation more "transparent" (Leutner, 1993).

Within instructional activities, feedback has an additional definition. Instructional feedback is information (often text) that is presented for a specific pedagogical purpose, usually informing the learner about the correctness of the learner's response. Feedback in this sense has many variables, it can be as simple as a single word (e.g., incorrect), a statement of the expected response (i.e., "answer c is correct"), or it can be more complex in explaining why a response is correct or not (sometimes called adaptive feedback). Leutner (1993) found that, under specific conditions, system-initiated adaptive (context sensitive) and on-demand non-adaptive learning support can improve learning of domain knowledge.

In their study investigating types of learner support in an instructional game, Moreno and Mayer (2005) examined the differential effect of simple feedback (telling the learner if their answer was correct or not) and more complex feedback that provided guidance to the correct answer. Guidance feedback included a discussion of why the player's response was incorrect based on a set of principles that were the main instructional content of the game. Moreno and Mayer (2005) found that guidance (explanatory feedback) resulted in higher transfer scores, fewer incorrect answers, and a greater reduction in misconceptions in learners.

Hsieh and O'Neil (2002) found that learners receiving feedback that indicates what is incorrect about a response (adaptive feedback) outperformed teams receiving only a correct/incorrect judgment. Chuang and O'Neil (2006) found that task-specific adaptive feedback, which indicates how a learner can improve their answer in addition to telling them what is wrong, results in additional performance gains.

Some theorists speculate that explicit teaching processes make games more instructionally effective. Some evidence exists that specific types of explanatory feedback improve performance, but more studies examining what types of instructional support are effective and what cognitive processes are affected are needed, especially in light of concerns about too much didactic content reducing the appeal and instructional effectiveness of games.

6. Social Interaction

Many learning theorists stress the value of social interaction in learning (e.g., Jonassen et al., 1999). By stressing multi-person or group play, games can be used to promote social skills as the targeted learning outcome. Group play can also be used as a motivating characteristic by playing with or against other humans.

Player preferences for game types and characteristics may be strongly influenced by social factors such as peer preferences and cultural influences. Carr (2005) found, via her observations of teenage girls' game playing habits, that a player's preferences for game types is partially determined by cultural and social aspects of the player's position in relation to gaming communities, peer group attitudes, and popular perceptions of gaming. "Different people will accumulate particular gaming skills, knowledge, and frames of reference, according to the patterns of access and peer culture they encounter – and these accumulations will pool as predispositions and manifest as preferences" (Carr, 2005, p. 15).

Various types of multi-user entertainment games can be found on the Internet. Socialization occurs through interaction among game players in clubs, Internet sites, etc., forming an affinity group for specific games (Gee, 2003). Multi-user entertainment game play is often competitive, but it can also be cooperative, with both goal formats increasing player' social interaction. Instructional games have the same social variables and can be used to support cooperative and collaborative learning. "Learning is a social activity, and computers can support the social construction of knowledge, with computer games providing environments which can demand collaboration of the people using them." (Becta, 2001, p. 9).

Game players understand that social interaction with other players will help support the learning of complex game rules. Discussion of game play within a community of players is a means of scaffolding game learning (i.e., it is a form of learning support). Game playing communities are examples of distributed knowledge and, as such, communities promote social interaction as players improve their game knowledge by interacting with others in the community. Learning to socially interact with others is a major side benefit of games. Gaining status within a group via game knowledge and skill is an additional motivation for playing games and may boost self-efficacy and self-esteem.

Gee (2003) makes a strong case that what people think and learn is determined by their interactions with other people who are members of their social group. This assertion implies that researchers may need to consider how student membership in various social groups affects motivation and learning when using games. Students are in some obvious social groups (Gee calls them affinity groups) such as being students in a specific grade in a specific school. Students are also members of various other social groups (e.g., those who play sports) and can form social groups in regard to games.

Malone and Lepper (1987) include interpersonal motivations in their taxonomy of intrinsic motivation in games. Interpersonal motivations (those depending on other people) include cooperation and competition. The strength of these motivations may be very dependent on individual differences or related to personal traits such as gender. As with game context, problems, and interactions, interpersonal game functions may or may not be directly tied to learning outcomes. Cooperation or competition may be an important targeted learning outcome, may be essential to the targeted learning outcome, or may be unrelated to the learning outcome. Malone and Lepper (1987) also identify recognition (e.g., by peers) as an important social motivator.

7. Summary and Conclusions

Ron Edwards (2004), an influential entertainment game designer, has formulated a theory which posits that game players fall into three categories: gamists, narrativists, and simulationists, who have quite distinct goals in playing a game. Gamists seek competition and challenge, narrativists seek story and characters, and simulationists like exploration and experience (Edwards, 2004). Edwards finds that games cannot be designed to meet the expectations of all three types of players. Edwards' player types can be viewed as each having a primary motivation tied to major categories in the framework we propose: gamists are primarily motivated by aspects of game goals (problems), narrativists by game context, and simulationists by game interactions and representations. These game player categories suggest a player-specific trait related to player motivation that also may have significance for the instructional games. Specific game goals will align with the motivations of only a subset of the target audience of the instructional game and may be less effective for those learners with a different goal orientation.

Dempsey et al. (1996) recognize that educational researchers will be increasingly asked how to incorporate games into learning environments and that they may be "perplexed" about how to answer the question through research studies. Some theorists and researchers also posit that the variables of the game–player interaction are so numerous, and so hard to measure and control, that a systematic study of gaming variables may not be productive or even possible (Dempsey et al., 1996; Quinn, 1997). We believe, like Gee (2005, p. 1) that "Good game designers are practical theoreticians of learning…" While good game design is currently an art more than a science, game designers intuitively attend to gaming variables that influence player affect and learning for specific audiences. Variables of interest to entertainment game designers can serve to indicate potential instructional game variables for systematic study.

A major obstacle in studying games is the practical consideration that complex simulation games are difficult and expensive to create. Game makers invest considerable resources in creating games in a highly competitive market and so are reluctant to provide game source code to researchers. Game makers can also be expected to be reluctant to share information gained about what affects players' motivations in playing games. Researchers, lacking resources to create their own games, do not have access to game variables that would allow systematic experimental investigation of gaming variables suggested by a conceptual framework. The resistance of game makers to support research is major barrier to understanding how game characteristics affect learning and there is no easy solution to this issue. Game makers may be convinced to provide source code for older versions of games or strategic alliances could be adopted by game makers and researchers in which early beta versions of games can be manipulated in experimental studies. Study results could be made available to the game maker prior to publication.

A more systematic study of instructional games can result from a conceptual framework that suggests categories of instructional game variables that are related to learning as suggested by a modern cognitive learning theory such as interactive cognitive complexity theory. We suggest that it is useful to view instructional gaming variables from the perspectives of virtual context, problem specification, interaction and control, learning support, and social interaction. More focused studies are required to guide instructional designers in creating instructional games that work. Clear rationales for the design of instructional games and evidence that they are effective learning resources will further the use of games in traditional learning environments.

References

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117–148.

Becta (2001). Computer games in education project report. Retrieved July 20, 2005 from www.becta.org.uk/page_documents/research/cge/report.pdf.

Carr, P. D. (2005) Contexts, gaming pleasures, and gendered differences. *Simulation & Gaming*, *36*(4), 464–482.

Chuang, S. & O'Neil, H. F. (2006). Role of task-specific adapted feedback on a computer-based collaborative problem-solving task. In H. F. O'Neil & R. S. Perez (Eds), *Web-based learning: Theory, research, and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cordova, D. I. & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88, 715–730.

Dempsey, D. V., Lucassen, B. A., Haynes, L. L., & Casey, C. S. (1996). *Instructional applications of computer games*. Paper presented at the *1996 Annual Meeting of the American Educational Research Association, NYC*.

Edwards, R. (2004). System does matter. Retrieved July 28, 2005 from http://www.indie-rpgs.com/_articles/system_does_matter.html.

Fabricatore, C. (2000). Learning and videogames: An unexploited synergy. Retrieved June 20, 2005 from www.learndev.org/dl/FabricatoreAECT2000.pdf.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, *33*(4), 441–467.

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).

Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005). Endogenous fantasy and learning in digital games. *Simulation & Gaming*, *36*(4), 483–498.

Henderson, L., Klemes, Y., & Eshet, Y. (2000). Just playing a game? Educational simulation software and cognitive outcomes. *Journal of Educational Computing Research*, *22*(1), 105–129.

Hsieh, I. G. & O'Neil, H. F. (2002). Types of feedback in a computer-based collaborative problem solving task. *Computers in Human Behavior*, *18*, 699–715.

Jonassen, D. H. (2002). Learning as activity. *Educational Technology*, *42*(2), 45–48.

Jonassen, D. H., Howland, J., Moore, J., & Marra, R. (1999). *Learning to solve problems with technology – A constructivist perspective*. Columbus, OH: Merrill Prentice Hall.

Kafai, Y. B. (2001). The educational potential of electronic games: From games-to-teach to games-to-learn. Retrieved May 16, 2006, from http://culturalpolicy.uchicago.edu/conf2001/papers/kafai.html.

Kirriemuir, J. (2003). The relevance of video games and gaming consoles to the higher and further education learning experiences (Techwatch Report: TSW 02-01). Retrieved June 30, 2005 from http://www.jisc.ac.uk/uploaded_documents/tsw_02-01.rtf.

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.

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 analyses* (pp. 255–286). Hillsdale, NJ: Lawrence Erlbaum.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction*, *3*, 113–132.

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 motivation for learning. In R.E. Snow & M. J. Farr (Eds), *Aptitude, learning and instruction: III. Conative and affective process analyses* (pp. 223–253). Hillsdale, NJ: Lawrence Erlbaum.

Mitchel, A. & Savill-Smith, C. (2004). The use of computer and video games for learning: A review of the literature. Retrieved June 30, 2005 from http://www.lsda.org.uk/files/PDF/1529.pdf.

Moreno, R. & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*(1), 117–128.

Northup, P. (2002). A framework for designing interactivity into web-based instruction. *Educational Technology*, *41*(2), 31–41.

O'Neil, H. F. & Fisher, Y. (2004). A technology to support leader development: computer games. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leader development for transforming organizations* (pp. 99–121). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal*, *16*(4), 455–474.

Ormrod, J. E. (2004). *Human learning* (4th edn.). Upper Saddle River, NJ: Merrill Prentice Hall.

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

Quinn, C. N. (1997). Engaging learning (available from the Instructional Technology Forum). Retrieved July 30, 2005 from http://it.coe.uga.edu/itforum/paper18/paper18.html.

Ricci, K., Salas, E., & Cannon-Bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology*, *8*(4), 295–307.

Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, *83*, 318–328.

Tennyson, R. D. & Breuer, K. (1997). Psychological foundations for instructional design theory. In R. D. Tennyson, F. Schott, N. Seel, & S. Dijkstra (Eds), *Instructional design: International perspective, volume 1: Theory, research, and models*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tennyson, R. D. & Breuer, K. (2002). Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior*, *18*, 650–668.

A FRAMEWORK FOR THE ASSESSMENT OF LEARNING GAMES

Eva L. Baker and Girlie C. Delacruz
University of California, Los Angeles

Abstract

This chapter describes issues associated with assessment and learning games, including design requirements, such as knowledge representations, task representations, instructional goals, and scoring models. Both outcome-focused measures and process measures are discussed. Throughout, approaches are proposed that support new conceptions of validity, that is, validity interpretations for different assessment purposes. In addition, the use of wrap-around approaches to evaluate existing games is discussed, as is the notion of using a game structure for assessment when the training intervention is not a game.

1. Introduction

Games have been a source of interest to scholars in learning, instruction, and computing at least since the publication of Weiner's (1954) *The Human Use of Human Beings*. The characteristics of games have been variously described (see Malone, 1981, for an early example), but they involve competition with oneself or others, in individual or team play, a set of explicit rules governing behavior and action, often (but not always) some skill, feedback, and the component of chance. When presented as a set of features, game characteristics are not unlike most instructions, including chance. It is clear that a principal purpose of games in general is to entertain, motivate, and energize us, and perhaps allow us to escape, when compared with the day-to-day options that guide the procedures in our lives.

Neither is the idea that games can teach particularly new. Those who play "thinking" games, like Go, chess, bridge, or even poker, understand that there are cognitive strategies, specific procedures, and social behaviors that may influence game outcomes and success.

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

More recent genre descriptions, primarily for games played on computers, include action, role-playing, adventure, strategy, simulation, sports, and racing games (Laird & van Lent, 2001), as well as location-based mobile games (de Souza e Silva & Delacruz, 2006; Squire & Jenkins, 2003). Although the underlying software of a game might be very simple, that is, having only a limited set of contingencies for each move, newer games involve complex artificial intelligence underpinnings to guide and anticipate conceptual moves by the user. This sort of substructure is in addition to the verisimilitude of graphics, motion, and other simulation displays.

This chapter, however, focuses on a highly specific use of games – the assessment or testing of the learner or player. The chapter considers analytical strategies to assure that the goals for assessment and the game are interrelated, approaches to assure adequate instructional and assessment sampling of content parameters, and the assessment purposes as they influence scoring rules and metrics. In short, the chapter addresses concerns of both design and validation of assessments in games intended to teach, which face the same critical issues as do assessments designed for other contexts. The issues of assessment validity can be dealt with by articulating the machinery, procedures, or engines used to make judgments about the acquisition, depth, retention, and utility of the knowledge, procedures, and strategies of the game topics from a learning and performance perspective, as opposed to only as mechanisms to advance game play. Of equal, if not of more, importance, is the extent to which the game (assuming it motivates considerable time investment) develops and sustains the transfer of skills, strategies, and social behaviors to settings other than the game topic domains.

Let us start with a noncomputerized example of a game that measures performance, such as baseball played by Little League teams. The game of baseball has complex rules, moderators, coaches, players, an easy way to determine success for individuals while adapting to a player role (e.g., Runs Batted In or Earned Run Averages), and individual metrics that apply to everyone (e.g., Errors), as well as clear metrics of success for the team (e.g., score differentials between teams or number of wins and losses). While the overall goal of the game is to "win," the instructional attributes of the practices and scheduled contests are intended to develop skills of the students in order to win. The collateral focus of assessment, in addition to winning, for Little League and other sports for children (like American Youth Soccer Organization soccer for 6-year-olds) is to evaluate students' performance and help them improve their performance. In the assessment trade, this is called formative assessment or evaluation (see, for example, Baker & Alkin, 1973; Black & Wiliam, 1998; Scriven, 1967). In team sports, qualifying trials may be held for selection or bracketing purposes, in which feedback is given and adapted to the individual, his or her level of skill, and the position(s) played, and it is done publicly. By the end of a season, or a second season, it is clear who has learned what by observing behavior in the game. Does the team member back up his colleague when appropriate? Can he or she choose where to hit the ball, assuming that level of skill? Do the players share pleasure in their collective efficacy, even if an individual is unsuccessful for a game or a streak? The validity of these games for learning is inherent because what is in part being taught are game-specific skills. In addition, however, general propensities, such as caution, risk-taking, enthusiasm, and perseverance can be observed to transfer across teams, sports, and other domains.

Our interest in this chapter is the use of games for assessment, both when the game itself embodies the goals of learning, and when the game serves as a proxy for typical tests. Our attention is directed to games that are used to make formative assessments (in-the-process-of-learning assessments), as well as those that might be used for criterion trials, either to determine the level of performance of an individual or to gauge the speed and agility with which a learner acquires a new set of skills in an unfamiliar game environment. As experts have reminded us (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education [AERA, APA, & NCME], 1999), validity inferences depend upon the use to which the findings will be put, with uses including classifying the learner (or group), determining rank order, making selections, determining next instructional sequences, and so on.

2. A Word About Validity

Each validity interpretation requires evidence relevant to its purported use. When the use of assessment in a game is formative, one would need to gather information from the player that could help direct the next step of instruction. For example, assessment may take the form of observing a golf player's stance while the player swings a golf club, or for an FBI trainee, what information sources are consulted while the trainee is solving cases. However, when the purpose of the assessment in a game is to certify that the player has developed and reached the set of objectives for the instruction, evidence would need to be gathered that the player can apply the set of skills and practices of the game domain within the appropriate conditions that would demonstrate mastery. For instance, to determine that a player has become proficient at golf, the golfer should demonstrate the ability to consistently score in the 80s on a variety of courses; and the FBI trainee would need to demonstrate solving cases of various complexities. Games relate to assessment in a number of ways that have validity implications. The assessment can be integrated into the game to provide guidance and feedback on the player's progress to attainment of learning objectives. The game may be used as an assessment task for some other educational intervention including a different game or noncomputational type of instruction. Or proficiency in a game may be assessed by means not integrated in the game itself because of feasibility or cost. We will consider all of these approaches. In all, the concern about what is to be learned and measured and how the interpretations guide other decisions are central precepts.

3. Design of Assessments to be Integrated into Learning Games

The design of assessments of learning in games faces the same technical issues as the design of assessments intended for use in other learning environments, such as classrooms or simulations. To ensure that the inferences one makes from the assessment are valid, one must take into consideration the learning objectives of the game and the purpose the assessment will serve. These key characteristics will drive the conceptual framework that underlies the assessment (AERA, APA, & NCME, 1999; Baker, 1997). Although, games invariably adapt options based on learner performance, it is not always the case that the basis for advancing or progression in the game (e.g., receiving points) maps directly to the learning objectives claimed for the game. To use a game as an assessment device that measures attainment of the game's goals, three questions must be initially answered: (a) What do we want trainees/students to learn? (b) How do we know that they have learned? – that is, what classes of actions against which criteria will suffice? A sub-issue of this concern is the third question: (c) How good is good enough? It should be clear that the assessment of performance in game settings conceptually addresses the same issues as assessment in other settings: Does the performance sample the range of the goal state with regard to content? What technical evidence is there that the inferences drawn from performance are appropriate and fair?"

The next section will describe key issues in the design and technical verification of assessments, especially concentrating on specifying the cognitive demands of the game, systematically representing the content from the domain of interest, and the gathering of evidence that the assessment is fulfilling its purpose, all with the aim of answering those questions. The issues will be presented in a sequence that might well describe the task order required for the assessment designer.

4. Technical and Strategic Features Required for Game Assessment

4.1. Cognitive Demands

What do we want trainees or students to learn? In the design of instructional interventions, the answer to this question is often referred to as the learning objectives or learning goals (Anderson & Krathwohl, 2001; Baker, 1974; Bloom, Engelhard, Furst, Hill, & Krathwohl, 1956). In addition to being a primary driver in instructional design, the learning objectives or learning goals shape the cognitive demands of the assessment. The cognitive demands of a task are the set of skills, knowledge, abilities, and behaviors to be assessed and should be a direct extension of what it is we want to be learned in our instruction (AERA, APA, & NCME, 1999; Baker, 1997). The characteristics of the assessment should not be considered independent of the desired learning outcomes of the game. Ideally, the desired learning outcomes and the cognitive demands of the assessment should be considered concurrently in order to assure that they correspond. Otherwise, the inferences one can make from the assessment may be either limited or, in the case of misaligned tasks, suspect.

Defining the cognitive demands of an assessment involves focusing on the domain-independent and domain-dependent set of knowledge and skills that an assessment is supposed to address. For example, Baker's (1997) model-based assessment (MBA) approach to assessment development identifies five families of learning that encompass the key learning outcomes of education and training. These five areas are content understanding, problem solving, teamwork/collaboration, communication, and self-regulation. Each of these families consists of a set of partially domain-independent requirements that serve as the framework of both instruction and assessment. Each of the families can incorporate task syntax that reflects research in the area. For example, measuring problem solving may involve creating tasks whose cognitive demands include making sense of information to identify a problem and deciding on an appropriate solution, whereas measuring content understanding could require the ability to explain why a procedure works (Baker & Mayer, 1999). Because these frameworks are largely independent of content, they can be reused as they are embedded in subsequent specific content domains. Thus, the problem-solving syntax chosen for one game (a role-playing game) might very well be embedded in a different content and game style, such as a quest game. Superficially, the games will look very different; it is our thought, however, that a common thread of problem-solving competence can be developed in different domains, an outcome that would support transfer of learning. Only after the cognitive demands of a task have been determined are these areas defined by specific subject matter. Baker, Niemi, and Chung (in press) have described the continuum that must exist among tasks that are peculiarly domain specific and those that may be applicable to a wider set of content domains. For example, even very domain-specific tasks, such as making an incision for an appendectomy, will have a range of contexts that require adaptive behaviors involving depth, speed, and consequences. Other tasks, such as figuring out an answer to a reading comprehension question, may have general strategies, such as the review of syntax and lexical context clues, to aid in the correct response. Thus one enormously important notion about the objectives to be taught, or perhaps only to be measured, involves the level of generality of the applicable cognitive learning and the domains to which it is supposed to apply.

4.2. Domain Representation

Defining the cognitive demands of a task requires that explicit representation of the content, knowledge, skills, attitudes, and other properties of the domains or constructs be assessed, as well as the behaviors or performances that reveal these constructs (AERA, APA, & NCME, 1999; Baker, 1997; Baker & Mayer, 1999; Mislevy, Steinberg, & Almond, 2002). A domain representation includes a representative description of the domain on which the assessment is focused and the range of eligible content. It makes explicit what is to be covered in the assessment, thus defining the universe of what is to be learned and assessed. Therefore, a domain representation can serve several functions by providing guidance for assessment developers, assisting teachers to focus on what is most important for instruction, and allowing external review of the assessment development process.

One function of a domain representation is to explicitly represent the cognitive outcomes to be measured. As stated earlier, part of the MBA approach to assessment development is to embed subject matter content into the cognitive demands of a task. For instance, for problem solving, the domain-specific aspects of the task would include the specific content matter, procedural knowledge, strategies, and discourse. That is, at one level of abstraction, problem solving could be defined as the act of looking at various information sources and identifying the problem. However, the actual practice of problem solving for firefighters, for example, may look very different from the problem-solving practices of, say, a historian in terms of specific content matter. While both domains may require identifying a problem from various sources of information, what serves as those sources (e.g., reports from other areas of the building for a firefighter versus original source materials for a historian) will differ from domain to domain.

Another approach to assessment that emphasizes the importance of domain representation is evidence-centered design (ECD). ECD shares with MBA its emphasis on the importance of being explicit about what is to be measured (Mislevy, Steinberg, Breyer, Almond, & Johnson, 1999). Information gathered from cognitive task analyses forms the basis of a student model (i.e., what set of knowledge skills or attributes are assessed). The student model specifies those variables in which the student is expected to demonstrate proficiency, using a mathematical structure in which a joint probability distribution expressing the relationship among the variables is represented (Mislevy et al., 2002). Both of these models demand that the process of assessment development include a clear definition of the nature of the assessment.

The other function of domain representations is to explicitly define the scope of the task by setting content limits for the domain. By explicitly defining the universe, domain representations assist assessment developers by forming the basis on which tasks are sampled over the domain, as well as the range for which skills are expected to generalize (Baker & Mayer, 1999). For instance, the range of content eligible for sampling from the domain representation of firefighting would be different for a domain representation that only targeted firefighting skills in buildings versus a domain representation that included firefighting in many types of scenarios, such as in forest fires and on naval ships. While the latter might subsume some of the knowledge and skills needed to fight fires in buildings, the scope of the content eligible for assessment might include additional factors that needed to be learned, and thus measured.

Various knowledge representation techniques have been explored as a means for domain representation. Knowledge representation techniques facilitate an explicit, precise, and externalized representation of the domain to be covered (Baker, Chung, & Delacruz, in press). Examples of earlier knowledge representation techniques focused on well-defined areas such as mathematics. These approaches used either set theory, where broad classes of performance using algorithmic rules generate items that could be sampled to yield representative instances of test items (e.g., Hively, Patterson, & Page, 1968), or Tatsuoka's rule space methodology, which represents items by well-defined skill attributes forming the basis of a hierarchy of skill (Katz, Martinez, Sheehan, & Tatsuoka, 1998; Tatsuoka & Tatsuoka, 1983). While applicable for well-defined areas such as mathematics, these approaches have been criticized for their inability to capture the more complex set of behaviors that are found in domains such as writing or history (see Baker & O'Neil, 1987, for an in-depth review). Recent methods of knowledge representation that can model more complex domains include the use of Bayesian networks to graphically model the dependencies between variables of interest in relation to actual student performance, such as in licensing dental hygienists (Mislevy et al., 2002), or to assess student performance in a biology-based problem-solving simulation (Chung, de Vries, Cheak, Stevens, & Bewley, 2002). Likewise, Schaffer's notion of *epistemic games* (Shaffer, 2005; Shaffer, Squire, Halverson, & Gee, 2005) calls for intense ethnographic studies to understand the practices of the professions out of which games are developed to ensure that actions and behaviors during game play are aligned with what happens in natural contexts.

4.3. Task Representation

Domain representations explicitly define what it is one wants the students or the trainees to learn, thus providing a basis for what is to be assessed. What is also required is to define what sort of evidence can substantiate the inferences made that learning has occurred in that particular domain. The task representation serves this function. In the context of games, the task representation is comprised of the domain representation, the materials with which the learners are expected to interact, the game narrative or scenarios, response expectations (e.g., the actions taken by and behaviors of the game player[s]), and the scoring method for the measures and performance. Whereas the domain representation defines the universe that will be covered, the task representation defines the conditions under which performance is to be interpreted. It is driven by the purpose or goal of the assessment and determines what classes of actions against which criteria will suffice. If the tasks are misaligned with the domain representation or if the domain representation does not adequately represent the target environment, then the inferences one makes about what is learned may not be well supported. For example, the act of deciding what formation is appropriate when faced with an enemy squad is important in both the natural context of battle and in the game *Full Spectrum Warrior*. However, in battle, the commands are issued verbally or visually with hand signals and are equipped with their own set of specific discourse and practices, whereas in the game, the commands are issued by moving the cursor, obviously reducing verisimilitude. As a result, game players can perform well, directing their squads into appropriate formations to meet various threats. However, because the players are not required to issue commands verbally or visually, the predictive validity of game performance is compromised as far as determining whether the player is ready for battle. Though this example can be argued to be simply an issue of the fidelity of the game, issues of fidelity are crucial when performance in a game is used to assess whether certain practices and skills have been mastered. That is, once a game is to be used for assessment purposes, a prediction of future performance is always implied. As a result, game tasks must require performance of critical behaviors or actions. Either the game should require actions that demonstrate mastery in the domain, if the purpose is certification, or elicit behaviors that can inform how the system should shape the next steps or courses of instruction.

Task representation may also differ as a function of the range of cognitive demands and priority of purposes. Consider designing assessments for a game intended to train FBI agents. A task representation for the part of the game used as a formative assessment to measure content understanding could be to determine whether the trainees had learned how to run basic database procedures. This part of the game would simply have the trainee run a set of comparable searches. Another example could be to determine whether or not FBI agents understood how to interview witnesses of various ethnicities, which would require the trainee to interview other players (either virtual or, in the case of massively multi-player online role-playing games, other actual players). Because the purpose of these assessments is formative, the trainee might be allowed to retry attempts, the game might provide feedback or assistance to the trainee, or the game might use the trainee's performance to determine the next situation to be directed to the trainee. A different task representation would be needed, however, to certify whether the FBI agent had learned the decision skills essential to investigations involving one or more victims. The game would need a task in which the trainee, when given a range of specified scenarios (for example, at an ambiguous crime scene), could identify the key problem by making sense of the evidence obtained, interpret medical examinations of the victims, conduct database searches, interview witnesses, seek new information sources to compensate for missing data, determine the conditions that apply to all victims, and choose a plausible next course of action once a suspect or suspect type is identified.

A mismatch can occur if the assessment extends beyond the content and skill boundaries of the learning tasks or inadequately represents them. That is, one cannot make inferences about whether or not trainees can solve criminal cases by having them do a task that only involves running a fingerprint match. Likewise, even if the assessment task represents the extended skills required to solve provided cases, support for a specific inference that a trainee knows how to match fingerprints can be provided only if the trainee was required to run a test to match fingerprints. In summary, the task representation defines the context from which one can observe and gather information about students' or trainees' performance in a domain in order to enable drawing inferences about their competency in the domain (AERA, APA, & NCME, 1999; Baker, 2002; Messick, 1995; Mislevy et al., 2002).

4.4. Scoring Models

Crucial to assessment design and validity interpretation is the step in assessment R&D of making explicit what scoring model is used. That is, one must reveal the method of translating observed performance into meaningful scores. Because scores are often used as the main indicator of performance, it is important to be concrete about the theory underlying the combination of observations as well as operational details, including how the observations are combined, how they are scored and scaled, and how they relate to other measures of the construct. The scoring model should include the measurement scale, the scoring criteria, performance descriptions of each criterion at each point on the scale, and sample responses that represent various levels of performance (AERA, APA,

& NCME, 1999; Baker, Aschbacher, Niemi, & Sato, 1992/2005; Baker, Chung et al., in press). The theory of the scoring models needs to be made explicit, either in terms of desired attributes of performance or empirical models supporting elements and combinations. These models also need game context validation, including contrasts among expert and novice judges and players, reproducibility, and evidence that irrelevant features are not inadvertently skewing scores.

In an ideal setting, the architecture for assessment, learning, and instruction is integrated such that summary scores or behavioral profiles have meaning and are designed in rather than noticed after the game has been developed.

5. Games as Assessment

Using the domain guidance above, a game itself can be used as the context for assessment. The underlying game engine invariably constrains the range or levels of adequate performance in a set of tasks. For example, player progress in the game could result either from reaching some predefined goal (e.g., get all the tokens), or achieving a certain score (e.g., reach finishing line under 60 s), or demonstrating mastery of a procedural skill (e.g., defeat an enemy using a strategy of mass attack). How one builds the underlying engine could limit expectations of players to counting, speed, combined metrics, or complex solutions within time and tools provided. Moreover, because most games are linear (Gredler, 2001) and interaction is dynamic, detailed subsets of learning objectives and how they are assessed can change over the course of the game. The power of the underlying game engine is that these changes in expectations of students can be built into the game enabling increases in challenge, complexity, or even in the cognitive demands required as the game progresses. Though intelligent tutoring systems (ITS) may share similar characteristics in that the system within which a student interacts may be modified based on performance, ITS are not generally "playful" in nature, that is, there is no element of play or competition, and the learner usually engages as a learner rather than a character or role. Therefore, ITS may often lack the motivational aspects that are inherent in many effective games.

Most games start off with "tutorials," which simply involve understanding the mechanics or basics of a required skill in the game. To return to the example of a game aimed at training FBI trainees, early stages of the game might focus on training in how to run database searches or the protocol of interviewing witnesses because the mechanics of executing these tasks can provide some insight into their purpose and utility. The middle stages of the game may involve solving very simple cases where identifying the correct suspect only requires demonstration of identifying which database search to run.

The last part of most games usually involves complex challenges that subsume all of the previously learned skills, similar to a cumulative final exam, which provides integrative and summative information about competency (e.g., did they master the full set of knowledge and skills across a range of critical contexts?). For instance, in fighting games, final completion of the game may often require fighting all-powerful bosses who have the powers of all of the previous bosses, such as in *Viewtiful Joe* (Squire, in press). Or in the case of the FBI training game, final completion of the game would require that the trainees be able to solve complex criminal cases, using all the skills that were tested during level progression (e.g., running tests, database searches, and interrogating witnesses).

5.1. Process Assessments

After the initiation of the game, interaction is often nonlinear, offering multiple paths to a solution. Assessments that capture details of the players' process, rather than just the solutions attained, can be exploited to get a finer grained picture of learning. Especially because of the interactive nature of game playing, process data may be the best way to capture proficiency, as it is often embodied in action. Because game playing is centered on activity, it is often necessary to see what people do to determine proficiency (Barab, Fajen, Kulikowich, & Young, 1996; Squire, 2005).

In other computer-based performance assessments, process data have been captured using various methodologies. Process data are most useful if the game is designed so that the task structure leaves little doubt about the content that the learner is viewing. That is, embedding features into the game to capture intent, such as requiring the learner to perform an explicit action to access information, can yield useful information about problem-solving strategies. For instance, examination of time-stamped logs of user actions such as screens visited or buttons pressed can be indicative of problem-solving strategies. Barab et al. (1996) divided such files into meaningful units using retrospective verbal protocols and using transitions between screens as the unit of analysis. Chung and Baker (2003) employed a "click-through" interface to capture intent by requiring the learner to perform an explicit action to access information (e.g., clicking on a button to see the time remaining). In IMMEX (Chung et al., 2002), each screen presents a single topic, rather than multiple topics; therefore there is little doubt about what the learner is attending to. The assumption is that the learner is aware of the availability of the information, and therefore clicking on the information is intentional, rather than some random act. Gathering process data is again tied to the appropriate domain and task representation in assessment design. One must consider what skills are to be assessed as well as what behaviors or responses need to be captured in order to properly infer that the skills have been learned. For example, in the FBI training game, as a formative assessment to evaluate student performance, it might be important to capture which information sources the trainees access or the number of failed attempts at running a database search. Simply looking at the metric of number of levels a trainee passed would be less useful in pinpointing areas of difficulty.

When using process data to gather observations about students' performance, the scoring techniques used are essential to performance interpretation. There are multiple techniques for scoring performance in complex tasks, such as expert performance-based scoring, modeling expert judgments, data-driven methods, and domain-modeling methods. Expert performance-based scoring is consistent with criterion-referenced measurement in which expert performance is considered to be the gold standard against which to evaluate student performance (Baker, 1997; Baker, Freeman, & Clayton, 1991; Chi, Glaser, & Farr, 1988). Using this technique to score game play, one could have subject matter experts play the game to serve as the expert referent and examine how close trainees are to targeted levels of performance. This is different from using expert ratings of student performance as the gold standard for evaluation (e.g., Clauser, Swanson, & Clyman, 1999; Margolis & Clauser, 2006; Williamson, Behar, & Hone, 1999). In this case, performance is evaluated using the expert ratings, possibly by having instructors or teachers watch the student play the game or evaluate the student process logs.

Moreover, process data can also be analyzed using various methodologies such as lag sequential analyses (Chung & Baker, 2003), neural nets (Vendlinski & Stevens, 2002), or Bayesian networks (Chung et al., 2002; Martin & VanLehn, 1995). Chung and Baker (2003) used lag sequential analyses to analyze the sequences of user actions in a bicycle pump simulation. The technique of artificial neural networks clusters students' performance based on their clickstreams, and raters examine characteristics of their clustered performance (e.g., Stevens & Cassillas, 2006; Vendlinski & Stevens, 2002). Interpretations of the performances represent the "score." Finally, process data can be scored via the use of Bayesian networks. For example, Chung, Delacruz, Dionne, and Bewley (2003) computed moment-to-moment behavioral indicators from clickstream data and synthesized the data using a Bayesian network that modeled the interrelatedness between the variables of the task and the behaviors that were evidence of the variables.

6. Assessing the Effectiveness of a Game

When there are technological constraints for game analysis (e.g., no access to the source code), or when the initial purpose of the game was not to enhance learning, the measurement of game effectiveness can be affected. When assessment cannot be embedded in the game itself, one solution is to use "wrap-around" assessment tasks, which may be given in a parallel setting (either electronic or paper) on a pre-game, during-the-game, and/or postgame basis. If the game has specific domain and task features specified, creating or choosing an appropriate wrap-around assessment task is easier. These tasks would require that the student use the same skills and exhibit the same class of behaviors as those used in the instructional intervention, as well as allow the demonstration of the ability to generalize those skills and behaviors to new (but comparable) situations (Baker & Mayer, 1999).

Examples of wrap-around options include structural assessments such as concept maps, where the player is asked to represent graphically (using pull-down menus and other tools) the features of the domain and skill set and the relationships of the components to one another. Such graphical representations of a domain can be conceived as a network of nodes that represent concepts and links. For example, to assess whether or not trainees learned the concepts in the game Space Fortress, Day, Arthur, and Gettman (2001) used Pathfinder to compare knowledge structures and performance in the actual game.

The primary objective of Space Fortress is to fly a spaceship and battle a fortress, while avoiding being destroyed by enemy mines. Trainees were asked to make similarity ratings of key concepts in the game such as the ability to control the ship and the destruction of a mine. The trainee ratings were compared to an expert referent structure. Results demonstrated that trainees whose ratings were more similar to an expert structure performed better in the game.

Retention or transfer tasks that target the content to be learned in the game can also be used as a wrap-around assessment of learning (e.g., Mayer, Mautone, & Prothero, 2002; Moreno & Mayer, 2000). For instance, to evaluate learning in the Design-a-Plant game, Moreno and Mayer (2000) used a retention task (e.g., "Write down all the roots you can remember") to measure whether or not students had learned basic declarative information from the game. However, to measure whether or not students had learned how to design a plant that can thrive under different conditions, the transfer task asked the students to apply their knowledge to a situation by designing plants for various conditions and identifying the conditions under which certain plants could thrive, and then to provide explanations to justify their answers.

The tasks used by Moreno and Mayer (2000) are a good example of choosing wrap-around assessments that are appropriate for the expected learning outcomes. The retention task measured whether or not the student remembered what was presented, whereas the transfer task measured the student's ability to apply the knowledge to new situations. When the goal of instruction calls for meaningful learning, a retention task is not adequate to determine that such learning has occurred because it can only show whether information was retained. Therefore, assessments that require students to actually apply knowledge in a problem-solving transfer task can be a more powerful indicator that meaningful learning has occurred (Baker & Mayer, 1999).

7. Deep Architecture of Games

Should games be designed one at a time? Should their assessments be idiographically developed? Although it is obvious that top-down, reusable components can be created as part of an overall game engine, and for tests as well, in reality one-at-a-time game design is the norm because of schedule and budgetary constraints. However, it is theoretically and instructionally desirable to consider the design of a suite of games and their assessments all at once. Designing games to include learning objectives can be accomplished by assuring that the game requires a recurrent set of cognitive demands. For example, if problem solving is thought to be an objective of importance, it could, as described above, be incorporated into the architecture or engine design of a game. This kind of fundamental design supports the actual steps of engine design and should reduce cost and time in subsequent developments, within the target game or for different games. Most importantly, such design can allow students' cumulative performance across games with common cognitive architecture to be monitored and reported back. We can see how problem solving progresses by capturing the critical incidents, decisions, and performances of the player, and feeding them back at intervals, to the player, the teacher or trainer, or to parents. This comprehensive and cumulative monitoring can be accomplished without compromising the variety or levels of play in which the learner engages, and in a way that is not intrusive to the learner. Of course, privacy provisions must be instituted if the data are to be used for purposes other than legitimate feedback in an educational setting.

8. Outcomes

We have sketched outcomes that involve cognitive tasks, such as problem solving, communication, and metacognition (planning, self-monitoring) as features that support game design and must also be explicitly selected or rejected as part of the game specification. For most designers, however, the salient outcomes involve content that the game is supposed to teach, such as content facts, or procedural knowledge – the how-to-accomplish a particular task – such as algorithms for solving algebra problems or step-by-step tasks to be coordinated among team members. In addition, any game will develop incidental knowledge that is not its primary goal. The extent to which a game develops useful auxiliary knowledge may be one way of judging its value.

Assuming the outcomes of a game are first tied to the general claims made for it, and to the specified objectives, performance can be judged along multiple axes that involve the degree of complexity, the richness or sparseness of content, the understanding of interrelationships, and the ability to perform criterion (or the desired final) tasks.

More important outcomes of a game, however, are typically external to the game play itself. These include the extent to which game outcomes can be applied in different contexts involving different constraints, whether high-quality performance can be achieved in speeded and nonspeeded conditions, and whether the player can compensate and fill in for missing information. A more remote level of transfer outcome could be determined to see whether the player adopts common problem-solving strategies as they are appropriate to very different settings and content objectives. It is our belief that assessing the performance of game players during the game and on the job is an essential component to games that claim to improve performance outside the game setting itself.

9. Affect and Motivation

It is also important to consider noncognitive factors that may contribute to learning. This is particularly crucial for games, which are often viewed as another tool to mediate learning, due to their intrinsic motivational properties (Gee, 2003; Malone, 1981). Factors that have been shown to predict academic achievement include achievement motivation (i.e., the enjoyment of surmounting obstacles and completing tasks in order to achieve academic success and excellence), goals (e.g., performance approach or avoidance with goal-directed behavior), self-efficacy (i.e., self-evaluation of one’s ability to be successful in a particular content area), and effort (i.e., the extent to which one works hard on a task)

(Hidi & Harackiewicz, 2000; Robbins, Lauver, Lee, Davis, Langley, & Carlstrom, 2004). In the context of games, these factors can be assessed by self-report or as a function of performance in the game itself. For instance, effort could be measured by the amount of time spent on a task in a game.

One of the motivating aspects of games is the presence of multiple goals (Malone, 1981). Game players may have a goal to complete a basic task, such as defeating an enemy, but they may also have another goal, such as efficiency (e.g., defeating an enemy quickly). Therefore, measures of goal-directed behavior could look at attempts to increase scores or to achieve a goal as fast as possible. For example, Lee, Luchini, Michael, Norris, and Soloway (2004) found that in a mathematics drill game, Skills Arena, students were increasing the difficulty of the game without direct instruction from the teachers or researchers by increasing the speed at which the questions would travel across the screen. This could be used as a measure of motivation.

Moreover, the increasing presence of massively multiplayer online games (MMOG) pushes for measures of extrinsic motivators such as competition, collaboration recognition, and material goods (Barab, Arici, & Jackson, 2005; Bonk & Dennen, 2005; Chen, Shen, Ou, & Liu, 1998). One of the key features of such games is the ability to customize one's avatar. The tools to change the avatar's appearance are often made available as a result of some achievement in the game. For instance, in Whyville, "clams" are used to purchase ways (e.g., makeup, change in hair color, or outfits) to personalize one's avatar, and clams are obtained by accomplishing some task in the game (Feldon, 2007; Fields, 2007). Therefore, while earlier studies on motivational aspects of games have focused on intrinsic motivators, MMOGs drive the need for assessment of extrinsic motivators for learning in games.

10. Summary and Conclusions

This chapter has presented a brief tour of game assessment requirements. These requirements include the cognitive and content representations needed to assure accomplishment of stated goals. We cannot emphasize enough that these must be designed into the game. They must also be explicitly measured during the game and in the field, using appropriate experimental designs. It is invariably true that games produce more outcomes than the developers intend, and many of them may be ultimately more important and long lasting. However, it is important to document these additional outcomes, whether they affect knowledge, strategy, or behavior, to help develop the models of what games can do under various conditions.

It is our intention to continue our experimentation with games and game-like environments in order to flesh out with adequate evidence guidelines for game designers. Our parting note is that if you expect anything to be learned through game playing, it has to be described, incorporated in the game in multiple trials, and measured during and after the game with appropriate tasks and settings.

Acknowledgment

Funding for this research was provided by the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Office of Educational Research and Improvement, US Department of Education, and by the Office of Naval Research award number N00014-06-1-0711, as administered by the Office of Naval Research. The findings and opinions expressed in this report do not reflect the positions or policies of the US Department of Education or the Office of Naval Research.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anderson, L. W. & Krathwohl, D. R. (Eds) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Baker, E. L. (1974). Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. *Educational Technology, 14*(6), 10–16.

Baker, E. L. (1997). Model-based performance assessment. *Theory Into Practice, 36*, 247–254.

Baker, E. L. (2002). Design of automated authoring systems for tests. In National Research Council, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education (Eds), *Technology and assessment: Thinking ahead: Proceedings from a workshop* (pp. 79–89). Washington, DC: National Academy Press.

Baker, E. L. & Alkin, M. C. (1973). Formative evaluation of instructional development. *AV Communication Review*, *21*, 389–418. (ERIC Document Reproduction Service No. EJ 091 462).

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992/2005). *CRESST performance assessment models: Assessing content area explanation* (CSE Tech. Rep. No. 652). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L., Chung, G. K. W. K., & Delacruz, G. C. (in press). Design and validation of technology-based performance assessments. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds), *Handbook of research on educational communications and technology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, E. L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M. C. Wittrock & E. L. Baker (Eds), *Testing and cognition* (pp. 131–153). Englewood Cliffs, NJ: Prentice-Hall.

Baker, E. L. & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*, 269–282.

Baker, E. L., Niemi, D., & Chung, G. K. W. K. (in press). Simulations and the transfer of problem solving knowledge and skills. In E. L. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds), *Assessment of problem solving using simulation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, E. L. & O'Neil, H. F., Jr. (1987). Assessing instructional outcomes. In R. M. Gagné (Ed.), *Instructional technology* (pp. 343–377). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barab, S. A., Arici, A., & Jackson, C. (2005). Eat your vegetables and do your homework: A design-based investigation of enjoyment and meaning in learning. *Educational Technology, 65*(1), 15–21.

Barab, S. A., Fajen, B. R., Kulikowich, J. M., & Young, M. F. (1996). Assessing hypermedia navigation through pathfinder: Prospects and limitations. *Journal of Educational Computing Research*, *15*, 185–205.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74.

Bloom, B. S., Engelhard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook I: Cognitive domain*. New York: Longmans, Green.

Bonk, C. J. & Dennen, V. P. (2005). *Massive multiplayer online gaming: A research framework for military training and education* (Tech. Rep. 2005-1). Washington, DC: Advanced Distributive Learning Initiative.

Chen, G. D., Shen, G. Y., Ou, K. L., & Liu, B. J. (1998). Promoting motivation and eliminating disorientation for web-based courses by a multi-user game. *Proceedings of the Ed Media/Ed Telecom 99 World Conference on Educational Multimedia and Hypermedia and World Conference on Educational Telecommunications*, Freiberg, Germany.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds) (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chung, G. K. W. K. & Baker, E. L. (2003). An exploratory study to examine the feasibility of measuring problem-solving processes using a click-through interface. *Journal of Technology, Learning, and Assessment, 2*(2). Available January 29, 2007 from http://jtla.org.

Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., & Bewley, W. L. (2003). Linking assessment and instruction using ontologies. *Proceedings of the I/ITSEC*, *25*, 1811–1822.

Chung, G. K. W. K., de Vries, L. F., Cheak, A. M., Stevens, R. H., & Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. *Computers in Human Behavior, 18*, 669–684.

Clauser, B. E., Swanson, D. B., & Clyman, S. G. (1999). A comparison of the generalizability of scores produced by expert raters and automated scoring systems. *Applied Measurement in Education, 12*, 281–299.

Day, E. A., Arthur, W., Jr, & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*, 1022–1033.

de Souza e Silva, A. & Delacruz, G. C. (2006). Hybrid reality games reframed: Potential uses in educational contexts. *Games and Culture, 1*, 231–251.

Feldon, D. (2007, April). *Mixed methods for mixed reality: Overcoming methodological challenges to understand user activity in a massive multi-user virtual environment*. Paper to be presented at the *Annual Meeting of the American Educational Research Association*, Chicago, IL.

Fields, D. A. (2007, April). *Learning by cheating? Investigating the science in cheat sites for informal educational multi-user virtual environments*. Paper to be presented at the *Annual Meeting of the American Educational Research Association*, Chicago, IL.

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

Gredler, M. E. (2001). Educational games and simulations: A technology in search of a (research) paradigm. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 521–540). Mahwah, NJ: Lawrence Erlbaum Associates.

Hidi, S. & Harackiewicz. J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*, 151–179.

Hively, W., Patterson, H. L., & Page, S. H. (1968). A “universe defined” system of arithmetic achievement tests. *Journal of Educational Measurement, 5*, 275–290.

Katz, I. R., Martinez, M. E., Sheehan, K. M., & Tatsuoka, K. K. (1998). Extending the rule space methodology to a semantically-rich domain: Diagnostic assessment in architecture. *Journal of Educational and Behavioral Statistics, 24*, 254–278.

Laird, J. E. & van Lent, M. (2001). Human-level AI’s killer application: Interactive computer games. *AI Magazine, 22*(2), 15–26.

Lee, J., Luchini, K., Michael, B., Norris, C., & Soloway, E. (2004). More than just fun and games: Assessing the value of educational video games in the classroom. *Proceedings from the Conference on Human Factors in Computing Systems*, 1375–1378. Available January 29, 2007 from http://portal.acm.org/citation/cfm?doid=985921.986068.

Malone, T. W. (1981). What makes computer games fun? *Byte, 6*, 258–277.

Margolis, M. J. & Clauser, B. E. (2006). A regression-based procedure for automated scoring of a complex medical performance assessment. In D. M. Williamson, I. I. Behar, & R. J. Mislevy (Eds), *Automated scoring of complex tasks in computer-based testing* (pp. 123–167). Mahwah, NJ: Lawrence Erlbaum Associates.

Martin, J. & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human–Computer Studies, 42*, 575–591.

Mayer, R. E., Mautone, P., & Prothero, W. A. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*, 171–183.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5–8.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the roles of task model variables in assessment design. In S. Irvine & P. Kyllonen (Eds), *Generating items for cognitive tests: Theory and practice* (pp. 97–128) Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing simulation-based performance assessment. *Computers in Human Behavior, 15*, 335–374.

Moreno, R. & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages*, Journal of Educational Psychology*, *92*, 724–733.

Robbins, S. B., Lauver, K., Lee, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*, 261–288.

Scriven, M. (1967). The methodology of evaluation. *AERA Monograph Series on Curriculum Evaluation, 1*, 39–83.

Shaffer, D. W. (2005). Epistemic games. *Innovate, 1*(6). Reprinted in *Computer Education* (in press). Available January 29, 2007 from 111.innovateonline.info/index.php?view=article& id=79.

Shaffer, D. W., Squire, K. D., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*, 104–111.

Squire, K. D. (2005). Educating the fighter: buttonmashing, seeing, being. *On the Horizon*, *13*(2), 75–88.

Squire, K. D. & Jenkins, H. (2003). Harnessing the power of games in education. *Insight*, *3*, 7–33.

Stevens, R. H. & Casillas, A. (2006). Artificial neural networks. In D. M. Williamson, I. I. Behar, & R. J. Mislevy (Eds), *Automated scoring of complex tasks in computer-based testing* (pp. 259–312). Mahwah, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement, 20*, 221–230.

Vendlinski, T. & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment, 1*(3). Available January 29, 2007 from http://www.jtla.org.

Weiner, N. (1954). *The human use of human beings: Cybernetics and society*. Boston, MA: Houghton Mifflin.

Williamson, D. M., Behar, I. I., & Hone, A. S. (1999). "Mental model" comparison of automated and human scoring. *Journal of Educational Measurement, 36*, 158–184.

A FORMATIVE EVALUATION OF THE TRAINING EFFECTIVENESS OF A COMPUTER GAME

Hsin-Hui Claire Chen and Harold F. O'Neil
University of Southern California

Abstract

According to previous studies, a computer game may be one of the most effective tools to improve problem-solving. However, despite computer games and simulations' potential power in instruction and training, research on their effectiveness for adults is limited, and a framework of evaluation is lacking. This research on the training effectiveness of a computer game was conducted in a research environment with reliable and valid measures of problem-solving; that is, the knowledge maps as content understanding measure the retention and transfer questions as problem-solving strategies measure, and the trait self-regulation questionnaire as self-regulation measure. Implications for further study were discussed.

1. Literature Review

1.1. Games and Simulation

As pointed out by Ruben (1999), researchers such as Abt (1970), Coleman (1969), Boocock and Schild (1968), Gamson (1969), Greenblat and Duke (1975), Pfeiffer and Jones (1969–1977), Ruben (1978), Ruben and Budd (1975), and Tansey and Unwin (1969) started to notice the potential effects of simulations and games in instruction decades ago.

The merits of computer games include facilitating learning by doing (e.g., Mayer, Mautone, & Prothero, 2002), and triggering motivation and enjoyment, although some researchers argued that it is the instructional strategies or methods in a medium instead of the medium itself that influences learning (e.g., Clark, 1983, 1994). In addition, computer

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

simulation games engage learners in a simulated experience of the real world, which makes learning potentially practical (Martin, 2000; Stolk, Alexandrian, Gros, & Paggio, 2001). Due to those merits, games and simulations have been applied in various fields and settings, such as that of business, of K-16 organizations, and of military organizations (e.g., Baker & O'Neil, 2003; O'Neil & Andrews, 2000). Furthermore, as pointed out by Stolk et al. (2001), for the training in some settings where practice and exercises in real situations are expensive and dangerous, computer games and simulations are helpful. For example, military settings applied computer-based training tools, such as war-games and simulators for task training. The same situation happens in the field of environmental crisis management; practicing dealing with natural disasters and industrial emergencies is usually very expensive and dangerous, therefore it is necessary to apply instructional gaming (Stolk et al., 2001).

According to O'Neil and Fisher, the effects of computer games and simulations can be generally divided into five categories: promotion of motivation, enhancement of thinking skills, facilitation of meta-cognition, enhancement of knowledge, and attitudes. However few studies have shown the empirical effects of games and simulations on training and learning (O'Neil & Fisher, 2004). They also indicated that despite the potential power of computer games on instruction and training, research on their training effectiveness is limited, and further, there is little gaming literature that is helpful in designing a formative evaluation of games. As pointed out by Ruben (1999), there is not enough research on the evaluation of games' instructional effectiveness and their validity and reliability. According to researchers (e.g., O'Neil et al., 2002; Quinn, 1996), one of the critical concerns is time and expense. Therefore, more investment should be put in the analysis and studies on computer game evaluation (O'Neil et al., 2002; O'Neil & Fisher, 2004; Quinn, 1996; Ruben, 1999).

Based on Gredler's (1996) categorization, SafeCracker, the puzzle-solving game used in this study had the characteristics possessed by both games and simulations: in the game, the players were transferred to a simulated environment; a player of the game was fulfilling a role as an expert safecracker; the event sequence of SafeCracker was branching (Gredler, 1996); actions and decisions made previously by the player would influence or result in the following situations and problems; the player was a component of the game's scenario and executed the tasks of his/her role of finding clues and tools, and breaking safes logically. Gredler (1996) defines a phrase that means the mixture of games and simulations' features; that is simulation games or gaming simulation. For the purpose of this study, we will use Gredler's definition of simulation game or gaming simulation.

1.2. Problem-Solving

According to research, problem-solving is one of the most critical competencies whether for lifetime learning or accomplishing tasks, whether in job settings, in academic settings (e.g., Dugdale, LeGare, Mathews, & Ju, 1998), or any other settings. Although there is substantial previous research which reveals the utility of problem-solving (e.g., Mayer, 2002), the methods to assess problem-solving ability are in need of refinement. For example, a common method used to assess problem-solving is the use of a standardized test consisting of unconnected multiple choice questions that fail to provide students and teachers with information on what problem-solving processes they should use and why. Although we can find the most valid measures for problem-solving competence in the cognitive science literature, these measures (e.g., think-aloud protocols), however, are inefficient to assess performance for diagnostic purposes, since their scoring is laborious and time-consuming (O'Neil, 1999). As a result, the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has developed a problem-solving assessment model to measure content understanding, problem-solving strategies, and self-regulation, the three elements of problem-solving.

Games and simulations have potential use in teaching and learning, and have been applied in the field of business, in academic organizations, and in military settings, and as argued by Quinn (1991), computer games may provide effective environments for problem-solving. But there is little research of games' training effectiveness. This researcher conducted this study focusing on the evaluation on a computer game with regard to its effectiveness in improving problem-solving.

Researchers (e.g., Clark, 1983, 1994, 2001) suggest that media themselves do not foster learning, therefore the theories and instructional strategies embedded in games and simulations are important. Most commercial games are not designed with an instructional strategy in mind. Experiential learning, discovery learning, and deductive learning are three instructional strategies to support training with computer games. Experiential learning (Ruben, 1999) is an important strategy focusing on increasing the student's control and autonomy, an aspect of constructivism, which relies heavily on student initiative and teachers' being available for guidance. Inductive/discovery learning is a strategy where learners acquire knowledge by exploring in environments by themselves with less or no guidance (Kalyuga, Chandler, & Sweller, 2001; Mayer, 2004; Sweller, 1994). Deductive strategy to teaching is an instructional design with worked examples (Kalyuga, Chandler, Touvinen et al., 2001). Worked examples are a type of guided instruction, which includes a problem statement and explanation that details the problem solution. As suggested by Touvinen and Sweller (1999), the selection of instructional strategies depends on the level of learners' prior domain knowledge.

The effects of computer games on training and instruction were claimed to be beneficial in some cases for instruction and training due to some of their characteristics. These effects can be generally divided into five categories: promotion of motivation, enhancement of thinking skills, facilitation of meta-cognition, enhancement of knowledge, and building of attitude. However, in general, there is limited empirical data on the effectiveness of games.

1.2.1. Evaluation Models

Evaluation is the process of determining achievement, significance or value; it is the analysis and comparison of current progress or outcome and prior condition based on the specific goal/objectives or standards (Woolfolk, 2001). While summative evaluation focuses on outcomes and is typically conducted at the end of a program (Kirkpatrick, 1994), another reason to do formative evaluations is to improve the program (Baker, 1998; Baker & Alkin, 1973; Baker & Niemi, 1996). Formative evaluation is normally conducted throughout the program to evaluate the process of a program development. There is limited evidence as to the training effectiveness of games for adults, so the framework of evaluation on the learning results of games needs to be used. Since different goal/objectives and games should be evaluated with different assessment measures, game developers should design appropriate assessment tools to find out if the game really helps learners or trainees achieve the learning goal and objectives, and its efficiency.

For example, this study evaluated a game with training/learning goal to increase learners' problem-solving, so the measures to assess problem-solving including content understanding, problem-solving strategies, and self-regulation were applied (O'Neil, 1999). In this study, we used O'Neil, Baker and Fisher's (2002) framework of formative evaluation in the pilot study. Their model differs from Kirkpatrick's (1994) in the sense that it is formative vs. summative. However, it is similar in that it focuses on what Kirkpatrick would call Level 1: Reaction and Level 2: Learning but not his Level 3: Behavior (i.e. transfer to job) or his Level 4: Results (e.g. cost-effectiveness).

2. Research Design

The research consisted of a pilot study and a main study. The pilot study focused on a formative evaluation to improve the implementation and ensure the success of the main study. The main study that followed the pilot study focused on the impact of the game on problem-solving.

3. Methods

3.1. Formative Evaluation

For the pilot study, the researchers applied the framework of formative evaluation (O'Neil et al., 2002). According to O'Neil et al., the purpose of formative evaluation is to find out the feasibility of a program of educational technology and improve the program by offering information on its implementation and procedure. The study followed a modified version of the O'Neil methodology to conduct a formative evaluation of a game. First, the pilot study was conducted to check the design of assessments for outcome and measurement. Measures were designed and tried out, such as the new-programmed knowledge mapper, the new-designed problem-solving questions, and the computerized trait self-regulation questionnaire. Second, the validity of instructional strategies embedded in the game was checked against the relevant literature. Third, the feasibility review was conducted with students, and finally, necessary revisions were implemented.

3.2. Participants

The participants of the pilot study were three college students from San Jose State University in Northern California, aged from 20–35. The pilot study was conducted after receiving the approval of the university's IRB. All participants were selected to have no experience of playing SafeCracker but have prior experience in playing other computer games.

3.3. SafeCracker: A Puzzle-Solving Game

SafeCracker is a computer puzzle solving game. When playing the game, players try to solve the puzzles in order to break into the safes located in different rooms of a mansion, applying all of the tools and clues hidden in those rooms found by them.

The selection of SafeCracker was based on a study by Wainess and O'Neil (2003). The researchers had conducted an evaluation on the research feasibility of potential 525 video games of three categories: puzzle games, strategy games, and educational games. The appropriate game was then sought among puzzle games, due to their properties and since they provide an appropriate platform for studying games' effectiveness of enhancing problem-solving. A participant in a puzzle-solving game is placed in a specific setting or story background, and tries to reason out possible task procedure and consequences, and a failure to solve a puzzle previously encountered may result in future problems in the game.

3.4. Measures

Content Understanding Measure. Knowledge Mapper was used to measure participants' content understanding of SafeCracker. The knowledge mapper used in previous studies (e.g., Chuang, 2004; Hsieh, 2001; Schacter, Herl, Chung, Dennis, & O'Neil, 1999) was reprogrammed to fit the needs of this study. Participants in the current study were asked to play SafeCracker twice and after each game to create a knowledge map in a computer-based environment. Participants were evaluated on their content understanding, based on their maps, after both the first and the second time of playing SafeCracker. Their maps were scored in real time by comparing the semantic propositions of a participant's knowledge map to those of three expert players maps. For example, if a participant made a proposition such as "Key is used for safe," this proposition would be then compared with all of the propositions in the three expert maps. Participants' content understanding scores were computed by comparing semantic propositions of a participant's knowledge map to those of three expert maps created by five experts of SafeCracker. The following description shows how these outcomes were scored. First, the semantic propositions were calculated based on the semantic propositions, two concepts connected by one link, in each of the three expert maps. Every proposition in a participant's knowledge map was compared against each proposition in the three SafeCracker expert maps. One match was scored as one point. The average score across all three expert maps would be the semantic score of the participant's map.

3.5. Problem-Solving Strategies Measure

In this study, the researcher measured domain specific problem-solving strategies by asking open-ended questions, using modifications of previous researchers' (e.g., Mayer, 2001; Mayer, Dow, & Mayer, 2003; Mayer & Moreno, 1998; Moreno & Mayer, 2000, 2004) assessments of retention and transfer. For example, we adapted Mayer and Moreno's (1998; also, Mayer et al., 2003) approach to measure a participant's retention and transfer by counting the number of predefined major idea units correctly stated by the participant regardless of wording. Two problem-solving strategy questions to measure retention and transfer were used in the pilot study. The retention question was, "Write an explanation of how you solve the puzzles in the rooms." The transfer question was, "List some ways to improve the fun or challenge of the game." An example of an idea unit for the problem-solving retention question is Recognize/compare room features. An example of an idea unit for the transfer question is Add characters to disturb/confuse or help.

3.6. Self-Regulation Questionnaire

The trait self-regulation questionnaire designed by O'Neil and Herl (1998) was used in this study to access participants' degree of self-regulation, one of the components of problem-solving skill. There was sufficient reliability of the self-regulation questionnaire, ranged from 0.89–0.94, reported in previous study (O'Neil & Herl, 1998). The 32-item questionnaire was composed of eight items of each of the four factors: planning, self-checking, self-efficacy, and effort. An example of an item to assess planning ability is, "I determine how to solve a task before I begin." An example of an item to assess self-efficacy is, "I check how well I am doing when I solve a task." Item response choices were *almost never* (1), *sometimes* (2), *often* (3), and *almost always* (4).

3.7. Procedure

The participants of the pilot study (2 males and 1 female, aged from 25 to 30) were selected to have prior computer game experience but have no experience of playing SafeCracker. They were tested individually on a PC. The process was: (a) 2–3 min for introduction, (b) 6–8 min for the self-regulation questionnaire, (c) 8 min for instruction of knowledge mapping, (d) 5 min for game introduction, (e) 20 min for the first game-playing, (f) 5 min for the first knowledge map, (g) 2 min for the first problem-solving strategy questions, (h) 20 min for the second game-playing, (i) 5 min for the second knowledge map drawing, (j) 2 min for the second problem-solving strategy questions, and (k) 2 min for debriefing. The entire session took a maximum of an 80-min period.

3.7.1. Training for SafeCracker

The researcher briefly introduced the puzzle-solving game to the participants, including the gist and mechanics of the computer game. In addition, the participants were told the task: they were trying any way they could think of and using every available resource to solve the puzzles to crack the safes located in the rooms.

3.8. Main Study

Thirty young adults, aged from 20 to 35, participated in the main study after receiving the university's approval of IRB. The volunteer had to be a college or graduate student who had prior experience of playing a computer game but had no experience of playing SafeCracker. Participants were paid for participating. Except for minor adjustments, the same methods and procedure used in the pilot study were applied in the main study.

Based on the pilot study results for the problem-solving strategy test, the transfer question, "Write an explanation of how you solve the puzzles in the rooms" was revised as, "List how you solve the puzzles in the rooms" to reduce the probability that participants would write an essay. In this study, the goal was for participants to write down only key words and phrases, not only to save time but also to make the idea units/propositions clear.

4. Results

4.1. Content Understanding Measure

As shown in Table 1, the mean scores of the 30 participants in pretest and posttest of knowledge mapping were 2.27 (SD = 1.23) and 3.44 (SD = 1.84) respectively. Further, the differences of the pretest and posttest were statistically significant, $t(29) = 4.32,\ p = 0.01$. In addition, the pretest and posttest of knowledge mapping test were significantly correlated ($r = 0.60,\ p = 0.01$).

Table 1
Descriptive statistics of knowledge mapping test.

	N	Mean	Std. Deviation
Pretest	30	2.27	1.23
Posttest	30	3.44	1.84

4.1.1. In Terms of Percent of Knowledge Learned

Table 2 further shows the percentage of expert's knowledge learned by the 30 participants. As may be seen in Table 2 the mean percentage of the knowledge learned by the 30 participants as measured by the pretest and posttest of knowledge mapping were 2%

(SD = 0.01) and 4% (SD = 0.02) respectively. Further, the differences of the pretest and posttest were statistically significant, $t(29) = 4.23,\ p = 0.01$. In addition, the pretest and posttest of knowledge mapping test were significantly correlated ($r = 0.53,\ p = 0.01$).

Table 2
Descriptive statistics of knowledge mapping test (in terms of percent of knowledge learned).

	N	Mean (%)	Std. deviation
Pretest	30	2	0.01
Posttest	30	4	0.02

4.2. Problem-Solving Strategy Measure – Retention

As shown in Table 3, the mean retention scores of pretest and posttest were 2.46 (SD = 1.13) and 4.43 (SD = 1.52) respectively. These differences were statistically significant $t(29)$=12.66, p=0.01. In addition, the pretest and posttest of retention were significantly correlated ($r = 0.83,\ p = 0.01$).

Table 3
Descriptive statistics of problem-solving strategy measure of retention.

	N	Mean	Std. deviation
Pretest	30	2.46 (8%)	1.13
Posttest	30	4.43 (15%)	1.52

However, the overall performance was low. The number of retention idea units of the experts was 29. Thus the participants' mean score of retention in pretest and posttest was 8 and 15% of the average expert knowledge respectively.

4.3. Problem-Solving Strategy Measure – Transfer

Problem-solving strategy was measured with the problem-solving strategy question. As shown in Table 4, the mean scores of transfer of pretest and posttest were 1.70 (SD = 0.98) and 2.76 (SD = 1.47) respectively. These differences were statistically significant, $t(29) = 7.05,\ p = 0.01$. In addition, the pretest and posttest of transfer were significantly correlated ($r = 0.84,\ p = 0.01$).

Table 4
Descriptive statistics of problem-solving strategy question – transfer.

	N	Mean	Std. deviation
Pretest	30	1.70 (7%)	0.98
Posttest	30	2.76 (12%)	1.47

However, the overall performance was low. The number of transfer idea units of the experts was 22. Thus the participants' mean score of transfer in pretest and posttest was 7 and 12% of the average expert knowledge respectively.

4.4. Interjudge Reliability

For each open-ended question of the problem-solving retention and transfer, the interjudge reliability was measured with Cohen's Kappa, using SPSS 11.0 for Windows (2002). One of the game experts, who had made one of the three expert maps was trained to do the categorization and scoring, and then asked to do it independently of the author. If the obtained value is less than 0.70, then the interjudge reliability is not satisfactory; otherwise, if it is greater than 0.70, the interjudge reliability is considered satisfactory (Frey, Botan & Kreps, 2000). Also, the closer the Cohen's Kappa value is to 1.0, the higher the accuracy of the data (Frey et al., 2000). The result showed the percentage of agreement between the two experts in the categorization and scoring for problem-solving retention pretest and posttest was satisfactory (Cohen's Kappa= 0.95). The result showed that Cohen's Kappa value, that is, the percentage of agreement between the two experts in the categorization and scoring for problem-solving transfer pretest and posttest, was satisfactory (Cohen's Kappa= 0.95).

4.5. Trait Self-Regulation Measure

The 30 participants' scores of self-monitoring, planning, self-efficacy and effort total scores of meta-cognition, which were the sum of the scores of self-monitoring and planning, and total scores of motivation, which were the sum of the scores of self-efficacy and effort. The sum of the scores of meta-cognition and motivation stands for their level of trait self-regulation.

As seen in Table 5 the mean item scores of self-monitoring of the 30 participants range from 2.25 to 3.87 with mean score of 3.09 and standard deviation of 0.50; the scores of planning range from 2.25 to 4.00, with mean score of 3.14 and standard deviation of 0.49; the scores of self-efficacy range from 2.25 to 4.00, with mean score of 3.28 and standard deviation of 0.55; and the scores of effort range from 2.12 to 4.00 with mean score of 3.15 and standard deviation of 0.51.

Table 5
Descriptive statistics of trait self-regulation scores.

	N	Minimum	Maximum	Mean	Std. deviation
Self-monitoring	30	2.25	3.87	3.09	0.50
Planning	30	2.25	4.00	3.14	0.49
Self-efficacy	30	2.25	4.00	3.28	0.55
Effort	30	2.12	4.00	3.15	0.51
Meta-cognition	30	4.5	7.87	6.23	0.90
Motivation	30	4.8	7.87	6.43	0.94
Trait self-regulation	30	10.00	15.62	12.67	1.60

4.6. Summary of the Results

According to the results of the pilot problem-solving strategy test, some revisions regarding phrasing of the questions were made for the main study.

The results showed that participants performed better on problem-solving strategy test of retention after the second round of game playing. Further, according to the paired-sample t-test, the difference between pretest and posttest was significant, $t(29) = 12.66$, $p = 0.01$). Therefore, participants' problem-solving strategy of retention was significantly better ($M = 1.52$) after the second round of playing SafeCracker than after the first round of playing the game ($M = 1.13$). Additionally, the percentage of agreement between the two experts in the categorization and scoring for problem-solving retention pretest and posttest was satisfactory (Cohen's Kappa= 0.95).

According to the results, participants performed significantly better on problem-solving strategy test of transfer after the second round of game playing. According to the paired-sample t-test, the difference between pretest and posttest was significant, $t(29) = 7.05$, $p = 0.01$. Therefore, participants' problem-solving strategy of retention was significantly better ($M = 2.76$) after the second round of playing SafeCracker than after the first round of playing the game ($M = 1.70$). In addition, the percentage of agreement between the two experts in the categorization and scoring for the problem-solving transfer pretest and posttest was satisfactory (Cohen's Kappa= 0.95).

5. Discussion

This study evaluated the training effectiveness of a computer game in terms of problem-solving strategy and content understanding. In the pilot study, O'Neil et al.'s (2002) framework of formative evaluation was used to find out the feasibility and implement revisions accordingly. In the main study, CRESST's measure of problem-solving was adopted to evaluate a computer game's training effectiveness. The treatment of 40-min game-playing experience was provided to 30 undergraduate and graduate students (aged

18–35). Except for game-playing mechanics such as the interface, no game strategy was explained to the participants; i.e., it featured pure discovery and experiential learning.

Results gathered from the study provide evidence that, first, participants performed significantly better on the knowledge mapping test after two times of game playing than they did before any game-playing. The average knowledge mapping score of pre-game-playing was 2.27 and the average score of after game-playing was 3.44. This mean difference between the pretest and posttest was statistically significant in favor of the performance after the game-playing. However, participants' performance was from 2 to 4% of the experts' knowledge. It was low as expected due to the lack of instructional strategy in the game.

Second, the participants performed significantly better on problem-solving strategy test of retention after the second time of game playing ($M = 1.52$) than they did after the first time of game playing ($M = 1.13$). Third, the participants performed significantly better on problem-solving strategy test of transfer after the second time of game playing ($M = 2.76$) than they did after the first time of game playing ($M = 1.70$).

Regarding the interjudge reliability, that is, the percentage of agreement between the two experts in the categorization and scoring for problem-solving retention pretest and posttest was high, i.e., 0.95. Also, the percentage of agreement between the two experts in the categorization and scoring for problem-solving transfer pretest and posttest was high, i.e., 0.95.

This study proposed to examine a computer game's training effectiveness in terms of problem-solving. The research question was, "Will participants increase their problem-solving ability after playing a game (i.e. SafeCracker)?" The answer of this research question was positive. However, the learning was low.

Regarding the results of problem-solving assessment, the results were consistent with several previous literatures: first, media has little influence on learning without appropriate instructional strategies (Clark, 1994, 2001, 2003). Since SafeCracker was a commercial off-the-shelf game instead of an educational game, it was not designed based on research-based instructional strategies. In addition, in the study, except for game-playing mechanics such as the interface, no game strategy was explained to the participants; i.e., it featured instructional strategies of discovery and experiential learning. Also, no instructional strategies on game playing and safe cracking, which were the content in this study, were given to the participants before, during, and after game playing.

Further, without effective instructional strategies, inductive/discovery and experiential learning are not effective (Clark, 2003; Mayer, 2004). Also, researchers (e.g., Kalyuga, Chandler, & Sweller, 2001; Mayer, 2004; Van Merriënboer, Clark, & de Croock, 2002) document that there is limited evidence to the effectiveness of these techniques. Furthermore, as pointed out by Mayer (2004), guided discovery learning is more effective than pure discovery learning. In this study, learning by playing SafeCracker is pure discovery learning; no guidance about puzzle solving and safe cracking was given to the participants. Also, the learning was inductive/experiential learning; participants were placed in a specific environment (game environment) to play the game to figure out the answers and strategies by themselves. Therefore, it may be assumed that the low extent of learning resulted from the lack of research-based instructional strategies.

Moreover, for beginners or learners with low background knowledge, research-based instructional strategies, e.g., learning with worked examples, are more effective than learning with problems to be solved (Kalyuga, Chandler, Touvinen et al., 2001; Touvinen & Sweller, 1999). In this study, the learners were beginners; none of the participants had experience of playing SafeCracker before the study, therefore their low performance may be due to the lack of worked examples.

In the main study, results indicated that, first, participants performed significantly better on knowledge mapping test than they did before any game-playing. However, participants' performance was low, only from 2 to 4% of the experts' knowledge.

Second, the participants performed significantly better on problem-solving strategy test of retention after the second session of game playing ($M = 1.52$) than they did after the first time of game playing ($M = 1.13$). Third, the participants performed significantly better on problem-solving strategy test of transfer after the second time of game playing ($M = 2.76$) than they did after the first time of game playing ($M = 1.70$). The interjudge reliability was 0.95 for all measures that involved ratings for the retention and transfer measures.

6. Conclusions

In conclusion, the present study indicated that playing a computer game had significant effects on college or graduate students' outcome performance in a computer-based problem-solving task. After playing the computer puzzle-solving game, students performed better than they did before playing the game. However, the improvement was low.

7. Implications

As pointed out by Ruben (1999), previous researchers have started to notice the potential effects of simulations and games in instruction decades ago, however, the scientific evidence for games' training effectiveness in adults has been lacking (O'Neil et al., 2002). The pilot study applied O'Neil's framework of formative evaluation on a game and the main study gave an example of using our problem-solving framework of evaluation with measures of content understanding, problem-solving strategy, and self-regulation to evaluate the training effectiveness of a computer game.

However, the learning was low in this study. Was it due to the lack of an embedded instructional design? A further research design using a control group with an embedded instructional design may be needed to answer this question. As pointed out by previous researchers it is the research-based instructional strategies or methods in a medium instead of the medium itself that influences learning (e.g., Clark, 1983, 1994, 2001; Wild & Quinn, 1998). Clark suggested that media themselves do not foster learning; they only serve as vehicles, delivering instruction, therefore the instructional strategies and teaching methods designed within a game or a simulation are important. Since neither effective instructional strategies were specifically designed in SafeCracker for the learning objectives nor did the participants receive any learning strategies before the two times of game-playing, the results of this study imply that the lack of appropriate instructional strategies may result in low training effectiveness of a game. Such results may also imply that it is the instructional strategies or methods in a medium instead of the medium itself that may influence learning (e.g., Clark, 1983, 1994, 2001).

Based on the results of this study, we could not conclude whether the instructional method (such as discovery learning in the present study) or the delivery medium (Safe-Cracker) itself has fostered the learning, or whether the computer puzzle-solving game enables an instructional method (learning-by-doing) that fosters learning. It should be noted that learning was low, i.e., from 2 to 4% of the expert knowledge of content understanding, from 8 to 15% of the expert knowledge of retention, and from 7 to 12% of the expert knowledge of transfer.

In a word, an effective instructional medium needs to be designed based on more effective instructional methods. As pointed out by Moreno and Mayer (2004), more research needs to be conducted to identify which instructional methods are most effective for which educational objectives and which learners, based on which an educational game should be designed.

Is learning with instructional guidance better than pure discovery learning? In previous studies, guided discovery has been more effective than pure discovery in helping students' learning and transfer (Mayer, 2004). In addition, learners need examples to connect new information in a learning task with information in their background knowledge, therefore instructional strategies should be provided to them if they fail to come up with appropriate examples (Clark, 1994, 2003). Further, according to the low performance of the participants, it may be inferred that the learning effectiveness of the pure discovery learning was not as effective as guided discovery learning or deductive learning with worked examples. In addition the research conducted by previous researchers (e.g., Ginns, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 2001; Touvinen & Sweller, 1999) showed that the most efficient mode of instruction depends on the level of experience of learners. That is, learning from worked examples is superior to learning from problem-solving at the beginning, but the superiority reverses at the later stage. In this study, participants were selected as novices of SafeCracker, so the low degree of learning may be due to inappropriate instructional strategies. Therefore, more research needs to be conducted to find out if previous instruction of game-playing strategies will improve the learning results, and whether intervention, such as giving participants pictorial navigational aids (e.g., Mayer et al., 2002) or worked examples, improves the training effectiveness of the game (e.g., Touvinen & Sweller, 1999).

According to the previous literature, personalized feedback messages are more effective than non-personalized feedback (Albertson, 1986; Moreno & Mayer, 2000, 2004; Sales, 1993), and "effective performance feedback must be focused on closing the gap between a student's learning goals and the student's current progress" (Clark, 2003, p. 18). However, the feedback in SafeCracker was a non-personalized one, the feedback given to every participant in SafeCracker was the same, therefore, the low performance of the participants in this study may imply the ineffectiveness of non-personalized feedback. However, more research needs to be conducted to find out if personalized feedback designed in a game results in more learning effects than non-personalized feedback/messages does/do.

In summary, although the learning was low, as expected, one outcome of this study is a research environment with reliable and valid measures of problem-solving; that is, content understanding (the knowledge maps), problem-solving strategies (the retention and transfer questions), and self-regulation (the trait self-regulation questionnaire).

References

Abt, C. A. (1970). *Serious games*. New York: Viking.

Albertson, L. M. (1986). Personalized feedback and cognitive achievement in computer-assisted instruction. *Journal of Instructional Psychology*, *13*(2), 55–57.

Baker, E. L. (1998). Formative assessment for learning: New goals, new models. In P. Black & A. Michel (Eds), *Learning from pupil assessment: International comparisons* (CSE Monograph Series in Evaluation, No. 12, pp. 39–48). Los Angeles: University of California, Center for the Study of Evaluation.

Baker, E. L. & Alkin, M. C. (1973). Formative evaluation of instructional development. *AV Communication Review, 21*(4), 389–418. ERIC Document Reproduction Service No. EJ091462.

Baker, E. L. & Niemi, D. (1996). School and program evaluation. In D. Berliner & R. Calfee (Eds), *Handbook of educational psychology* (pp. 926–944). New York: Simon & Schuster Macmillan.

Baker, E. L. & O'Neil, H. F., Jr. (2003). Evaluation and research for technology: Not just playing around. *Evaluation and Program Planning, 26*(2), 169–176.

Boocock, S. & Schild, E. O. (1968). *Simulation games in learning*. Beverly Hills, CA: Sage.

Chuang, S. (2004). The role of search strategies and feedback on a computer-based collaborative problem-solving task. Unpublished doctoral dissertation. University of Southern California.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–30.

Clark, R. E. (Ed.) (2001). *Learning from media: arguments, analysis, & evidence*. Greenwich, CT: Information Age.

Clark, R. E. (2003). What works in distance learning: Instructional strategies. In H. F. O'Neil, Jr (Ed.), *What works in distance learning*. Los Angeles: University of Southern California; UCLA/National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Coleman, J. S. (1969). Games as vehicles for social theory. *American Behavioral Scientist, 12*, 2–6.

Dugdale, S., LeGare, O., Mathews, J. I., & Ju, M. (1998). Mathematical problem-solving and computers: A study of learner-initiated application of technology in a general problem-solving context. *Journal of Research on Computing in Education, 30*(3), 239–253.

Frey, L. R., Botan, C. H., & Kreps, G. L. (2000). *Investigating communication: An introduction to research methods*. Needham Heights, MA: Allyn & Bacon.

Gamson, W. A. (1969). *Simulated society*. New York: Free Press.

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*(2), 229–251.

Gredler, M. E. (1996). Educational games and simulations: A technology in search of a (research) paradigm. In D. Jonassen (Ed.). *Handbook of Research for Educational Communications and Technology* (pp. 521–540). New York: Macmillan.

Greenblat, C. S. & Duke, R. D. (1975). *Gaming-simulation*. New York: John Wiley.

Hsieh, I. (2001). *Types of feedback in a computer-based collaborative problem-solving Group Task*. Unpublished doctoral dissertation. University of Southern California.

Kalyuga, S., Chandler, P., Touvinen, J., & Sweller, J. (2001). When problem-solving is superior to worked examples. *Journal of Educational Psychology, 93*(3), 579–588.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*(1), 5–23.

Kirkpatrick, D. L. (1994). *Evaluation training programs: The four levels*. San Francisco, CA: Berrett-Koehler Publishers.

Martin, A. (2000). The design and evaluation of a simulation/game for teaching information systems development. *Simulation & Gaming, 31*(4), 445–463.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem-solving. *Computer in Human Behavior, 18*, 623–632.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19.

Mayer, R. E., Dow, G. T., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology, 95*(4), 806–813.

Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171–185.

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social Cues in Multimedia Learning: Role of Speaker's Voice. *Journal of Educational Psychology, 95*(2), 419–425.

Moreno, R. & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92*(4), 724–733.

Moreno, R. & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173.

O'Neil, H. F., Jr. (1999). Perspectives on computer-based performance assessment of problem-solving. *Computers in Human Behavior, 15*, 225–268.

O'Neil, H. F., Jr & Andrews, D. (Eds) (2000). *Aircrew training and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr, Baker, E. L., & Fisher, J. Y.-C. (2002). A formative evaluation of ICT games. Los Angeles: University of Southern California.

O'Neil, H. F., Jr & Fisher, J. Y.-C. (2004). A technology to support leader development: Computer games. In V. D. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leadership development for transforming organization: Growing leaders for tomorrow*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr & Herl, H. E. (1998). *Reliability and validity of a trait measure of self-regulation*. Los Angeles, University if California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Pfeiffer, J. W. & Jones, J. E. (1969–1977). A handbook of structured experiences for human relations training (Vols. 1–6). La Jolla, CA: University Associates Press.

Quinn, C. N. (1991). Computers for cognitive research: A HyperCard adventure game. *Behavior Research Methods, Instruments, & Computers, 23*(2) 237–246.

Quinn, C. N. (1996). Designing an instructional game: Reflections on "Quest for independence." *Education and Information Technologies, 1*(3& 4), 251–269.

Ruben, B. D. (1978). *Human communication handbook: Simulations and games* (Vol. 2). New York: Hayden.

Ruben, B. D. (1999, December). Simulations, Games, and experience-based learning: The quest for a new paradigm for teaching and learning. *Simulation & Gaming, 30*(4), 498–505.

Ruben, B. D. & Budd, R. W. (1975). *Human communication handbook: Simulation & games*. New York: Hayden.

Sales, G. C. (1993). Adapted and adaptive feedback in technology-based instruction. In J. V. Dempsey & G. C. Sales (Eds), *Interactive instruction and feedback* (pp. 159–176). Englewood, NJ: Educational Technology publications.

Schacter, J., Herl, H. E., Chung, G., Dennis, R. A., & O'Neil, H. F., Jr. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior, 13*, 403–418.

Stolk, D., Alesandrian, D., Gros, B., & Paggio, R. (2001). Gaming and multimedia applications for environmental crisis management training. *Computers in Human Behavior, 17*(5–6) 627–642.

Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*(4),295–312.

Tansey, P. J. & Unwin, D. (1969). *Simulation and Gaming in education*. London: Methuen.

Touvinen, J. E. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning. *Journal of Educational Psychology, 91*(2), 334–341.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-Model. *Educational Technology Research & Development, 50*(2), 39–64.

Wainess, R. & O'Neil, H. F., Jr. (2003, August). *Feasibility study: Video game research platform*. Manuscript: University of Southern California.

Wild, M. & Quinn, C. (1998). Implications of educational theory for the design of instructional multimedia. *British Journal of Educational Technology, 29*(1) 73–82.

Woolfolk, A. E. (2001). *Educational Psychology* (8th ed.). Needham Heights, MA: Allyn and Bacon.

FOUNDATIONS FOR SOFTWARE SUPPORT OF INSTRUCTION IN GAME CONTEXTS

Allen Munro
University of Southern California

Abstract

Experimental evidence indicates that active instruction is necessary for learning to occur. If an instructional object is to teach in the context of a game, it will require certain services of the game engine, such as highlighting a game object on request, or inserting a game state for instructional purposes. Eight learning objectives that can be pursued in game contexts are described, together with the services a game should provide for instruction. The architecture of iRides and the application programming interface (API) of its simulation engine are offered as a model for a universal API for utilizing games in instruction.

Many game-makers and some adult learning authorities seem to believe that simply playing a game will result in learning. For instance, Prensky (2001) claimed that

> Business people are slowly "getting it." Schools "get it" here and there. But the U.S. military "gets it" big time. The military has embraced Digital Game-Based Learning with all the fervor of true believers. Why? Because it *works* for them. And trust me, the guys in charge of training at the Pentagon are a very sharp group. They have seen and evaluated *everything* (p. 295).

Whether these purported evaluations assessed actual learning (as opposed to enjoyment of the activity, say) is very much in question. When Hays (2005) undertook an extensive review of the effectiveness of instructional games for the Naval Air Warfare Center, he found that of 274 documents on games for learning, only 105 articles could be used at all. A few dealt only with design; some used simulations that were not games; 36 had significant methodological problems and 77(!) "only provided the author's opinion on the potential of instructional games" (p. 5). Of the 105 usable articles, only 48 provided

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

empirical data on the effectiveness of games in instruction. Hays states that "unfortunately, many decisions about whether to use games for instruction are based on unfounded assumptions about the ability of games to provide effective and efficient instruction" (p. 9).

In order to play an electronic game well eventually, novice players must learn. How could there be improvements in game performance without learning? For purposes of education and training, however, the primary desired outcome is not that the game user will become a better game player. The intended educational or training result is that the learner will acquire knowledge that transfers to other, non-game environments.

Repeatedly playing a game can clearly be an effective way of learning to play that game. But what else can be learned in a game? What can be learned in a game context that will transfer to other aspects of life, such as work performance? And how can games best support adult learning needs? Could there be a universal architecture for utilizing games in instruction and learning? Recent work conducted at USC's Center for Cognitive Technology, in collaboration with scholars at UCLA's Center for Research on Evaluation, Standards, and Student Testing, throws light on the answers to some of these questions.

1. Games and Simulations

There is a long history of simulations being employed to promote learning, with a special emphasis on computer-based simulations[1] (Aldrich, 2004; de Jong & van Joolingen, 1998; Hollan, Hutchins, & Weitzman, 1984; Munro, Johnson, Pizzini, Surmon, Towne, & Wogulis, 1997; Rieber, 1992;). In this chapter, the term *simulation* is used to refer to an interactive experience that represents a class of real-world experiences or phenomena. By *interactive*, I mean that the learner can take actions that affect subsequent events in the simulation. For example, reducing the throttle on a simulated aircraft control system may result in a simulated stall, possibly followed by a dive and a crash. Simulations are to be distinguished from mere *animations*, which present a sequence of events that is invariant. In an animation, the user cannot take actions that will affect subsequent events. Whether a presented animation is a Quicktime movie or an invariant Flash sequence, it is not a simulation if it can only show a series of events in one order.

In what way are interactive games related to simulations? Many, perhaps most, electronic games simulate some aspect of the real world. First-person shooter games such as Doom and all its sophisticated competitors and descendents simulate warfare using projectile weapons, for example. (A few games, such as Tetris™, do not seem very close to any real life activity.) For practical purposes, the electronic games that are most likely to have applications in education and training will be those that simulate some aspect of reality. Facer (2004) claims that simulations are sometimes referred to as "games without

[1] In the field of medical training, the term *simulation* most often refers to a training experience in which human actors present standard symptoms to the medical trainee who is learning to diagnose disease and injury. This is not the meaning used in this chapter.

pleasure" (p. 1), but one could as well say that *games are simulations that are enjoyable*. Hays (2005) reports that the following characteristics of games have been noted by Bright and Harvey (1984) and by Leemkuil, de Jong, and Ootes (2000): voluntary play, pursuit of a goal state, constraints and/or rules, and competition—which may include competition with oneself. It may be that some or all of these characteristics are responsible for the enjoyable quality of games. More rigorously, a simulation may be considered a game if it incorporates these characteristics – learners choose to use it, they pursue a goal in the simulation, there are constraints to be met, and the users can compete, either trying to best others or to best their earlier performances. There are games that are not simulations (such as Tetris™ or WebSudoku™) and there are simulations that are not games. Of the games that can be used for instruction and training about real world subjects, however, most have a number of structural features in common with simulations. In both, the user has considerable autonomy, is initially unaware of all the likely outcomes of actions, and learns about outcomes as a result of making decisions and carrying out actions.

In a summary of the 2004 Serious Games Summit, participants are described as admitting that initial expectations for automatic learning from games have not been realized (Michael & Chen, 2004).

> In the panel "How Can Games Shape Future Behaviors?" James Paul Gee, author of *What Video Games have to Teach Us about Learning and Literacy*, talked about the amount of "transfer" (of ideas, skills, etc.) that happens between games and players. The views on the subject have shifted over the years, from "transfer is easy" to "transfer is impossible" to the current position of "transfer is possible"–but it isn't cheap, quick or easy (p. 1).

It is worth remembering that commercial video games and computer games are not developed primarily with the goal of teaching facts or skills. Instead, their purpose is to provide entertainment. Would a game provide good value if it could be quickly mastered? How many game buyers would be pleased to find that they saw everything and attained maximum performance in short order? Entertainment value requires that the goals of the game not be achieved too easily, so that the consumer's expenditure on the game provides the value of many hours of diversion. Consider a puzzle-solving exploration game such as Myst™. The player moves among locations on a virtual island, finding clues, searching for controls, and trying combinations of actions. Correct action combinations lead to new locations and/or new clues. The process is sometimes laborious, but the user learns a good deal about the locations of objects on the island and how to create certain effects using controls. This knowledge could be taught much more directly and more quickly, but the entertainment value of the game would be much reduced.

In contrast, a normal goal of teaching and learning institutions is to impart knowledge or skill as rapidly as possible. In the real world of adult training for job performance, employers want to provide effective learning experiences in as little time as possible. The US Department of Defense, arguably the largest training and learning organization in the world, is continuously looking for ways to train better at lower cost. The most effective means of lowering costs is by reducing time spent in training. Even in schools, budget constraints weigh in favor of approaches that make it possible to meet learning objectives in shorter, rather than longer, periods of time. So, on the face of it, the design of commercial games has goals incompatible with efficient teaching and learning.

2. Learning in Guided and Unguided Contexts

There has been a frequently recurring belief among a portion of the educational research community (e.g., Brunner, 1961; Jonassen, 1991; Jonassen, Howland, Moore, & Marra, 2002; Papert, 1980) that certain approaches to learning that take a good deal of time are nonetheless worthwhile, because the acquired knowledge is better understood or will be more productively utilized by the learners. These proponents of *discovery learning* – now reincarnated as *constructivism* (Jonassen, 1991; Jonassen et al., 2002) – believe that the process of acquiring knowledge in much the same way that it was first discovered will help students to understand and make effective use of that knowledge. The efficiencies that result from guided learning are less important than the quality of learning in discovery environments, in their view.

Mayer (2004) and Sweller (1999, 2004) have pointed out that discovery-learning environments impose a heavy cognitive load on learners, and would therefore be expected to interfere with learning of subject matter. In fact, numerous studies have either failed to find a significant advantage for discovery learning or have found that the discovery-learning treatment disadvantaged students. For example, Clark (1989) reported that a review of many studies showed that lower aptitude students actually have posttest scores that are *lower* than their pretest scores after experiencing unguided or discovery learning. It could be that, without guidance, students carried out series of actions that produced results they found confusing, causing them to erroneously doubt or modify previously correct portions of their understanding of the subject matter.

Kirschner, Sweller, and Clark, in a 2006 analytical survey, show that research on human cognition should lead to an expectation that relatively unguided instruction should be inferior (at least for all but the most able students) to well-designed guided instruction. They also point out that controlled research studies on the topic show, on balance, that no reliable measurable advantages for unguided instruction can be found.

What does this result imply for learning in game contexts (and, more generally, for learning in any interactive simulation context)? Clearly, it means that conventional game experiences alone are unlikely to promote effective and efficient learning. Despite the beliefs of the discovery-learning/constructivist community, there is little evidence to support any claims of learning superiority in the data, as shown by Kirschner et al. (2006). Guidance and instruction can be used to ensure that students are rapidly exposed to crucial concepts, important features, effective procedures, and other important information about the subject matter, rather than being left to ‘discover’ these facts for themselves. Active guidance results in better learning in less time.

3. Evidence for Learning from Games

The questions of whether students can learn from games and under what conditions they are most likely to do so have been addressed in a number of studies. Many of these studies have, in turn, been evaluated and analyzed from different perspectives in two recent papers. O'Neil, Wainess, and Baker (2005) look at the classification of learning outcomes in multi-player computer-based games. Their approach looks at learning outcomes very broadly, considering a variety of different approaches to measuring learning from games and classifying these approaches using two different systems, Kirkpatrick's (1994) *levels of evaluation* and the CRESST model of learning (Baker & Mayer, 1999). Of the thousands of published articles on games, only 19 met the empirical research criteria that O'Neil et al. (2005) set. Most of these studies sought to measure learning of problem-solving knowledge and skills in particular contexts.

Hays (2005) looked at studies with single-player games, and so had a larger number of studies to examine. In addition, 26 review articles were surveyed in this paper. Among the original empirical studies, only a few sought to study the effects of instruction versus non-instruction in game contexts. Leutner (1993) compared two types of instructional support with no guidance in a role-playing simulation. The results indicated that instructional support is essential if the goal of the game is to support the learning of content. De Jong and von Joolingen (1998) found that successful discovery learning required instructional support for the learner. Hays concludes that some games have been shown to have instructional value, if they are designed to meet specific instructional objectives, but that this is rarely the case. In fact, the interdisciplinary design teams that create learning games never included instructional designers or developers.

Both of these analytical reviews of research on games and learning commented on the surprisingly limited results. O'Neil et al. (2005) concluded that "the empirical evidence for the effectiveness of games as learning environments is scant" (p. 468). They state that their position is "that games themselves are not sufficient for learning, but there are elements in games that can be activated within an instructional context that may enhance the learning process" (p. 465).

Is the importance of guidance and explicit instruction recognized in conventional simulator training, such as pilot training by the armed forces? Traditionally, yes. Although most military simulators do *not* have automated guidance or instructional features built in, the simulators are typically used in the presence of human instructors. In some cases, more instructors and knowledgeable technicians are present than learners. These instructors often provide introductory demonstrations and other orienting experiences. When they judge that the cognitive task load imposed by the simulation is not too great, they may recommend actions, point out salient features, or ask pointed questions about the simulated environment in order to guide the student's understanding of the simulated task. In almost every case, instructors participate in a post-simulation debriefing for the student, sometimes including a simulation replay with commentary and discussion.

The border between games and interactive simulations is not always sharp. Are military flight simulator-trainers games? Some are very engaging, even though many of the entertainment features of commercial games are absent. For example, because real air-to-air engagements are likely to take place over such large distances, there is ordinarily no requirement for the visually spectacular and noisy explosions that would be part of a commercial fighter game. Whether a simulation is also a game or not, instructional guidance is very likely to enhance learning in that context.

It is even a question whether first-person shooters like *America's Army* (http://www.americasarmy.com/) are learning games. Like other first-person role-playing games, *America's Army* presents threats, a variety of means for responding to those threats, and generated outcomes that depend on the choices made by the player. It is a game, but is it an effective or efficient learning experience? Judging from the results reported in the studies by Clark (1989), Kirschner et al. (2006), Mayer (2004), and Sweller (2004), it seems very unlikely that such a game would be as effective as one that contains explicit instruction. In fact, Belanich, Sibley, and Orvis (2004) assessed learning in the context of this game. They found, according to Hays (2005), that "the players only learned procedures, experiences, and facts about the game. There is no indication that any of the learned information is relevant to real-world Army requirements" (p. 40).

Would it still be a game if there were demonstrations with commentary on appropriate tactics to use in different contexts? Would it be a game if play were sometimes interrupted to point out relevant features or to remind the player of the rules that apply in the situation? What makes a simulation entertaining enough to be a game is not the topic of this chapter. Our goal as educators and educational researchers is to teach or train as effectively as possible. If doing that makes a simulation no longer fun enough to be a game, so be it. It would make sense to eschew guidance only if explicit instruction is so demotivating that learners will refuse to use the simulation, or will use it so little that they would be better off in an unguided game. Yet the results of the studies referred to above provide no evidence of such an effect occurring due to the use of active instruction.

4. Ways to Guide and Teach in Game Contexts

Since the evidence shows there is reason to provide instruction and guidance in simulation or game environments, it now makes sense to ask how this should be accomplished. In what ways can guidance and instruction be added to games in order to support learning and transfer of knowledge to the real world? Fortunately, there is a body of research and development in the area of simulation-based learning that can be applied to learning in game contexts. For the past 17 years, my colleagues and I at USC have been working on the development of generic authoring tools and delivery systems for providing education and training in simulation contexts. The major issues we have encountered for simulation-based learning are likely to apply as well to game-based learning, which, as I have suggested, can be thought of as a near-subset of simulation-based learning. Before considering these issues, however, let us consider what kinds of knowledge are best conveyed in game contexts.

4.1. How Can Content Be Explicitly Taught in Game Contexts?

Consider eight possible types of learning objectives for game contexts, listed in List 1. These are not exhaustive, but they illustrate a variety of learning objectives that are likely to be attempted in game contexts. We can ask what services an instructional component would need to request of a simulation in order to provide effective instruction for each of these types of learning objectives in the simulation context.

List 1
Eight types of learning objectives in game contexts.

1. Learning to recognize surface game cues
2. Learning to recognize the meanings of objects and events in the real world that are simulated in the game; learning typologies, instances of a concept, etc.
3. Learning causal relationships qualitatively
4. Learning causal relationships quantitatively
5. Learning about relative effects, nested effects, feedback effects and other complex behavioral relationships
6. Learning action sequences, including conditional actions
7. Learning goal hierarchies
8. Learning to solve novel problems in the game context by applying strategies

These range from learning objectives that apply only to game interactions, to learning abstract principles and techniques that can be applied to problem-solving in related real-world contexts. For many of these learning objective types, one or more of the *effects* noted by Sweller (in press) can be applied to select appropriate instructional techniques. The techniques selected for achieving these objectives can be implemented using a set of low-level action types, such as presenting text, highlighting a simulation object, and so on. Some of these low-level instructional actions (such as highlighting a simulation object) will consist of a service request from an instructional component to the game. A set of all the common types of service requests to games by instructional components could be the basis for a universal interface between games and instruction control objects. I now consider each of these eight example learning objectives for game-based learning contexts, and ask what low-level services will be required of a game in order to effectively achieve learning through instruction and guidance in the game context.

1. *Learning surface game cues*. Many games are mastered by practicing—learning to quickly recognize each graphical or auditory cue and learning the appropriate action to take in response to that cue.

Although one might hope that games designed to promote learning for application to real problems would focus on teaching transferable skills in recognition, reaction, carrying out procedures, and making decisions, there is no a priori reason to expect that learning surface game cues would have wide application outside of games. Although this type of learning is not a conventional educational goal, it is sometimes a goal of training software, such as the ship recognition training developed by Hollan and his colleagues in the early 1980s. That program presented the silhouettes of naval vessels of both friendly and potentially hostile nations as they might be observed on the horizon at sea, in random order, requiring that the learner determine the class of each ship.

This sort of recognition training has a game-like quality, in that scores can be run up and compared, students can compete for time, and so on. Even here, however, there is scope for useful training guidance. Features that distinguish one class from another can be explicitly pointed out and described. When a student makes a classification error, the student can be shown the differences between the features of the class chosen and those of the class of ship displayed. Explicit guidance thus can play a crucial role even in simple recognition training games.

Sweller's *split-attention effect* predicts that when the same perceptual channel is used to present two types of information on the same topic (such as text and graphics that require each other for interpretation), cognitive load will increase and learning will decrease. Rather than use text labels and explanatory presentations, therefore, it would be better to exploit Sweller's *modality effect*, presenting names and explanations with audio, while highlighting the game objects under discussion. Low-level game services utilized:

highlight object
unhighlight object

(Presentation of instructional audio is the responsibility of a component of the instructional system, not of the game. Although both components may share an execution environment, the separation of modules promotes the maintainability of both software and content.)

2. *Learning to recognize the meanings of objects and events and learning to respond*. In many first-person shooter games, the player must learn to classify threats so as to choose the appropriate weapon for use against that threat. In law enforcement, *Shoot/Don't Shoot* exercises are frequently used to give the learner opportunities to practice rapidly evaluating the threat posed by possible assailants. Other, more reflective games of strategy or problem-solving also require that salient cues be attended to, so that users will select appropriate responses or take necessary initiatives in the game.

For this type of learning, explicit instruction and guidance can play a major role. First students can be presented with demonstrations of correct responses to different events. The instructional system can provide explicit information about salient features of the example. Possible errors of classification can often be detected and remediated. An instructional monitor can also notice when the wrong action set (or tactic) is used after a correct classification. This, too, can provide an opportunity to teach general principles in the context of specific examples. As with the simpler recognition objective in (1), the application of Sweller's *modality effect*, especially for novices (and the avoidance of the *split-attention effect*) should help to promote the attainment of this learning objective.

Low-level game services utilized:

highlight object
unhighlight object

report occurrence – informs instruction when a condition of interest has occurred. This is used to detect game states of interest.
register occurrence interest – dynamically adds an occurrence descriptor to the list of conditions that should be reported to the instructional component whenever they occur.

3. *Acquiring a qualitative understanding of causal relationships* and
4. *Acquiring a quantitative understanding of causal relationships*. Games may provide a natural context for conveying to students a qualitative understanding of causal relationships. There is a long history of research on the use of qualitative models and scaffolding in teaching basic concepts in electricity (White, 1984; Gutwill, Frederiksen, & White, 1999).

Effective game-based training about causal relationships among events should usually provide one or more demonstrations of the effect. Sweller (in press) identifies this instructional tactic as one that exploits the *worked example effect*. In order to conduct an effective demonstration, an instructional component must be able to draw the attention of the learner to objects in the game, and it must be able to make the game behave. This can be done by setting the value of some variable in the game, or by imitating a set of user actions. In the latter case, there may be visual cues, such as a moving mouse pointer, to convey the idea that the demonstration is being carried out by an invisible instructor, acting as a second user of the game.

Low-level game services utilized:

highlight object
unhighlight object
set game internal value – the instructional object asks a game to set an attribute or variable to a specific value
emulate user action – instruction asks the simulation to emulate a user action, resulting in a chain of simulation effects

5. *Learning about complex behavioral relationships*. Games, like simulations, may be well-suited to helping users learn about complex relationships that may hold between objects and events. Complex concepts like relative motion, additive and subtractive effects, and positive and negative feedback can be illustrated interactively in a well-designed game context. The same types of instructional primitives that are employed for learning objective types (3) and (4) would be utilized for this type of learning objective, as well.

In addition, however, the game itself may need to be pre-programmed with special modes to help convey complex concepts. For example, a game that includes multiple freely moving entities with a realistic through-the-windshield view may sometimes be able to clarify a situation by also presenting a schematic overhead view of the evolving situation. The game would have the capability of presenting this view, and would do so when the instructional system made a particular *set game internal value* request. In some cases, the use of two such views would illustrate Sweller's *modality effect*. When incorrectly applied, it might constitute an illustration of the *redundancy effect*, in which an additional source of available information increases the cognitive load and reduces learning. Another approach to coping with the cognitive load imposed by understanding complex interactions is to explain portions (or *elements*) of a complex system in isolation. Sweller (in press) calls this the *isolated interacting elements effect*. This approach is most likely to be feasible when the game was designed for pedagogical purposes and so can support views and simple interactions in which only a subset of the full game elements are involved. An instructional component can utilize this feature of the game by setting game internal values.

Low-level game services utilized:

highlight object
unhighlight object
set game internal value
emulate user action

6. *Learning a sequence of actions*. Procedures for accomplishing tasks can be learned as sets of actions and observations in a game context. Such sequences may be ordered, unordered, or partially ordered. An example of a partially ordered sequence is "A, B, C, and D in any order, except that step B must be carried out before step D."

Although many first-person role-playing games are mastered by learning a sequence of actions, including conditional actions, it would be very inefficient to try to get students to learn a real-world task by having them blunder through it over and over again, seeking the correct sequence. Games like Myst™ include puzzles that can only be solved by repeatedly trying different combinations of lever settings or valve positions. This is a fine way to spend a lot of time, but it is not a very efficient way to learn how to carry out a task.

Instead, learners need to be given a structured sequence of actions, together with the steps needed to carry out each of them. Explanations should be given for constraints in the order in which a set of actions must be carried out. Students should be given the opportunity to practice, and instructional feedback should be presented on how to improve performance.

Sweller's (in press) *guidance fading effect* can be an effective approach to teaching about procedures in a game context, as described by Van Merriënboer, Schuurman, de Croock, and Paas (2002). First, the procedure is demonstrated one or more times for the student. Then the student is guided through a substantial portion of the procedure and then asked to complete the remaining steps. Guidance can be progressively weened as the student learns. When a student gets stuck, it may be useful to demonstrate how the task is done, while providing an explanation. Then a new problem can be presented. This process can continue until the student exhibits a desired level of proficiency.

Low-level game services utilized:

highlight object
unhighlight object
set game internal value
emulate user action

register occurrence interest
report occurrence

7. *Learning goal hierarchies*. Students or trainees can learn how to flexibly carry out complex tasks that can be characterized as a set of nested goals. The terminal nodes of such a tree are actions and observations that define sub-goals. Learning complexly structured procedures specified by goal hierarchies involves many of the same activities described for simple procedures in (6). An extension of Sweller's *isolated interacting elements effect* is to separate sub-goal elements, teaching them in isolation.

When each new sub-goal is to be demonstrated, explained and practiced, the game will need to be placed into a state in which necessary prerequisite goals have already been met. Although this could be done with a sequence of *set game internal value* requests, some games support the installation of a complete block of values in a single service request. This notion is conveyed by the use of the instructional service *set game internal value(s)*.

Low-level game services utilized:

highlight object
unhighlight object
set game internal value(s)
emulate user action
register occurrence interest
report occurrence

8. *Learning to solve novel problems in game contexts by applying knowledge and strategies*. A deep understanding of a subject matter can be utilized to define new goal hierarchies for carrying out new tasks, or for accomplishing standard tasks in novel contexts, such as tasks with defective initial states.

Presenting worked examples to the learner will facilitate this type of learning. Instruction can guide the student by pointing out how the particulars of a given problem map to the roles referenced in a strategy, tactic, or other problem-solving system. In this type of training, the instructional text may provide an explicit typology of problem types in the simulated domain. The *set game internal value(s)* service would be used to present problems in the game context. Learning aids may be made available to help students appropriately classify novel problems and choose potentially useful solutions. Exactly the same set of service request types can be utilized to carry out this type of advanced training in a game context, as well.

Low-level game services utilized:

highlight object
unhighlight object
set game internal value(s)
emulate user action
register occurrence interest
report occurrence

A naïve approach to learning in simulation contexts would be to expect students to acquire all of these types of learning objectives simply as a result of being exposed to the simulation. In fact, some proponents of games for learning (Aldrich, 2004; Prensky, 2001) have at least implicitly made the claim, "If you simulate it, they will learn!" Perhaps so, but without any instruction, it can take a long time.

Instructional developers tend to think in terms of "pages." Game-makers ordinarily do not. In ordinary web-based instructional systems, a learner is presented with a webpage with text and graphics. The users sometimes are allowed to choose which page to view next; in other cases, the answer to a question about the just-presented content determines what page will be presented. In contrast, in graphical games and simulations, the user is engaged with an environment that represents some part of the world – or some part of a possible world. Interacting with that world is not seen as bringing up a new page, but rather of moving about in that world, taking actions that have effects, observing those effects, and then planning and carrying out new actions. To be effective in game contexts, the authors of instruction often avoid thinking in terms of a sequence of pages that will be presented to teach the learner. Instead, they think about the opportunities for learning in the game context. There are many such opportunities. The simplest ones are due to the graphical nature of games. The positions of objects can be pointed out to learners, including relative proximities, lines of sight, and so on, and the consequences of these relationships can be explained. When a procedure is being learned, an error is an opportunity to show the undesired consequences of the erroneous action. An instructional system can detect anticipated error types by using the *register an occurrence interest* service. When the game notifies the instruction system that one of the registered conditions has occurred, that naturally occurring error can be used as an opportunity to instruct.

The use of games and simulation contexts is *not*, by itself, a sufficient condition for learning. Self's (1995) study examining the effectiveness of computer-based discovery environments found that only rare, highly motivated, intelligent learners with high tolerances for ambiguity and uncertainty will ordinarily learn spontaneously in a discovery environment. Most students need guidance, instruction, hints, and practice with instructional feedback.

5. A Universal Application Programming Interface for Interfacing Games to Interactive Instructional Systems

Munro, Surmon, and Pizzini (2006) identified a possible universal set of services that computer-based simulations might provide for collaborating instructional control objects. The core elements in that set of low-level services are identical to those identified above as being necessary for effective interactive instruction in a game context. There are many advantages to not entangling the code for a particular game or simulation with the code for teaching in the context of that game or simulation. The most important of these is that the encapsulation of function promotes code maintainability. If either the game is to be modified or the doctrine about how the game should be played must change, the necessary modifications will be much more expensive in an entangled system. A second reason is that a well-constructed general-purpose instruction engine can be built just one time, and it can interpret specific instructional scripts to control guidance an instruction in many different simulations or game contexts. This has the advantage of turning over specific instructional development to subject matter experts (SMEs) and instructional designers, without having to make use of systems programmers every time anything has to be changed about a game. Not only will this reduce costs in a training development process, but it will also greatly reduce the chances that programming errors will result in thread lockups and other software glitches when instruction is delivered in the context of a running game or simulation. For such an approach to work, however, each game that is to collaborate with the independent instructional engine must provide a set of services that can be called upon by that engine.

By using its own text (or text-to-speech, recorded speech, video, etc.) interfaces to present language content and calling upon the game for instructional services such as those listed above, complete sequences of introductory, remedial, practice, and assessment vignettes can be delivered in the context of a game. If games offer these services through a consistent API (application programming interface), then instructional engines can use that API to interact with those games for the benefit of the learner.

The successful implementation of the iRides system (Munro, 2003; Munro, Surmon, Johnson, Pizzini, & Walker, 1999) demonstrates the feasibility of such an approach, although in the context of interactive graphical simulations, rather than games. In the iRides architecture, there are two major components, an instructional system with its presentation interfaces, and a collaborating simulation object. The iRides instructional engine interprets instructional specifications in an XML file and presents instruction to the student while requesting services from the simulation engine. A simulation engine reads appearance and behavior specifications from a different type of text file. These specifications determine the look and feel of iRides simulations. In the following discussion, the design of iRides is presented as an example of an integration of an instruction system with another interactive component, a simulation. The architecture may be sufficiently robust to also support the integration of training with appropriately designed game engines, as well.

The suggestion of a universal architecture for delivering instruction in the context of games can only apply to the design of future games. At the time of game design, the game engine must be designed and programmed to include the basic services that instructional systems could make use of. Existing games could be modified to work with such instruction systems only by being, at least in part, reprogrammed to offer the specified services.

Higher level instructional specifications. Patterns of service requests and presentations are utilized repeatedly in game-based instruction. For example, when the function or nature of an object in a game is introduced to a learner, the instructional system often asks the learner to touch (click on) the object as a way of determining that the learner knows what is being discussed, and it can be used to pace the instructional process. The instruction system does this by carrying out the following sequence of actions:

- It asks the game to begin ignoring actions by the user. (This makes it possible for the student to indicate an object without causing the game response that would normally occur when the object is clicked.)
- It presents text (or voice, etc.) introducing the object.
- It waits for the student to click on something. If too much time goes by, it asks the game to highlight the object and tells the learner that the object to click is the one that is highlighted.
- If the student clicks on an incorrect object, it offers remedial feedback. This may include a hint, followed by another invitation to indicate which object is under discussion. After a specified number (sometimes only one) of erroneous clicks, it has the game highlight the object and asks the learner to click on it.
- After the correct object is indicated by the student, the instruction system tells the simulation to turn off the highlighting (if highlighting was used), and it asks the game to resume responding to the user's game actions.

This standard set of low-level activities can be gathered into a simple instructional template, which can be repeatedly used with different object and textual content during instruction. In iRides, instructional templates can be built of instructional primitives, like *highlight object, present text, set internal value*, and so on. The iRides template that consists of the primitives described above is called the *Find* template, and it is used to ascertain that students can identify objects in a game or simulation by pointing to them. Templates can also reference other templates, so higher level instructional patterns can be built up of lower-level templates, together with primitive instructional actions.

Implementing ways to guide and teach in simulations. Consider the eight types of learning objectives for game contexts that were introduced in List 1 above. For each of these types of instruction, one or more teaching or training strategies can be devised that can be re-used for almost any instance of that type. As an example, consider learning objective type (6), learning action sequences. Clark (2004) proposes a consistent, evidence-based approach to teaching procedures. Could his design be implemented as a template in iRides?

An approach for designing learning-to-do instruction. Whether teaching someone how to carry out a task in the context of a game or in a less entertaining simulation context, research studies suggest a principled approach may be very effective. Clark's (2004) *GEL* system for Guided Experiential Learning Design proposes that effective training on how to carry out tasks should always have five components, which are presented in a fixed order:

1. Goals – A clear and precise description of what trainees will be able to do after training is complete;
2. Reasons – A description of the benefits and risks of the training;
3. Overview – The background information needed;
4. Demonstration of the procedure;
5. Practice and feedback.

Clearly the last two steps of a GEL training system will often make use of some kind of simulation or other practice environment. In addition, however, it may often be useful to introduce background information in the context of the game or simulation. The interaction of elements in the game (which presumably emulate the interaction of real-world entities) can be illustrated graphically and/or with animation, as well as being explained propositionally.

Authoring instruction for learning-to-do. Learning to carry out a task (the third and fourth entries in List 1) is an important use of games for learning. My colleagues and I at USC have constructed several authoring applications that can be used to build lessons on how to carry out tasks. Munro and Pizzini, (2004) describe an application for building interactive graphical simulations and training that can be offered in simulation contexts. In Figure 1 a sequence of screen shots from *Rivets* editor is shown. *Rivets* is an authoring environment that produces simulations (which may have game-like features) that can be delivered by *iRides*. Here, a mini-lesson on achieving a particular goal state by carrying out a sequence of actions is authored. In the left panel of Figure 1, the author has specified that a Goal Drill is being authored. A list of possible initial states for the

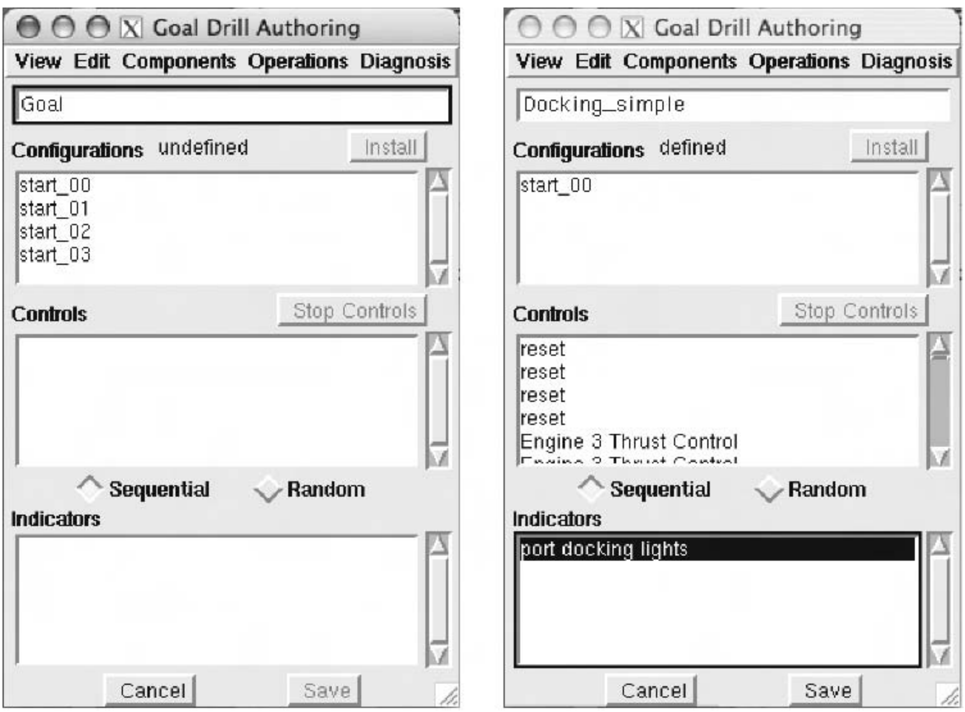

Figure 1. A Rivets (Munro & Pizzini, 2004) interface for authoring a demo by performing actions that are recorded.

game is shown (*start_00, start_01*, etc.). The author selects one of these (*start_00*) and hits the Install button to place the game in that initial state. The author also enters a name for the mini-lesson, "Docking_simple." (This simulation-game was about docking a space craft at a space station – an environment that offers opportunities to learn about Newtonian mechanics, including the relationships among force, mass, inertia, time, speed, and acceleration.)

After the initial state is selected, the author simply carries out the steps of the task, as it should be performed under the condition specified. Every time an object that is marked in the game/simulation-environment as a *control* is manipulated, a record is made of that action, and the name of the object is listed in the Controls panel. When the task was complete, the author clicked on *Stop Controls*. At this point, the Indicators panel became active. The author clicked on the object whose visible condition reflected the goal state of the task. (In this case, the docking lights of the target spaceport hanger had switched from flashing to steady when the spaceship was successfully docked. It is also possible to point to multiple objects whose states taken together represent the goal condition.) See the second panel in Figure 1.

When the author clicks the Save button, the record of his or her actions is converted into a mini-lesson specification, and a lesson editor window opens to display that automatically constructed lesson specification (see Figure 2). This generated lesson can be used to demonstrate to a game learner how to carry out the set of actions that the author did. While it presents the demonstration, the lesson delivers comments about its actions, using names of the objects that were provided by the simulation during the task recording process.

Authors can modify generated lessons like that shown in Figure 2. Steps can be deleted, changed in sequence, or modified. Additional text (or speech) presentations can be inserted in the lesson. Objects can be graphically highlighted to draw the learner's attention, and the simulation can be put into new states under the control of the lesson. Naturally, authors can change the names of the steps in a lesson so that they will make more sense, which supports long-term lesson maintainability.

The right side of Figure 3 shows such an edited lesson, after modifications. (The left side of the figure shows part of the prototype game environment for the spaceship-docking game.) In this lesson, the author has modified a simulation so that events are spaced by authored pauses of predetermined length. Additional explanatory text has been inserted, and the demonstration sequence is now initiated when the learner shows that he is attending by clicking on the spaceship icon when directed to do so.

We have found that many authors find it easy to modify generated text, to modify the timing of instructional actions, and to provide meaningful names for the steps in an instructional sequence. It is much more difficult for them to generate the basic control structure of the lesson, starting with an empty lesson editor. This is why the task-step recording environment (Figure 1) is a useful tool for quickly roughing out a demonstration lesson with the appropriate structure.

Rivets lessons can be delivered in three different modes. In *demonstration mode*, the lesson performs the actions required to carry out a task, while explaining the steps. In *practice mode*, learners are asked to perform the task. The lesson can operate in either

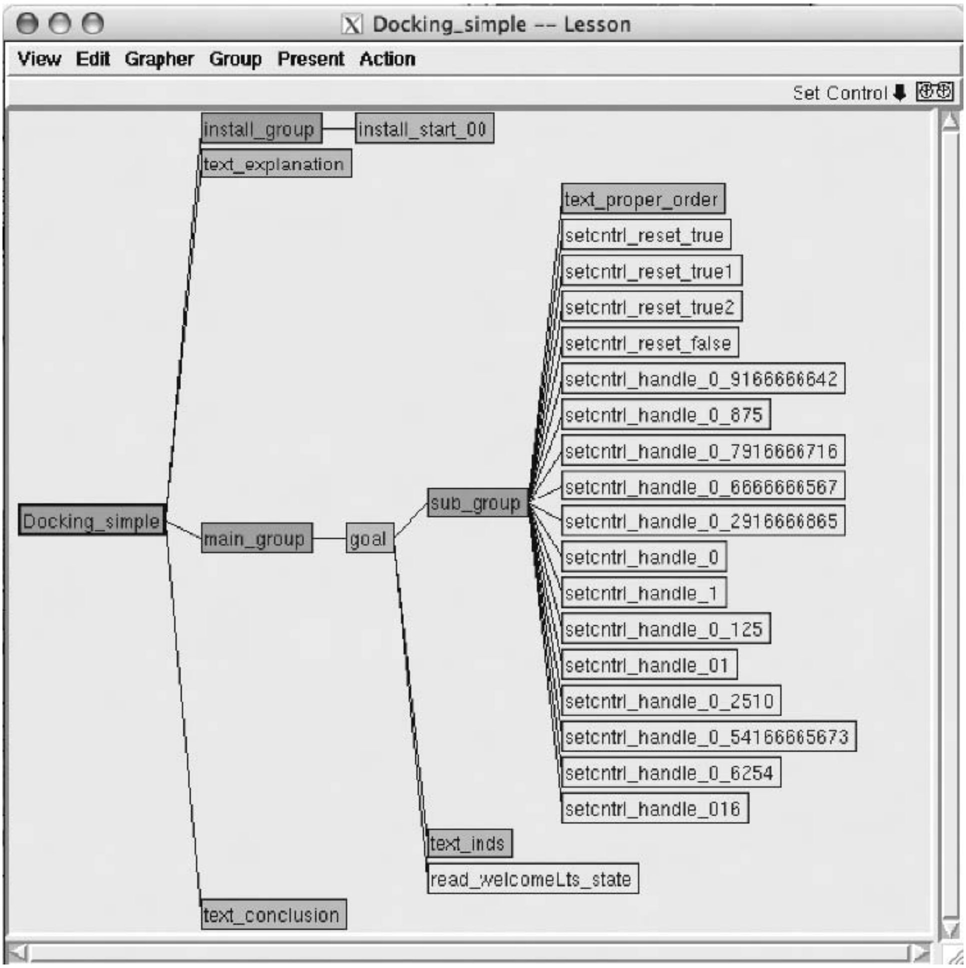

Figure 2. The structure of the generated mini-lesson.

of two ways in practice mode, at the option of the instruction author. One way is to require that each step be carried out in sequence, and to permit no deviation from the authored/recorded sequence. When a learner tries to deviate from the required sequence, the lesson shows which step must be taken next. The other way practice mode can work is to permit any set of actions, and to notice when the goal state is achieved. If the student asks for help (or, at the option of the author, if the student takes too much time), the lesson can present the *demonstration mode* to show how the task can be carried out. The third mode is *test mode*, which is similar to practice mode, but does not include detailed remediation or guidance. These modes of delivery provide a basic sequence of instruction that implements the GEL framework.

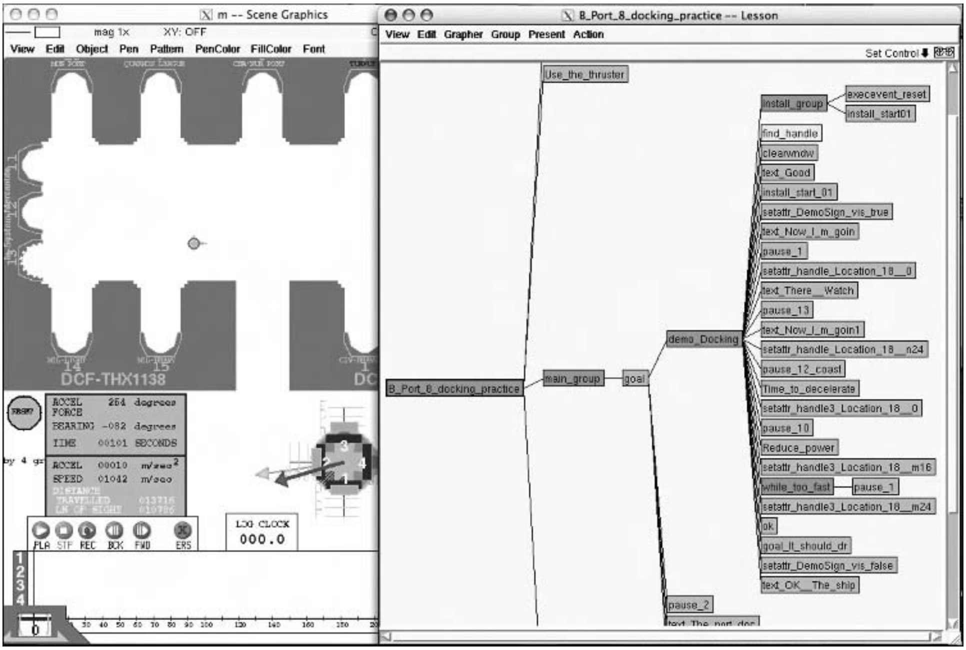

Figure 3. Authoring instruction about a procedure.

6. Conclusion

Games can play a number of roles in learning, but they are most effective as learning environments when they have well-integrated guidance and/or instructional support. Using separate instructional control software and game software systems will reduce total cost and improve software reliability, and, in concert with appropriate authoring tools, may it may promote the evolution of training and instruction in game contexts. The iRides system constitutes a demonstration of the feasibility of the instructional services concept for interactive environments. Its authoring tools, including Rivets, support the development of separate executable specifications of simulations and games and of instructional interactions that can be delivered in the game contexts.

Acknowledgments

This research was supported by the Office of Naval Research Grant N00014-06-1-0711 through and through the Center for Research on Evaluation, Standards and Student Testing (CRESST) of the University of California, Los Angeles (UCLA) under subaward 0070 G HC649.

References

Aldrich, C. (2004). *Simulations and the future of learning*. San Francisco: Pfeiffer.

Baker, E. L. & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*, 269–282.

Belanich, J., Sibley, D. E., & Orvis, K. L. (2004). *Instructional characteristics and motivational features of a PC-based game* (Research Report 1822). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. Cited in Hayes (2005).

Bright, G. W. & Harvey, J. G. (1984). Computer games as instructional tools. *Computers in the Schools, 13*, 73–79. Cited in Hays (2005).

Brunner, J. S. (1961). The art of discovery. *Harvard Educational Review, 31*, 21–32.

Clark, R. E. (1989). When teaching kills learning: research on mathemattics. In H. N. Mandl, N. Bennett, E. de Core, & H. F. Freidrich (Eds), *Learning and instruction: European research in an international context*. Vol. *2*, 1–22. London: Pergamon.

Clark, R. E. (2004). *Design document for a guided experiential learning course*. Technical Report. Los Angeles: University of Southern California. Retrieved on November 17, 2006 from http://cogtech.usc.edu/publications/clark_gel.pdf.

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.

Facer, K. (2004). Computer games: the most powerful learning technology of our age? *Vision Magazine*. Abstracted retrieved on December 11, 2006 from http://www.ltss.bris.ac.uk/interact/31/facer.doc.

Gutwill, J. P., Federiksen, J. R., & White, B. Y. (1999). Making their own connections: students understanding of multiple models in basic electricity. *Cognition and Instruction, 17*, 249–282.

Hays, R. T. (2005). *The effectiveness of instructional games: a literature review and discussion* (Technical Report 2005-004). Orlando: Naval Air Warfare Center Training Systems Division.

Hollan, J., Hutchins, E., & Weitzman, L. (1984). Steamer: an interactive inspectable simulation-based training system. *AI Magazine, 5*(2), 15–22.

Jonassen, D. H. (1991). Objectivism vs. constructivism: do we need a new paradigm? *Educational Technology Research and Development, 39*(3), 5–14.

Jonassen, D. H., Howland, J., Moore, J., & Marra, R. M. (2002). *Learning to solve problems with technology: a constructivist perspective* (2nd ed.). New York: Prentice Hall.

Kirkpatrick, D. L. (1994). *Evaluating training programs: the four levels*. San Francisco: Berrett-Koehler. Cited in O'Neil et al. (2005).

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.

Leemkuil, H., de Jong, T., & Ootes, S. (2000). Review of educational use of games and simulations (Knowledge management Interactive Training System Project No. IST-1999-13078). The Netherlands: University of Twente. Retrieved on January 11, 2007 from http://kits.edte.utwente.nl/documents/D1.pdf.

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3*, 113–132.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*, 14–19.

Michael, D. & Chen, S. (2004). *Serious Games Summit 2004 Report*. Retrieved on December 27, 2006 from http://www.gamedev.net/reference/articles/article2173.asp.

Munro, A. (2003). Authoring simulation-centered learning environments with Rides and Vivids. In T. Murray, S. Blessing, & S. Ainsworth (Eds), *Authoring tools for advanced technology learning environments*. Dordrect, The Netherlands: Kluwer Academic Publishers.

Munro, A. & Pizzini, Q. (2004) *Rivets Reference Manual*, Behavioral Technology Lab, University of Southern California.

Munro, A., Johnson, M. C., Pizzini, Q. A., Surmon, D. S., Towne, D. M., & Wogulis, J. L. (1997). Authoring simulation-centered tutors with RIDES. *International Journal of Artificial Intelligence in Education, 8*, 284–316.

Munro, A., Surmon, D., Johnson, M., Pizzini, Q., & Walker, J. (1999). An open architecture for simulation-centered tutors. In S. P. Lajoie & M. Vivet (Eds), *Artificial intelligence in education: open learning environments: new computational technologies to support learning, exploration, and collaboration* (pp. 360–367). Amsterdam: IOS Press,.

Munro, A., Surmon, D., & Pizzini, Q. (2006). Teaching procedural knowledge in distance learning environments. In H. F. O'Neil & R. S. Perez (Eds), *Web-based learning: theory, research, and practice* (pp. 255–278). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474.

Papert, S. (1980). *Mindstorms: children, computers, and powerful ideas*. New York: Basic Books.

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

Rieber, L. P. (1992). Computer-based microworlds: a bridge between constructivism and direct instruction. *Educational Technology, Research, and Development, 40*(1), 93–106.

Self, J. (1995). Problems with unguided learning. *Proceedings of the International Conference on Computers in Education*. ICCE'95.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Australia: Acer Press.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*, 9–31.

Sweller, J. (in press). Human cognitive architecture and educational technology. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds), *Handbook of Research on Educational Communications and Technology* (3rd ed.).

Van Merriënboer, J., Shuurman, J., de Croock, M., & Paas, F. (2002). Redirecting learner's attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*, 11–37. Cited in Sweller (in press).

White, B. Y. (1984). Designing computer games to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1*(1), 69–108.

PART 2

TEAM LEARNING

ELICITING AND EVALUATING TEAMWORK WITHIN A MULTI-PLAYER GAME-BASED TRAINING ENVIRONMENT

Talib S. Hussain
BBN Technologies, Cambridge, MA

Shawn A. Weil, Tad T. Brunyé, Jason Sidman, and Amy L. Alexander
Aptima, Inc., Woburn, MA

William Ferguson
BBN Technologies, Cambridge, MA

Abstract

Modern multi-player computer games (or multi-player games) show great potential for enabling distributed training of a variety of skills, including effective teamwork, in a variety of simulated environments. We explored the utility of a fantasy-based multi-player game for training teamwork skills. Forty members of the United States Army Infantry participated in the study. The lessons learned indicate that multi-player game-based training systems can elicit teamwork behaviors and provide a viable environment in which those behaviors may be practiced and improved. Our methodology, design, exercise, and results are discussed.

1. Introduction

The recent rise in the popularity and technological robustness of multi-player games has led to a variety of efforts to use such applications for training groups of people in general, and members of the military in particular (Bonk & Dennen, 2005). These efforts seek to exploit the ability of multi-player games to engage players in an immersive, enjoyable simulated environment to elicit complex behaviors and improve performance on targeted tasks and goals. As training systems, multi-player game technologies have the potential to facilitate learning by providing opportunities that are readily available, low cost, widely distributed, and engaging. In order to explore this potential, we conducted a study, named *Gorman's Gambit*, to explore the use of multi-player game technologies to support military training of teamwork skills (Hussain & Ferguson, 2005; Weil, Hussain, Diedrich, Ferguson, & MacMillan, 2004; Weil, Hussain, Brunyé, Sidman, & Spahr, 2005a,b).

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

Our work is based upon the thesis, put forward by General Paul Gorman (US Army, Ret.), that teamwork skills can be taught effectively to military personnel using modern commercial off-the-shelf multi-player games and that there is no need for the game to be realistic with regard to modern military operations to be effective (Gorman, 2003). Although the level of weaponry and the obstacles faced are very different, the elements of teamwork required for a "siege of Camelot" are similar to those for many military operations.

In an 8-month effort, we developed the foundations for a training system targeted to teams of 20–40 individuals using a multi-player game based in a fantasy world, developed a framework for evaluating the training value of this system, and conducted a two-day exercise to validate our design. In the following sections, we first discuss some related research. We then describe the methodology we followed in creating our training system and the methods we developed for assessing the level and types of teamwork exhibited by players. We discuss the specific results as well as their general implications and summarize both the lessons learned about designing and conducting empirical studies of large teams (i.e., "meta-level" lessons), as well as the lessons obtained from our study on the value of game-based training of teamwork skills for large teams.

2. Background

2.1. Game-Based Training for Teams

The rise of global network connectivity, together with the development of multi-player games that exploit that connectivity, has made it possible for individuals distributed around the world to interact in common virtual environments. For instance, commercially available "massive" multi-player games allow dozens to hundreds of geographically distributed participants to work together in simulated locations, often displaying high degrees of coordination to accomplish complex goals. The potential of this type of interaction to be used for training purposes has not gone unnoticed, and a number of people have explored the use of multi-player games as training tools (Alexander, Brunyé, Sidman, & Weil, 2005; Bonk & Dennen, 2005).

In particular, many game-based training technologies have been fielded within the US military (see http://www.dodgamecommunity.com for a complete listing). For example, the US Army uses *America's Army* (registered trademark of the US Army) and *Full Spectrum Warrior* (registered trademark of Pandemic Studios). Several fielded games also exist for the Air Force, Navy, Marines, and Joint Forces, such as *Quickstrike Time-Sensitive Targeting Trainer* (developed by MÄK Technologies), the *SOCOM: US Navy Seals* family of games (copyright Sony Computer Entertainment America Inc.), *Close Combat Marines* (based on Close Combat, a trademark of Atomic Games/Destineer Publishing), and *Joint Force Employment*, respectively.

DARWARS Ambush! (Diller, Roberts, Blankenship, & Nielson, 2004; Diller, Roberts, & Willmuth, 2005), currently used for convoy-operations training, is one example of successfully developed and deployed game-based training. Specifically, the system trains soldiers skills required to anticipate and react to ambushes and improvised explosive devices and facilitates the transfer of effective strategies during personnel rotations. The system, based on commercial gaming technology, enables the collection and dissemination of lessons in a contemporary environment, ultimately providing timely and low-cost training to improve warfighter adaptability.

DARWARS Ambush! has been adapted and integrated into training regimes across the US military, starting with its initial Iraq deployment with the 1st Brigade, 25th Infantry Division, Stryker Brigade Combat Team. Further, the Fort Lewis Mission Support Training Facility has adapted the system to platoon-level training with upward of 60 trainees interacting within a single training simulation, while other users have adapted it to mounted infantry tactics, dismounted operations, rules-of-engagement training, and cross-cultural communications training.

The widespread use of gaming technologies across services is testament to military demand for game-based training systems. The challenge is for research, design, and development communities to collaborate toward producing relevant and effective training technologies that are portable, engaging, extensible, and easy to use. In fact, many training technologies have been met with harsh criticism for not fulfilling one or more of these important characteristics (e.g., Brian, 2005; Erwin, 2000). Furthermore, a number of factors have been identified regarding the transfer of knowledge and skills acquired in multi-player games to the operational environment, namely, fidelity, immersion, presence, and operator buy-in (Alexander et al., 2005).

Fidelity in this context can be described as the extent to which the multi-player game emulates the real world. Fidelity can be measured on multiple continua, and a large number of subcategories (e.g., physical, functional, psychological) have been described in the literature (e.g., Andrews & Bell, 2000; Nystad & Strand, 2006).

Across the currently fielded US military training systems, the focus has been placed on developing multi-player game-based systems that play particular attention to the graphic realism, or physical fidelity, of the tactical environment and equipment. An implicit premise has been that a game-based simulator that holds to how the soldiers will experience the real world is necessary and sufficient to produce a training effect.

However, it has been shown more broadly that fantasy-based games, like simulation-based training (SBT) environments, provide training capabilities by conceptually and functionally capturing and representing the critical characteristics of military environments, known as functional fidelity, to facilitate training (Hollenbeck, Colquitt, Ilgen, LePine, & Hedlund, 1998; Wickens & Hollands, 2000). Further, learning by analogy is an effective means of obtaining skills that will be applied in unrelated contexts (Gentner, Holyoak, & Kokinov, 2001).

Immersion is another factor which has been shown as relevant to the transfer of knowledge and skills – immersion refers to the degree to which an individual feels absorbed by or engrossed in a particular experience (Witmer & Singer, 1998). The degree of immersion experienced within a multi-player game may contribute to the amount of information acquired, skills developed, and subsequent transfer of knowledge to real environments. Many activities are considered immersive (e.g., a game of chess, a book, a conversation), but the term is particularly germane to videogames, as evidenced by the rising phenomena of videogame addiction (Young, 2004).

While immersion may be viewed as the objective description of a multi-player games capability to draw the user into the act of playing the game, presence – also referred to as situated immersion – refers to the subjective experience of actually existing within the computer-mediated environment even when one is physically situated in another (Slater & Steed, 2000; Witmer & Singer, 1998). The amount of human–computer and avatar–avatar interaction required by many multi-player games, compounded by high levels of fidelity on a number of dimensions, may lead to a more vivid feeling of presence in the experience.

Varying the levels of fidelity or increasing the feelings of presence and immersion are thought to impact the degree of transfer from the multi-player game to the operational setting. However, there is another construct that may influence transfer: user acceptance or "buy-in." In the present context, buy-in refers to the degree to which a person recognizes that an experience or event is useful for training. The conjecture is that higher levels of buy-in imply that the user will invest more effort to extract generalizable lessons from training and more effort to transfer those lessons to the real world. Transfer is consequently more frequent and successful as a result.

2.2. Teamwork

Psychological, sociological, and anthropological research has clearly distinguished between "teamwork" and "taskwork" (Salas & Cannon-Bowers, 2001). According to this literature, taskwork involves an individual using a particular set of skills to perform a job; for instance, knowing the sequence of button-presses to successfully draw, aim, and fire a weapon within a combat-oriented multi-player game. In contrast, teamwork skills are not only a compilation of several individuals' skills, but the ability to use these interactively to support team functioning. Orasanu and Salas (1993) note several fundamental characteristics that facilitate team functioning, including multiple members, multiple information sources, interdependence among members, clearly defined roles, and common goals. Team functioning is clearly critical for most military operations, and teamwork behaviors are the observable processes that support effective team functioning, and ultimately increase mission effectiveness.

There is a large body of work demonstrating that these teamwork skills exist, and that they can be defined, trained, and assessed (e.g., Salas & Cannon-Bowers, 2001). Among others, these skills include adaptability, monitoring and backup, communication, and leadership, as detailed in Table 1 (e.g., Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Freeman, Diedrich, Haimson, Diller, & Roberts, 2003; Serfaty, Entin, & Johnstonn, 1998; Sims, Salas, & Burke, 2004; Smith-Jentsch, Johnston, & Payne, 1998; Smith-Jentsch, Zeisig, Acton, & McPherson, 1998).

These teamwork skills form the basis of our training development; we ask whether they can be observed in multi-player game environments, and how these environments can be used in the future for training.

Table 1
Teamwork skills supporting team effectiveness.

Leadership: The ability to direct and coordinate the activities of other team members, assess team performance, assign tasks, motivate team members, plan and organize and establish a positive atmosphere
Monitoring: The ability to develop common understandings of the team environment and apply appropriate task strategies and processes in order to accurately monitor teammate performance
Back-Up Behavior: The ability to anticipate other team member's needs through accurate knowledge about their responsibilities. Includes the ability to shift workload among members to achieve balance during high periods of workload pressure
Adaptability: The ability to adjust strategies based on information gathered from the environment through the use of compensatory behavior and reallocation of intra-team resources; altering a course of action or team repertoire in response to changing conditions (internal or external)
Team Orientation: Propensity to take other's behavior into account during group interaction and the belief in the importance of team goal's over individual member's goals
Closed Loop Communication: The practice of confirming receipt and understanding of others' communications. This practice builds trust in the communication skills, knowledge, and intent of others and ensures that information is accurately conveyed
Team Mental Models: The ability to accurately represent the capabilities of others, their responsibilities, and their perception of the state of the world
Coordination: The practice of planning, preparing, organizing people and/or tasking to accomplish a goal
Communication Push: The practice of sharing or sending information with/to others
Communication Pull: The practice of seeking information from others or other data sources; asking questions; attempts to gather intelligence

3. Methodology

Gorman's Gambit not only represents an attempt to examine the utility of multi-player games for training, but also an opportunity to assess the process of multi-player game-based training system development (Hussain & Ferguson, 2005). In order to achieve the desired learning in training system users, both technical and pedagogical assumptions need to be periodically examined. We adopted an iterative development methodology based on best practices from both software development and the psychological sciences.

The *Gorman's Gambit* project had a 4-month high-level planning phase (April–August, 2004) in which we explored different commercial game solutions and identified the general form of our study (including design approach and basic system design). The development of the training system and assessment tools took place in three main developmental phases over a period of 4 months (September–December, 2004), culminating in a final exercise held at Fort Benning, GA (December 15 and 16, 2004).

4. Goals and Development

In the *Gorman's Gambit* project, we had multiple high-level goals that influenced our design. We sought to elicit a variety of observable teamwork behaviors from a large team using a multi-player game-based training system and gather information on the utility of such systems for training teamwork. However, we also sought to identify processes for rapid development of multi-player game-based training, gather information on the utility of such systems for training in general, and investigate the importance of issues such as the fidelity of the simulation and immersion produced by the game for multi-player game-based training.

Given these high-level goals, we did not have highly specific "training objectives" as such. However, we did focus on developing conditions and measures intended to capture data on a variety of different issues pertaining to multi-player game-based training. Rather than simply list our goals out of context, we introduce them below as the supporting reasons behind a variety of design and development decisions.

4.1. Basic Technical Design

During our planning phase, we made several basic design decisions in order to achieve our high-level goals of eliciting observable teamwork behaviors from a large team while minimizing development effort.

To avoid the potential of restricting teamwork behaviors, we adopted a weakly scripted approach in which the terrain and non-player characters were made relatively simple, but the gameplay options open to the human players were plentiful (i.e., rather than heavily scripted approach that carefully guides players using in-game cues, events, and non-player character actions). In particular, we chose to create a scenario in which two large teams (or "platoons") of players would play against each other in a "capture-the-flag" mission, described in the next section.

To encourage a variety of interactions among teammates, we chose to divide each platoon into multiple squads, give different players game characters (or *avatars*) with different strengths and weaknesses, and to limit communication channels among squads.

To facilitate development and minimize risks, we chose *Neverwinter Nights* (Trademark of Wizards of the Coast, Inc.) to be the basis for our training environment. *Neverwinter Nights* supports up to 64 distributed players; provides extremely stable operation, with straightforward and robust setup and execution of multi-player sessions; allows for the customization of avatar skills and inventory; includes a powerful scenario development environment, the *Aurora Toolset* (Trademark of Bioware Corp, copyright 1997–2005); allows for non-intrusive observation and control via a "Dungeon Master" avatar; and includes pre-scripted game and interface tutorials.

4.2. Iterative Development

Throughout our development process, our goals guided our decisions. We present here a brief summary of this process to illustrate factors to consider when developing fielded

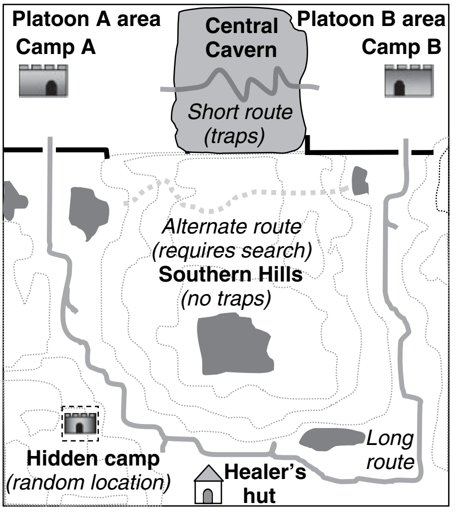

Figure 1. Simulated world for teamwork training scenario.

training systems. We describe three development phases, each culminating in a pilot study that was used to validate whether we were achieving certain goals.

In the first development phase, each platoon had a separate "camp" in the simulated world (largely the same as shown in Figure 1), and avatars were distinguished primarily by the items carried by those avatars (e.g., weapons, healing items, magic items) and their intrinsic skills (sword-fighting, trap detection, spell casting). To limit communications, we augmented *Neverwinter Nights* with a hierarchical chat-based communication capability based on the third-party Talus Speech System software (developed by Josh Dalton (Lanthar D'Alton); includes code copyright 2003 by Ingmar Stieger and Jeroen Broekhuizen; this product includes software developed by the Politecnico di Torino and its contributors; available at http://nwvault.ign.com/) so that each platoon had distinct communication channels, members of one squad could only talk to their leader and squad leaders could talk to each other and to the platoon leader. The communication hierarchy restricted unrealistic chat channels in order to increase functional overlap between the voice-over IP system and true military communications. To ensure that players could actually use the game and to minimize the amount of time spent designing and developing interface training, we chose to use the in-game tutorial to train players. We conducted a pilot run on October 29, 2004 with 12 game-novice personnel from the author-affiliated institutions comprising two "platoons" of six players each, organized in two "squads" of three players each. Following this, the scenario was tested with the mission objective of occupying the enemy camp, operationalized as having two of a team's players alive in the enemy camp for any 2-min period. The pilot showed that the text-based chat interfered significantly with gameplay, that the in-game tutorial was lengthy and ineffective, and that our mission goal resulted in a very fast-paced game that did not encourage players to exploit their various capabilities. The pilot also provided an opportunity to test our assessment techniques and revealed that several improvements were required in our data collection tools.

In the second development phase, we made large modifications to individual avatar characteristics, such as endurance and strength, in an effort to promote teamwork through necessitating dependence on teammates' skills sets. We incorporated the BBNTalk Voice-over-IP system (copyright BBN Technologies) that was developed as part of and used successfully in the *DARWARS Ambush!* system, and which is capable of readily handling 20–40 players with hierarchical communication channels. The voice-over-IP software was set up external to the game and players wore a headset and microphone. To improve the tutorial process, we developed a guide for participants to follow in order to minimize non-essential aspects of the tutorial (e.g., aspects of an entertainment-based storyline, teaching skills that were not to be used by that avatar in the *Gorman's Gambit* scenario). We also developed two additional semi-scripted tutorials, the Tutorial Arena and the Tutorial Mission (described later) that emphasized experiential learning. We conducted a second smaller pilot with seven personnel from the author-affiliated institutions which validated our changes.

In the final development phase, we added a third hidden camp and adapted the mission objectives in order to encourage differentiated strategies and avoid simple race conditions. We also incorporated several capabilities to enable an in-game observer (via a "Dungeon Master" class) to keep better track of game activities. We conducted a final pilot using 32 personnel from the author-affiliated institutions as participants, and using a reasonably fast game server machine (Pentium M 1.7 GHz) and a variety of client machines (Pentium III to Pentium M 2.0 GHz). The pilot showed that our system scaled well; only low-end client machines (i.e., Pentium III) exhibited significant slow-downs under heavy load. It also validated our measurement tools and showed that the new conditions elicited a variety of teamwork behaviors, such as resource management, communications, strategizing, and working toward common goals.

5. Conditions

5.1. Final Training System Design

The world used in the *Gorman's Gambit* training system is illustrated in Figure 1. A new mission would begin with each platoon stationed within its own home camp (i.e., Platoon A in Camp A, Platoon B in Camp B). Travel between the two camps was constrained. There were two routes that the players would be informed about in advance, a "long-but-safe" route and a "short-but-dangerous" route. The long-but-safe route traversed the Southern Hills, a safe area with no traps, while the short but dangerous route traversed the Central Cavern, an area filled with hidden traps of varying severity. There was also a third route that was both safe and of medium length, but which players would not be informed

Table 2
Exercise characters and their respective roles and resources.

Character class	Role in scenario	Resources
Platoon Leader Platoon Leadership	Coordinates activities directly with the Squad Leaders	Small dagger
Squad Leader Squad Leadership	Communicates with the Squad, other Squad Leaders, and the Platoon Leader. Moderate combat ability, moderate speed, vulnerable to injury	Long sword
Archer Long-Range Weapons	Provides long-range weapons support with low-moderate close-range combat abilities	Long bow, crossbow, and several arrow types
Artillery Long-Range Magic	Capable of short and long-range attacks. Low physical combat ability, high resistance to spells	Scrolls, potions, wands, staves and rods. No armor
Medic Health Support	Provides health support via resurrection and healing. Very low combat ability, but high speed	Healing magic, small weapon
Scout Primary Reconnaissance	Provides fast reconnaissance, stealthy maneuvering, trap setting, detection and disarming, and mild offensive fighting capabilities	Short sword and claw blade. Able to set traps (i.e., mines)
Tank Primary Close-Combat	Provides powerful fighting capabilities, but cannot travel quickly	Long sword, short sword, and great sword

about and would therefore need to discover. Upon death, avatars would respawn (after a 1-min period) to the Healer's Hut, at a significant distance from both home camps.

To encourage teamwork, mission-essential skills were distributed among five primary avatar types (archer, artillery, medic, scout, and tank), as detailed in Table 2. For example, disabling a trap in the Central Cavern would require the assistance of a Scout, while healing a serious wound quickly would require a Medic. The platoon and squad leader were given avatar types with general skill sets.

A third camp ("Hidden Camp") also existed which was heavily protected by non-player characters. At the beginning of each mission, the location of this camp was randomly determined (within the Southern Hills or Central Cavern areas), and the players would not be informed of that location. The third camp was considered neutral territory until captured and defended by a team.

Within all three camps was a pillar of colored light (i.e., the "flag") that indicated the current "possessor" of the camp. To obtain the possession of a camp for his team, a player's avatar would have to pull a lever placed adjacent to the flag; the territory would then remain in that team's possession until an avatar from the opposing team pulled the lever.

The general goals of a mission were for each platoon to defend their own flag, capture the enemy's flag, and find and capture the hidden flag. We defined two distinct winning conditions. The first was to hold the most flags at the end of a given period of time. The second was to hold as many flags as possible for the longest total amount of time by the end of a given period of time.

5.2. Tutorial Process

Three structured tutorial steps were used to ensure that all players would be proficient with the *Neverwinter Nights* game mechanics prior to using the training system. The tutorials progressed in their levels of complexity and interactivity, and are described below. In particular, though, there was no explicit tutorial on teamwork skills.

The first tutorial step used the existing tutorial exercise contained within *Neverwinter Nights* to train novice players on game and interface basics. The in-game tutorial would place the player's avatar into a virtual environment containing virtual instructors that would engage in simple training dialogs. Under the auspices of these virtual instructors, the player could practice the essential gameplay fundamentals, such as using their avatar's inventory, weapons, and basic functions. The players were given specific written materials to help them focus on learning the gaming and character skills that were of particular use for the *Gorman's Gambit* training system.

The second tutorial step provided each team with a relatively free-form "tutorial arena" within which the players could gain detailed specifics about their avatars' individual skills and abilities (provided by several avatar-specific non-player characters with which players could hold brief dialogs) and engage other teammates to practice those skills (e.g., combat, casting spells, healing others, etc).

The third tutorial step provided a squad with a "tutorial mission" that required all squad members to work together toward locating and capturing a hostage princess, located somewhere within a "tutorial village." In order to reach the princess' location, the squad needed to perform a variety of tasks. These tasks included finding hidden keys to unlock doors along their path, detecting hidden doors, obtaining information from villagers, engaging enemies, and ultimately defeating a dragon. The tutorial mission was designed to emphasize the importance of each individual avatar's unique skills, attributes, and abilities (e.g., certain tasks could only be accomplished by certain avatars; for example, a hidden door or key could only be detected by a scout; while others could only be accomplished together; for example, the dragon could only be killed through a collaborative effort of several different avatar types).

5.3. Gorman's Gambit Exercise Structure

The *Gorman's Gambit* exercise was structured to involve three successive sessions. In order to facilitate effective teamwork, we adopted a session design that enabled each team to privately coordinate together in person prior to playing a mission (planning period) and then to privately discuss the events occurring immediately after conducting a mission (debriefing). We determined that both of these sessions would be more effective if done in person, occur more quickly than if performed online (using, say, the in-game text chat), and have the benefit of being easily observable. At the end of the third and final session, all participants would complete a post-experiment questionnaire and then participate together in a final AAR. While debriefings were intended to be used to reflect on specific instances in specific sessions, the AAR was intended to be used as an opportunity for the participants to reflect on the larger questions pertaining to the potential of using multi-player games to train teamwork.

6. Measures

The primary dependent measures for the *Gorman's Gambit* effort were observations of teamwork skills. Accordingly, the first step was to ensure that our exercise was designed to elicit teamwork (as described in the previous sections). The second step was to develop a technique to capture the behaviors that are indicative of teamwork within the context of a multi-player game. It is important to note that our aim was not to capture every instance of teamwork, but rather to obtain a sample of representative behaviors to support the notion that multi-player games can be used for training. Given the broad nature of our goals, we chose to capture data through both observer-based and self-report techniques. We chose not to capture data via systems-based techniques. On the one hand, the level of data we could capture easily within the training system (e.g., number of casualties, traps disarmed, buttons pressed, or menus accessed) would not have provided direct support for our goals. On the other hand, significant software development would have been required to capture data that would have supported our goals (e.g., instances of decision making, communication of situation reports, issuing of commands).

As such, we developed a suite of measurement techniques and tools that would enable us to capture a broad array of data:

- observer in-game evaluation forms;
- post-session team debriefings;
- post-exercise questionnaires; and
- final AAR participant feedback.

Each of these measurement techniques is described in turn, below.

6.1. Observation Tool

Our primary tool to capture data on the teamwork elicited during gameplay was the *Observer In-Game Evaluation Form*, which was filled-in by experimenter/observers using one of three techniques. Our primary technique was over-the-shoulder observation of player activities; a second technique was the use of the in-game observation capability (provided by a "Dungeon Master") which enabled viewing of other players' activities without them being aware (the "Dungeon Master" avatar was invisible to other avatars); a final technique was the use of a video camera to capture gameplay (over-the-shoulder) as

#	Time	Participants involved	Description of event	Teamwork skills exhibited			Comment
				Skill shown	**Y/N**	**Quality (1–5)**	
1	08:00	Platoon A Archer and Platoon B Tank	Tank and archer engage in battle. Archer calls up squad members for back up. Squad confirms message receipt, arrives and destroys Tank.	Monitoring	Y	3	
				Back-up	Y	4	
				Coordination	Y	3	
				Comm – Push	N	N/a	
				Comm – Pull	Y	5	
				Other… (Closed Loop Comms, Leadership, Orientation, Backing Up, Adaptability	Y	5	

Figure 2. An event captured on the Observer In-Game Evaluation Form. Quality scale ranged from 1 to 5; 1 indicates a poor example of the behavior; 5 indicates an exemplary example.

well as more general player activities. These observation techniques were adopted in order to allow us to monitor participants' behavior (e.g., see actions, hear communications) without disrupting gameplay.

The observation form allowed for the recording of event details (i.e., event #, time, players involved), a detailed description of the event, the presence or absence and quality rating of several key teamwork skills, and comments. An example of an event recorded onto the Observer In-Game Evaluation Form can be seen in Figure 2.

A critical component of the observer form was the ability to record instances and quality of observed teamwork skills. Toward this end, the form included a subset of skills identified by contemporary teamwork psychology literature (monitoring, back-up, team orientation, communication push and pull, coordination; see Sims et al., 2004). A content dictionary (Table 1) provided operational definitions for each teamwork skill.

Observers sampled across multiple participants and throughout the entire exercise duration. Sampling rates varied across observers, given differences in teamwork frequencies and durations, and participants' locations; average sampling rates were one observation per minute, per observer. Observers were instructed to actively gather information from multiple participants regarding a variety of teamwork behaviors; it is possible, of course, that certain teamwork behaviors were unobserved, or that particular behaviors were more salient than others. Following the exercise, the video was analyzed to capture additional instances of teamwork that had not been noted previously. All observations made during the exercise, including planning periods, missions, debriefings and the AAP, and from the video were collected into a single multi-data source (Figure 3).

6.2. Team Debriefings

The team debriefings were intended to be used for actively capturing measures, with the teams explicitly being instructed to discuss such topics as goals, strategies, performance, and lessons learned; this debriefing format is quite similar to that used in actual military

		Teamwork behaviors							
Teamwork observed	Specific teamwork skill/s	Monitoring	Coordination	Comms, Push	Comms, Pull	Leadership	Orientation	Data source	Team and/or mission details
At start up of mission. PL assessed map, coordinated group, pushed commands and instructions, discussed setting up booby traps. PL also described environment and instructed platoon to move slow.	Monitoring, coordination, comms-pushed and pulled, leadership, team orientation	1	1	1	1	1	1	LS observation sheet	Mission #1, Platoon B, 8:40 AM, observing Archer S1 and PL

Figure 3. An example of teamwork captured and archived in the multi-data source.

operations. Observers used this opportunity to ask specific questions related to ambiguous or incomplete teamwork recordings from the mission, elaborating on observations, and filling in any gaps in knowledge.

6.3. Post-Experiment Questionnaire

We designed a *Post-Exercise Questionnaire* to be presented to and completed by the participants before the final AAR (see Appendix A). The *Post-Exercise Questionnaire* was of two pages and was comprised of 21 questions addressing several areas, such as game experience (e.g., pleasurable, stressful), knowledge about their avatar, team interactions, self and team assessment, situational awareness of self and team, and practical applications of the game as a learning tool. Nineteen of the 21 questions asked participants to give their answers on a seven-point Likert scale relevantly anchored at low, midpoint, and high. Five of the 21 questions also included comments fields where participants could further detail their experiences and opinions.

6.4. Final AAR

The final after action review was a critical part of our measures since it was a communal activity in which all participants could offer their views. In particular, the AAR was intended to address experiences and elicit opinions on the following six topics:

1. positive and negative feedback regarding gameplay;
2. challenges faced by the team;
3. examples of teamwork used to overcome challenges;
4. generally adopted strategies and adaptations;
5. suitability for army training needs; and
6. comparison to other multi-player games.

Appendix B lists the prepared AAR questions.

7. Exercise

The entire exercise was conducted over the course of 2 days, the first of which was dedicated to the tutorial process and the second to the main *Gorman's Gambit* exercise.

7.1. Participants

Forty members of a US Army Infantry Platoon, between the ages of 19 and 33 years ($M = 23.6$), participated in the exercise. Military rank and experience varied from E-2 (Private) to O-1 (Second Lieutenant), and from 1.5 to 174 months, respectively. Overall, the participants reported both computer game ($M = 3.7$ hours/week) and console game ($M = 4.0$ hours/week) experience within the preceding year (see Appendix C for demographic form).

Participants were randomly divided into two groups of 20, each roughly representing a platoon. Each group contained one Platoon Leader and three similarly composed squads, with six or seven players each, as defined in Table 2. Critically, roles, resources, and responsibilities varied widely within squads as a function of avatar type. Figure 4 illustrates the composition and communication hierarchy of each platoon. All participants used a personal computer (PC) with a 15 inch CRT monitor, keyboard, optical mouse, and voice-over-IP headset.

7.2. Tutorial Day

On the first day, the participants were presented with the tutorial materials relevant to their avatar, and then all participants executed all three steps of the tutorial process. We ended each step of the process only when everyone had completed it. The first step (in-game tutorial) lasted approximately 45–50 min, and staff were available to field questions as necessary. The second step (tutorial arena) lasted approximately 30 min. The third

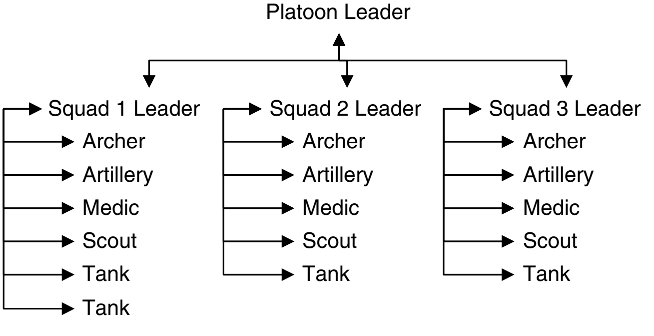

Figure 4. Platoon composition and communication hierarchy.

step (tutorial village/mission) lasted approximately 40 min. During the tutorial mission, it was clear that teamwork suffered and many avatars were injured and killed, until the participants became aware of, and leveraged, each avatar's unique characteristics. This mission also provided a last opportunity for participants to practice using their complex array of skills and resources (e.g., voice-over-IP) before the main *Gorman's Gambit* exercise took place.

7.3. Main Exercise Day

The same participants returned on the second day for the main *Gorman's Gambit* exercise. The exercise involved three successive sessions, each with a 15-min planning period, 30 min mission, and 15-min debriefing. Different winning conditions and communication mediums were used in different sessions:

- possession of at least two levers at the end of 30 min of play using voice-over IP;
- possession of at least two levers for the longest duration in 30 min of play using text chat; and
- possession of at least two levers at the end of 30 min of play using voice-over IP.

The use of alternating conditions was intended to encourage different strategies across missions. For gameplay, all participants were instructed to behave as they would in a true military mission, for example, they were to maintain contact with their leaders, assist their teammates, and engage the opposing team as necessary.

The planning periods were used effectively by the platoon leaders to devise a strategy for that session and to convey it to the team. It was interesting to note that the two platoon leaders followed very different styles (one discussing a plan with the entire team on a white board and the other collaboratively discussing the plan with his squad leaders, who then relayed the decisions to their squads). During mission execution, the two platoons were physically separated; one (Platoon A) was seated in cubicles within a single large room, while the other (Platoon B) occupied four separate rooms due to facility restrictions. During the debriefing, the platoon members discussed their execution of the mission, focusing positive and negative aspects of the platoon's performance, as well as potential strategy improvements. At the end of the third and final session, all participants completed the post-experiment questionnaire and participated together in the final AAR. The questionnaire enhanced the quality of the data received through the AAR and enabled those who are less vocal an anonymous medium for sharing their experiences and opinions. During the semi-structured AAR, the staff facilitators presented several core questions probing for global information regarding overall game experiences and the potential utility of multi-player games for use in military training. Some of the questions had been determined in advance of the exercise (Appendix B), but others had been identified during the exercise itself. The AAR produced a rich discussion and many valuable insights into the utility of multi-player game-based training systems for military training, as perceived by the participants.

8. Results

The results of our study are highly varied in form. We present some of the objective measures here and introduce more subjective results primarily in the context of our discussions in the next section.

8.1. Scenario Outcomes

The overall results for each of the three scenarios were as follows:

- At the end of mission one, Platoon B possessed the majority of levers.
- At the end of mission two, Platoon A had the longest total lever control time.
- At the end of mission three, Platoon B possessed the majority of levers.

8.2. Teamwork Observations

A total of 550 instances of specific teamwork skills were extracted from over 115 recorded observations, the post-session team debriefings, and final AAR. The teamwork behaviors seen in these instances are enumerated in Table 3. Critically the number of teamwork skills observed *during* each mission (mission 1 = 114; mission 2 = 136; mission 3 = 158) increased substantially over the course of the exercise [$\chi^2(2) = 7.12, p < 0.05$]. Most events involved multiple teamwork skills (five to six skills), as teamwork skills often work in conjunction with each other, such as leadership and monitoring. Leadership involves

Table 3
Instances of teamwork observed in the current exercise.

Behavior	Number of observations
Monitoring	75
Coordination	88
Pushing information	57
Pulling information	35
Leadership	69
Orientation	66
Backing-up	53
Adaptability	43
Closed loop	6
Mental models	58
Total	**550**
Total mission 1	165
Total mission 2	176
Total mission 3	209

directing and coordinating the activities of others, assessing performance and/or assigning tasks (Sims et al., 2004). Monitoring involves understanding the team environment and applying appropriate strategies (Sims et al., 2004). The two skills have overlapping definitions that would make it probable, although not without exception, that where you have leadership you also have monitoring and vice-versa.

The events that were observed varied highly, even when considering a single type of teamwork skill. For instance, the 43 occurrences of adaptability included:

- revising strategies in between missions based on lessons learned from previous missions;
- adjusting to a breakdown in communication medium (e.g., by setting up a message relay protocol between the avatars in the game);
- adopting new roles (e.g., by recognizing that an avatar with a typical non-combat role was best suited for a particular offensive attack since it was faster than the enemy);
- changing the team organization to improve offensive capability (e.g., by splitting up into pairs of avatars with complementary skills);
- modifying the mission plan dynamically upon discovering that an avatar had a useful capability they had not been notified of (e.g., discovering that they could summon a flying goblin led to its use for remote surveillance, much in the manner of a modern unmanned aerial vehicle).

Although we could present detailed examples of each of the behaviors observed, space precludes this. Rather, we will introduce some illustrative behaviors in the context of our discussion of the lessons learned. Further, note that it would be possible to extract several hundred more examples of teamwork from the data sources. However, our intent is not to document all of the instances of teamwork, but, rather, to demonstrate that teamwork is readily elicited (and, therefore, that multi-player game environments may provide at least some of the conditions necessary for teamwork skills training).

8.3. Questionnaire Responses

Table 4 summarizes the responses of the 40 participants, giving the mean and standard deviation for all questions. Some of the original questions had multiple parts and these

Table 4
Summary of questionnaire results ($N = 40$).

Short wording	Mean	StdDev
Understand avatar?	6.2	1.2
Game exciting?	4.1	1.6
Game interesting?	4.2	1.6
Game stressful?	1.7	1.8
You help teammates?	5.1	1.9

(*continued on next page*)

Table 4
(*Continued*)

Short wording	Mean	StdDev
Teammates help you?	4.5	2.2
Push info?	5.6	1.8
Receive info?	5.4	1.8
Provide non-combat help?	2.8	2.5
Receive non-combat help?	4.1	2.2
Overall teamwork of team?	5.4	1.5
Overall performance of team?	5.7	1.2
Coordinate with teammates?	5.3	1.5
Monitor teammates?	4.7	2.0
Anticipate teammate needs?	4.9	1.5
Your situational awareness?	5.2	1.4
Workload level?	3.6	1.7
Good teamwork tool?	3.9	1.8
Learn to strategize?	3.9	1.8
Enough tutorial training?	4.9	1.6
Aware of need to manage resources?	0.9	0.2
Aware of lack of resources?	0.8	0.4
Aware of real-world parallels?	0.8	0.4
Playing leads to better teamwork?	3.5	2.0

are separated out here. For improved readability, Table 4 uses highly shortened versions of the questions. We discuss most of these responses in our discussions.

9. Discussion

Based on our experiences in developing the training system, the results obtained from observing the participants during the tutorial and exercise sessions, and the comments and suggestions experienced during the AAR, we are able to identify a number of lessons that validate the goals of the *Gorman's Gambit* project. In this section, we discuss the training utility and fidelity of multi-player games and convey lessons learned for efficiently developing and studying game-based training.

9.1. Training Utility

The results of the exercise show, in several ways, that multi-player games may be useful for supporting team training. First, it is clear that a variety of teamwork skills

(550 instances) were demonstrated, which is highly suggestive that multi-player games can support large-scale exercises in which many individuals with differing but complementary skills work together toward a single effort.

Second, when asked, soldiers were able to see several positive aspects to the game they participated in. From the questionnaire, we see that soldiers rated playing games like the *Gorman's Gambit* scenario as having a moderate ($M = 3.4$) ability to help them become a better team member. Moreover, the most frequent positive verbal comments provided by soldiers were related to teamwork. Soldiers viewed the game and scenarios as a reasonable tool to train teamwork ($M = 3.9$) and for promoting development and adaptation of strategies ($M = 3.9$).

Interestingly from the observer's standpoint, the soldiers appeared to demonstrate better and better team coordination as time passed and over successive missions. In fact, the number of teamwork instances increased substantially over the course of the three missions, with the fewest and most occurring during the first and third missions, respectively. In other words, it appears that the participants were learning better ways of working together within the context of the game. However, during the AAR, the soldiers did not acknowledge that they had learned any teamwork – rather that they were already experts in teamwork and that the game would have training utility primarily for novices. Thus, the results argue that properly developed games can have favorable training outcomes, but that participants may learn while using a game-based trainer without consciously realizing its training value.

9.2. Fidelity

In terms of seeing functional similarities between a game environment and a military environment, a large majority of the sample (80% of the soldiers) reported verbally in the AAR that they saw such similarities. This sub-sample of soldiers went on to enumerate several examples where the game environment provided a functional relationship to a military environment; the examples were: teamwork, specializations, communications, hierarchy, fighting in squads, coordination, setting defensive positions, short/long range weapons, healing times, traveling in formations, strategizing, and conserving resources.

The *Neverwinter Nights* software is largely limited to ground-based interaction, congruent with the operational requirements of the infantry. There were several powerful (unexpected) analogies that occurred during the exercise. For instance, *Neverwinter Nights* allows the artillery avatar to summon a small flying goblin. The soldier's point of view could be changed to take on the goblin's perspective and fly quickly through dangerous areas, avoiding traps and enemy engagement, and allowing for reconnaissance. This was quickly recognized as being analogous to the real-world function of a UAV and exploited accordingly.

Game-based simulators operating on a laptop or desktop do not have the same physical or visual fidelity as large simulators, but they can still, if designed appropriately, reflect real-world characteristics to a high degree. For instance, an important component of teamwork skills is communication and coordination. We explicitly attempted to provide an adequate and realistic communication capability by augmenting the game with hierarchical text and voice-over-IP communications. Our efforts were rewarded, for the soldiers commented that the communication system was a good analogue to their typical hierarchical system. The questionnaire results show that soldiers thought they were able to push information at a high level to others ($M = 5.6$), receive a high level of communications ($M = 5.4$), and coordinate at a high level ($M = 5.3$).

It is clear that, as predicted by General Gorman, the use of a fantasy setting afforded the exhibition of the teamwork behaviors. However, the departure from their usual operational environment also makes it more difficult for buy-in from military participants. Some soldiers reported having difficulty taking on a military mindset while playing the game, and the most frequent negative comment voiced during the AAR by soldiers was poor realism (43%). This indicates that a sizable minority of soldiers is conflicted by the game environment and believes that more operational realism is required for adequate transfer of training to occur. However, the non-operational setting could easily be used to train critical thinking skills and challenge participants to engage appropriately in unfamiliar environments.

9.3. Developing Games for Training

Several lessons were learned regarding the development of games for training and the development of effective empirical studies of such games. Developing our training system so that it would elicit the desired teamwork behaviors was not always straightforward. As shown in the pilot studies, subtle imbalances in gameplay led to a complete breakdown in the effect we were targeting. The iterative development strategy we adopted and the authoring capability provided by *Neverwinter Nights* turned out to be crucial for rapid development and for achieving the effects we wished in the game.

When conducting a study of games for training, time for an exercise is a limiting factor – both in terms of availability of participants (i.e., it is hard to get soldiers to participate for extended periods of time) and cost (i.e., if paying the participants, costs will increase with time required). One area to which it is critical to devote sufficient time, but in which it is also important to minimize that time, is the tutorial process for playing the game itself. As revealed in the pilot studies, this was a key factor. With too little tutorial time, many players were ineffective during the actual scenario and therefore less immersed and poor at teamwork. With the wrong type of tutorial activities, the players spent much of the time during the actual scenario playing independently or trying to learn capabilities, which in turn resulted in poor teamwork. The three-step tutorial process we developed was validated since, mostly, the soldiers “hit the ground running” on the first scenario played. In the questionnaire, soldiers clearly understood what their avatars were capable of ($M = 6.2$) and rated the tutorial process as moderately to highly effective ($M = 4.9$).

The ease of conducting and maintaining a training session is critical for successful participation within a game-based training system. When managing the interaction of 40+ players, simple connectivity that is mostly transparent to the user is a necessity. Within *Neverwinter Nights*, setting up the client and getting started playing the game took a matter of seconds. Further, if a restart of the server (e.g., to start a new mission) or of a client (e.g., machine crashed) was needed, the users could easily join or re-join the game (within seconds). Without this capability, the exercise would have been constantly delayed and participant buy-in would have been negatively impacted.

Likewise, the ease with which changes to the system may be made can impact buy-in. For instance, platoon leaders reported increased levels of satisfaction and effectiveness as a result of the following rapid manipulations.

- At the end of the first day, the platoon leaders indicated that they felt that the written map materials were insufficient for them to maintain situational awareness. After an hour's effort, we introduced an in-game item that performed localization analogous to a compass.
- After the first session of the second day, both platoon leaders expressed dissatisfaction with their avatar's slow speed and low combat abilities and believed these deficiencies detracted from their ability to lead their platoons effectively and gain their soldiers' respect. In only a few minutes, we improved the speed and effectiveness of their avatars.

Despite all design efforts and interventions, it is of course possible for software vulnerabilities to be exposed, especially by inventive gameplayers. For example, during the second mission several soldiers found a way to bypass an intentional game manipulation; by exiting and restarting the *Neverwinter Nights* application upon death, players were able to immediately respawn, fully healed, at the location where they had died (i.e., thereby circumventing the mandatory visit to the Healer's Hut). Even though both platoon leaders subsequently ordered their teams not to do that, it continued to happen increasingly toward the end of the final mission. (We note, though, that several simple technical solutions would have been available if we had realized the extent of the abuse earlier; one such solution being as simple as changing a password.) In addition, by inviting opponent team members into chat groups, several players were able to view enemy communications during the text chat gaming session. These examples of soldier resourcefulness led to frustration and reduced effectiveness of the opposing team (until they figured out how to do it too. . .) and undermined the facilitator-imposed scenario structure. For these reasons, while the training product must be robust enough to allow for flexibility, it must also guard against user or environmental factors that could mitigate training value.

10. Conclusions

The military is interested in supporting effective large-scale distributed simulation-based training that will enhance and expedite soldier instruction. In the *Gorman's Gambit* effort, we focused upon clearly demonstrating the potential of multi-player games for training of teamwork skills. The training system we developed was shown, in an exercise, to elicit a variety of teamwork behaviors and was perceived by both the participants and the observers to have some potential value for training. Collectively, the lessons learned from this exercise demonstrate that multi-player games may be good models to emulate for military training. However, there are inherent challenges with using commercial, off-the-shelf gaming systems in assessment-intensive applications such as military teamwork training that were not directly addressed in *Gorman's Gambit* which must be considered in future efforts.

For instance, the observer-based sampling methodology used in this study was effective in capturing instances of teamwork and sufficient for our needs. However, if we had desired to achieve a truly representative sample of teamwork behaviors under the dynamic conditions of our multi-person environment, it is unclear how many observers would have been required, whether the fields of our Observer In-Game Evaluation Form are the ideal fields, and whether the scales used were sufficient to capture the complexity of interaction. Furthermore, as the number of simultaneous users increase, the sustainability of using human observers diminishes.

Effective, large-scale multi-player game-based training systems will clearly require mechanisms to capture participant performance directly within the training system. However, existing commercial multi-player games, being originally designed for entertainment purposes, do not generally have the components critical for enabling monitoring of training progress and objectives fulfillment. This includes, for example, the ability to readily specify in-game performance measures as well as the ability to automatically capture low-level and high-level data that may be used to identify complex, interactive teamwork behaviors, support after action reviews, and aid trainers in making performance assessments.

Acknowledgements

This work was supported by DARPA, under the direction of Ralph Chatham, via ONR contract N00014-03-C-0279. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the Department of Defense or the US Government unless so designated by other authorized documents. We thank General Paul Gorman (US Army, retired), Harold F. O'Neil, Richard Wainess, Bruce Roberts, Len Gittleman, Lisa Spahr, Frederick Diedrich, and Elliot Entin. Particular acknowledgements are given to the soldiers who participated in the final exercise at Ft. Benning, GA.

Appendix A – Post-Exercise Questionnaire

Gorman's Gambit Exercise – Post-Exercise Questionnaire
Ft. Benning, December 2004

Participant Number: __________
Character Name/Type:____________________
Team: **A** or **B** (Circle One)

Please answer the following questions by circling the point on the line that corresponds to your answer and/or writing short answers in the space provided.		
1.	How well did you understand what your character was supposed to do?	●--------●--------●--------●--------●--------●--------●--------● Not Understood Midpoint Well Understood
2.	Regarding the game, how…	
	…exciting was it?	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
	…interesting was it?	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
	…stressful was it?	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
3.	How often did you help another teammate with a combat task?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
4.	How often were you helped by a teammate with a combat task?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
5.	How often did you pass information to another teammate during the game?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
6.	How often did a teammate pass information to you during the game?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
7.	How often did you provide non-combat resource/service to another teammate (e.g., healing spell, trap detection)?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
8.	How often did another teammate provide you with non-combat resource/service (e.g., healing spell, trap detection)?	●--------●--------●--------●--------●--------●--------●--------● Very infrequently Midpoint Very often
9.	How would you characterize the overall teamwork exhibited by your team?	●--------●--------●--------●--------●--------●--------●--------● Very Poor Midpoint Very good
10.	How would you rate the overall performance exhibited by your team?	●--------●--------●--------●--------●--------●--------●--------● Very Poor Midpoint Very good

11.	How well did you coordinate with your teammates? If midpoint or lower, why? (For example, Distractions, Difficult to do in Game, Hardware wasn't working, etc…). Use the space below for your answer.	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
12.	How well did you monitor (pay attention to) the behavior of your teammates?	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
13.	How well did you anticipate the needs of your teammates?	●--------●--------●--------●--------●--------●--------●--------● Not Very Midpoint Very
14.	Overall, how good was your situational awareness during the missions?	●--------●--------●--------●--------●--------●--------●--------● Very Bad Midpoint Very Good
15.	Overall, how would you describe your workload?	●--------●--------●--------●--------●--------●--------●--------● Very Low Midpoint Very High
16.	Overall, how would you characterize the game as a tool to learn teamwork?	●--------●--------●--------●--------●--------●--------●--------● Very Poor Midpoint Very good
17.	To what extent did playing the game help you learn how to develop and adapt strategies to accomplish your mission?	●--------●--------●--------●--------●--------●--------●--------● Not Helpful Midpoint Very Helpful
18.	Did you receive enough training to play the game well? How would you change training to make it more effective? (Use the space below to answer)	●--------●--------●--------●--------●--------●--------●--------● Not Adequate Midpoint More than Adequate
19.	This game was designed to force team members to manage limited resources (i.e., the skills of the archer or medic, limited ammunition). Were you aware of this? Yes or No (Circle One). Did you notice the lack of resources? Yes or No (Circle One). If yes, how did you overcome these limitations? (Use the space below to answer)	

20.	Although the fantasy setting of this game was not realistic, many aspects of the game were designed to have parallels to the military. Were you aware of any similarities between the team structure or the scenario and the characteristics of the Army? Yes or No (Circle One). If yes, which aspects of the game? If no, how do you think we could strengthen this relationship? (Use the space below to answer)	
21.	Overall, to what extent do you believe playing a game like this for some period of time can help you to become a better team member of a fire team, squad or platoon? Why or why not? (Use the space below to answer)	●--------●--------●--------●--------●--------●--------●--------● Not Much Midpoint Very Much
	Do you have any additional comments about the exercise? Please use the back of this page. We appreciate any thoughts you might have.	

Appendix B – AAR Questions

Gorman's Gambit Exercise – AAR Questions
Ft. Benning, December 2004

1. What aspects of the game did you like the most and why? Which would you change? How and why?
2. Can you think of any specific instances of good teamwork? What were they? What made it a good example? (use observer forms as a guide). Can you think of instances which you could have done better? What would you have done differently?
3. What difficulties did your team encounter in completing the mission? Were any of these difficulties teamwork related?
4. What were successful and unsuccessful strategies? Why do you think they were successful or unsuccessful? Did these strategies require teamwork to be successful?

5. What adjustments or changes did your team make during the scenario? Which appeared to be most effective? Why?
6. How would you improve the game?
7. What did you learn from the game?
8. Do you think what you learned in this game would help in other army tactical games? What would and what wouldn't?
9. Do you think what you learned in this game would help in actual army tactical operations (e.g., Baghdad)? What would and what wouldn't?

Appendix C – Demographic Form

Gorman's Gambit Exercise – Demographic Survey
Ft. Benning, December 2004

Please fill in the following information to the best of your ability.

Participant Number: _______________ Rank: _____________ Age: ___________

Assignment for the Demonstration:

1. Team A or B (Circle One)

Education and Military Experience

2. Formal Education (in years): _________
   (high school diploma/GED = 12, 2 years college = 14, etc.)

3. Current MOS: _______________

4. **Months** of experience in Infantry-related MOS: ______

5. **Months** of military service: ______

6. If you have been an infantry fire team leader, squad leader, platoon leader, platoon sergeant, etc., list the number of **months** you served in these leadership positions:

   a. Position: _____________ Months: ______ b. Position: ____ Months: _____
   c. Position: _____ Months: ______

7. Number of **months** of military deployment for peacekeeping, peace enforcement, stability operations or combat: _____
   If so, where? ____

Computer and Gaming Experience

8. In the past year, on average, how many **hours** per week have you used a computer for any activity (e.g., internet, school, work, etc.)?

   0 1–5 6–10 11–15 16–20 21–25 26–30 31–35 36–40 More than 40

9. In the past year, on average, how many **hours** per week have you spent playing any type of video game (e.g., PC-based, Nintendo, Playstation, arcade, etc.)?

   0 1–5 6–10 11–15 16–20 21–25 26–30 31–35 36–40 More than 40

10. Have you played the PC-based game *Neverwinter Nights*™? **YES** **NO** (Circle One)
    If yes, approximate number of hours played: __________

References

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in PC-based simulations and games. http://www.darwars.com/downloads/DARWARS Paper 12205.pdf#search=%22alexander%20 brunye%20sidman%20weil%22.

Andrews, D. H. & Bell, H. H. (2000). Simulation-based training. In S. Tobias & J. D. Fletcher (Eds), *Training and retraining: A handbook for business, industry, government, and the military* (pp. 357–384). New York, NY: Macmillan Reference.

Bonk, C. J. & Dennen, V. P. (2005). *Massive multiplayer online gaming: A research framework for military training and education* (Technical Report 2005-1). Washington, DC: Advanced Distributive Learning Initiative.

Brian, M. (2005). Stop playing games. *Military Training Technology, 10*(5). http://www.military-training-technology.com/article.cfm?DocID=1202.

Cannon-Bowers, J. A., Tannenbaum, S.I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds), *Team effectiveness and decision making in organizations* (pp. 330–380). San Francisco, CA: Jossey-Bass.

Diller, D. E., Roberts, B., Blankenship, S., & Nielson, D. (2004). DARWARS Ambush! – Authoring Lessons Learned in a Training Game. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL: I/ITSEC.

Diller, D., Roberts, B., & Willmuth, T. (2005). DARWARS Ambush! A case study in the adoption and evolution of a game-based convoy trainer by the U.S. Army. Presented at the *Simulation Interoperability Standards Organization*, 18–23 September.

Erwin, S. I. (2000, November). $65k flight simulator draws skepticism from military buyers. *National Defense*.

Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral representations for training tactical communication skills. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*, Scottsdale, AZ.

Gentner, D., Holyoak, K. J., & Kokinov, B. K. (2001). *The analogical mind: Perspectives from cognitive science*. Cambridge, MA: The MIT Press.

Gorman, P. (2003, July). Comments at DARWARS program meeting. Washington, DC.

Hollenbeck, J. R., Colquitt, J. A., Ilgen, D. R., LePine, J. A., & Hedlund, J. (1998). Accuracy decomposition and team decision making: Testing theoretical boundary conditions. *Journal of Applied Psychology, 83*, 494–500.

Hussain, T. S. & Ferguson, W. (2005). Efficient development of large-scale military training environments using a multi-player game. In *Proceedings of the 2005 Fall Simulation Interoperability Workshop* (Paper 05F-SIW-060, pp. 421–431). Orlando, FL, 18–23 September.

Nystad, E. & Strand, S. (2006). *Using virtual reality technology to include field operators in simulation and training*. Paper presented at the *Canadian Nuclear Society 27th Annual Conference* (Report No. IFE/HR/E-2006/008).

Orasanu, J. & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zambok (Eds), *Decision making in action: Models and method*. Norwood, NJ: Ablex Publishing.

Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471–499.

Serfaty, D., Entin, E. E., & Johnston, J. H. (1998). Team coordination training. In J.A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 221–245). Washington, DC: American Psychological Association.

Sims, D. E., Salas, E., & Burke, C. S. (2004). Is there a "Big Five" in teamwork? Presented at the *19th Annual Conference of the Society for Industrial and Organizational Psychology*, Chicago, IL.

Slater, M. & Steed, A. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments, 9*(5), 413–434.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 271–295). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 271–295). Washington, DC: American Psychological Association.

Weil, S. A., Hussain, T., Brunyé, T., Sidman, J., & Spahr, L. (2005a). Assessing the potential of massive multi-player games for military training. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL: I/ITSEC.

Weil, S. A., Hussain, T. S., Brunyé, T., Sidman, J., & Spahr, L. (2005b). The use of massive multi-player gaming technology for military training: A preliminary evaluation. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, Santa Monica, CA: HFES.

Weil, S. A., Hussain, T. S., Diedrich, F. J., Ferguson, W., & MacMillan, J. (2004). Assessing distributed team performance in DARWARS training: Challenges and methods. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL: I/ITSEC.

Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Witmer, B. & Singer, M. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.

Young, K. S. (2004). Internet addiction: A new clinical phenomenon and its consequences. *American Behavioral Scientist, 48*, 402–415.

THE EFFECTIVENESS AND EFFICACY OF INTELLIGENT AGENTS AS TEAM TRAINING PARTNERS IN THE ACQUISITION OF COMPLEX SKILLS IN A GAMING ENVIRONMENT

Jonathan H. Whetzel,
Sandia Laboratories

Winfred Arthur, Jr and Richard A. Volz
Texas A&M University

Abstract

We investigated the comparative effectiveness and efficacy of human and intelligent agents as team training partners in terms of trainee complex skill acquisition (on Space Fortress), affective reactions to the training, and levels of task-specific self-efficacy. Using a dyadic training protocol in which 92 participants were paired with either a human or intelligent agent, trainees engaged in two consecutive days of 2-h training sessions on Space Fortress, a research-based computer game designed to simulate a complex and dynamic aviation environment. We found no difference in performance or self-efficacy for individuals who trained with a human partner versus those who trained with an intelligent agent. Although trainees with a human partner had more favorable affective reactions, the two conditions did not differ on their training utility perceptions. These findings suggest that individuals may train with intelligent agent partners without experiencing losses in learning and self-efficacy. Thus, the use of intelligent agents can, amongst others, address some of the team composition and administrative and scheduling challenges of team training.

The objective of the present study was to investigate the efficacy and comparative effectiveness of human and intelligent agents as team training partners in terms of trainee

An earlier version of this paper was presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, Texas, May 2006.

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

complex skill acquisition (on Space Fortress), affective reactions to the training, and levels of self-efficacy. With the rapid pace of technology advancements and complexity, many work tasks now consist of cognitive and physical demands that are too diverse and complicated for most individuals to successfully accomplish single-handedly. Consequently, reliance on teams is now a pervasive reality in many military and civilian settings. With this surge in the use of teams in organizations, there has been a concurrent interest in how to efficiently and effectively train teams (e.g., Arthur, Edwards, Bell, Villado, & Bennett, 2005; Arthur, Villado, & Bennett, in press; Day et al., 2005; Salas & Cannon-Bowers, 2001; Salas, Dickinson, Converse, & Tannenbaum, 1992; Swezey & Salas, 1992).

1. Intelligent Agents as Team Training Partners

Team training can have one of two foci – the performance of the team as a collective (i.e., team performance) or the performance of individuals within the team (i.e., individual performance) with the latter more accurately described as cooperative learning instead of team training. Regardless of the foci, training partners play an important role in both team and individual performance because they can assist the trainee in learning and performing the task. Thus, for example, Day et al. (2005) reported a strong influence for trainee general mental ability on team and individual performance on Space Fortress such that at the individual level, high-ability trainees acquired more skill when paired with high-ability partners instead of low-ability partners, but low-ability trainees benefited very little from being paired with high-ability partners. High-ability teams also performed better than low-ability teams.

In addition to ability, other self or team member factors that have been shown to influence trainee learning and performance in team training contexts include aggressiveness (Bowler, Woehr, & Rentsch, 2006) and social interaction anxiety (Arthur, Young, Jordan, & Shebilske, 1996). Specifically, Bowler et al. showed that the presence of an aggressive individual was detrimental to the knowledge acquisition and performance of the trainee, as well as their partner, such that teams that included an aggressive individual demonstrated significantly lower levels of taskwork schema congruence between teammates and subject matter experts. And from the self-characteristic perspective, Arthur et al. demonstrated that some individuals may benefit more from team training than others by showing that assignment to a dyadic training protocol adversely affected the performance of high-social interaction-anxiety trainees but appeared to be advantageous for low-social interaction-anxiety trainees. In summary, the results of the preceding studies all suggest that who one trains with makes a big difference to the achievement of the specified outcomes of interest.

From a more administrative and program management perspective, whether the focus is the performance of the team or the performance of the individuals within the team, one critical challenge in team training contexts is gathering and scheduling team members for the training sessions. Scheduling team training sessions can be a challenge because the absence of any team member may make it impossible to proceed with the training session. One potential solution to both the team composition and scheduling problems is to use intelligent agents as virtual teammates. Intelligent agents are software programs that can autonomously make decisions and act to achieve goals in dynamic environments. Thus, they are goal oriented, reactive, and autonomous (Wooldridge & Jennings, 1995). Specifically, they are proactive and seek to achieve assigned goals; they are situated in dynamic environments in which they take actions to change the state to achieve their goals; and they can make decisions without human intervention. These software programs may be programmed to complete a particular task, such as operating a piece of software (e.g., games), and to model the behavior of a novice, expert, or any other level of performance. In the present study, using neural networks (Mitchell, 1997), we developed and used an agent that learned its game strategy and play by observing a human expert play and subsequently, played the game like a human expert (Whetzel, 2005).

Specifically, we used a supervised learning approach (Russell & Norvig, 1995) where we designed an agent that developed strategies for handling the task by observing a human player. Supervised learning begins by collecting training examples, data that describe scenarios that the agent will see while performing the task. Every training example the agent views has a classification attached to it. As the agent views each training example, the agent makes predictions about the example's classification. For each training example the agent classifies incorrectly, the agent makes adjustments to its learning structure. The agent continues viewing the training examples, adjusting its learning structure when necessary, until its accuracy in classifying the training examples can no longer improve. Once the agent enters the task environment, the agent is able to recognize the scenarios from the training examples and make the proper classifications. For an unknown scenario, one not included in the training examples, the agent classifies it based upon which known scenario is most analogous to the unknown one. Hence, the agent can make educated, and often correct, classifications about every scenario without exposure to all possible scenarios beforehand.

For learning human player behavior, the agent learned to classify what actions a player takes at different moments in the game. We started by recording data of the player's performance in the game environment. These data described the game environment at certain times along with the actions that the player took. From these data, we selected key features and generated training examples from them. These features included information that described the environment as seen on the game screen (e.g., distance between objects, objects present) as well as "traits" that characterize the player's behavior (e.g., reaction time, the last time an action occurred). Each training example that was produced was classified by what action the player took at that point in time. During training, the agent observed the training examples and learned what action the player took. Once placed into the game environment, the agent interpreted the input it received and determined the scenario that best fit its current state. The agent returned a classification for the scenario, the action to be performed, and executed it within the game. If trained correctly, the agent should mimic the actions that a player takes in the known scenarios. In addition, the agent should have reasonable predictions for unknown scenarios and take actions similar to what the human player would do at those times.

Like human partners, intelligent agents can also serve as tutors who assist trainees in learning and performing the task. They have a potential for training that goes beyond traditional intelligent tutoring systems (Anderson, Boyle, Corbett, & Lewis, 1990) because they are able to dynamically interact with the problem-solving environment, actively participating with trainees to solve the specified problem (Ioerger, Sims, Volz, Workman, & Shebilske, 2003). Consequently, if intelligent agents could effectively replace human partners in team training contexts, then there are several advantages that would accrue from this. However, the critical question is whether intelligent agents can be as effective as human partners in terms of trainees' complex skill acquisition and learning, their affective reactions to training, and their levels of task-specific self-efficacy.

A major difference between human and intelligent agent team training partners is the absence of social contact associated with the latter. On the basis of social learning theories (e.g., Bandura, 1986), this absence of social contact might lead one to posit that intelligent agents may be less effective than humans as team training partners. However, this may also be mitigated by certain facets of social contact still present or available with intelligent agents such as the opportunity to still observe and model a partner's task strategy and performance and the absence of evaluation apprehension (Arthur et al., 1996) when paired with an intelligent agent. In addition, assuming that the performance level of the agent is within the trainee's zone of proximal development (Vygotsky, 1978, 1987) the tenets of scaffolding might be sufficient to compensate for the loss of social contact. Scaffolding refers to a variety of instructional approaches that involve collaboration and providing just enough assistance to extend the trainee's current capabilities. Assistance can come in the form of providing hints or asking probing questions but, generally, scaffolding refers to modeling advanced response patterns that are within reach of the trainee's current capabilities.

The challenge then is to design an agent whose performance level is higher than the trainee's but still within their zone of proximal development to capitalize on the advantageous effects of scaffolding. However, it is also important to note that given the current state of the artificial intelligence architecture and technologies, the use of intelligent agents in training focuses almost exclusively on the taskwork (i.e., behaviors involved in the execution of team tasks; these behaviors are typically task specific) and not teamwork (i.e., non-task-specific general behaviors required for cooperative functioning and focuses on team process variables such as communication, team cohesion, team spirit, and morale) facets of team training and performance (Arthur, Villado, & Bennett, in press).

2. Reaction and Behavioral/Performance Training Evaluation Criteria

The choice of training evaluation criteria (i.e., the dependent measure used to operationalize the effectiveness of training) is a primary decision that must be made when evaluating the effectiveness of training – in this particular instance, the comparative effectiveness of intelligent agents and humans as team training partners. Although newer approaches to, and models of, training evaluation have been proposed (e.g., Day, Arthur, & Gettman, 2001; Kraiger, Ford, & Salas, 1993), Kirkpatrick's (1976) four-level model of training evaluation and criteria continues to be the most popular (Arthur, Bennett, Edens, & Bell, 2003). We used this framework in the present study because it is conceptually the most appropriate for our purposes. In addition, given the nature of the study, specifically a lab-based study, our evaluation criteria were limited to reaction and behavioral/performance criteria.

Reaction criteria represent trainees' feelings, impressions, or attitudes to training – in the present study, being paired with either a human or intelligent agent team training partner. Thus, reaction criteria can be conceptualized in terms of trainees' affective reactions toward the training (how much they liked it) and their perceptions of the usefulness or value of the training. We assessed both affect and perceptions of utility in the present study. Although there is very little empirical reason to believe that how trainees feel about a training program or whether they like it tells researchers much about how much they learned from the training and their subsequent performance (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997; Arthur, Tubre, Paul, & Edens, 2003; Noe & Schmitt, 1986), information about how trainees' feel about a training program or intervention can be useful. For instance, it can be used to inform future attempts to recruit individuals into and retain them in training.

Behavioral/performance criteria, in contrast to reaction criteria, pertain to actual measures of task or job performance. Thus, whereas reaction criteria are operationalized via trainee self-reports, behavioral/performance criteria are based on assessments of trainee task or job performance. In the present study, we operationalized this criterion as performance on Space Fortress and we compared the performance of trainees paired with an intelligent agent to those paired with a human team training partner.

3. Self-Efficacy

A third criterion that we used to assess the comparative effectiveness of human and intelligent agents as team training partners was task-specific self-efficacy. Self-efficacy is defined as individuals' beliefs about their capabilities to produce designated levels of performance on a specified task (Bandura, 1994). Historically, the relationship between self-efficacy and training performance has been explained from a social cognitive theoretical perspective based on the evaluative and agentive properties of human self-regulation (Bandura, 1997). Consistent with this, the relationship between self-efficacy and task performance would seem to have been extensively documented (Bandura & Locke, 2003; Sadri & Robertson, 1993; Stajkovic & Luthans, 1998). Conceptually, it is posited that high self-efficacy causes individuals to set higher goals, increases their goal commitment, and subsequently results in increases in performance (Bandura, 1997; Bandura & Wood, 1989; Locke & Latham, 1994). However, an accumulating body of recent empirical research has espoused the counter position that self-efficacy is redundant with prior performance (e.g., Ackerman, Kanfer, & Goff, 1995; Arthur, Bell, & Edwards, in press; Heggestad &

Kanfer, 2005; Richard, Dieffendorf, & Martin, 2006; Vancouver, Thompson, Tischner, & Putka, 2002; Vancouver, Thompson, & Williams, 2001). Specifically, it would seem that the observed relationship between self-efficacy and performance is due to the effect of past performance on self-efficacy instead of self-efficacy on subsequent performance. Consequently, from this perspective, self-efficacy can be used as a training effectiveness criterion such that in the present study, we investigated the differential effects of training with an intelligent agent versus a human partner on trainees' levels of post-training self-efficacy.

4. Performance Task Platform and Training Protocol

The performance task platform used in the present study was Space Fortress, a video game-based task that has a long history as an excellent research tool in the area of complex skill acquisition (Donchin, 1989; Gopher, 1993; Mane & Donchin, 1989). Specifically, Space Fortress is "an experimental game which was designed to simulate a complex and dynamic aviation environment" (Gopher, 1993, p. 299). Space Fortress (Gopher, 1993; Shebilske, Regian, Arthur, & Jordan, 1992) is a complex perceptual-motor skill task that represents important information processing demands that are present in aviation and other complex tasks (Gopher, Weil, & Bareket, 1994; Hart & Battiste, 1992). These processing demands include short- and long-term memory load, high workload, and dynamic attention allocation, decision-making, prioritization, resource management, discrete motor responses, and difficult manual control elements (Gopher, Weil, & Siegel, 1989).

The training protocol used in the present study was the active interlocked modeling (AIM)-dyad protocol (Shebilske et al., 1992). The AIM-dyad protocol requires trainees to simultaneously perform each half of the Space Fortress task components (i.e., pilot-gunner functions versus copilot mine-missile manager functions) while being interlocked with a teammate who performs the other half. Therefore, the AIM-dyad protocol requires that team members share taskwork, creating a highly interdependent team training environment so that successful overall performance requires the proper coordination and effective performance of all components (e.g., Day et al., 2001). For instance, the actions of a team member require reactions from the other and vice versa. Task workflow ratings provide additional evidence for the high interdependency of Space Fortress task components (Arthur, Villado, & Bennett, in press). Day et al. (2005) have also shown that due to the high interdependency (resulting from the fact that partners are yoked in the AIM-dyad protocol), a trainee's learning and performance is influenced by the ability and performance level of their teammate. From an applied perspective, the AIM-dyad protocol's time and resource-saving efficiencies served as the impetus for the design of innovative team training protocols for Israeli Air Force pilots, US navigators, and Aer Lingus airline pilots (Johnston, Regian, & Shebilske, 1995; Shebilske, Goettl, & Regian, 1999).

In summary, the objective of the present study was to investigate the comparative effectiveness and efficacy of intelligent agents and humans as team training partners on individual trainee complex skill acquisition, affective reactions to the training, and levels of task-specific self-efficacy. The intelligent agent used in our study was trained to model the performance of an expert. The study objective was accomplished using Space Fortress (Gopher et al., 1994) as the task performance platform and the AIM-dyad protocol (Shebilske et al., 1992) as the training protocol. Therefore, our protocol used a dyadic training team and a game which was designed to emulate combat aviation environments such as F-16B/D and F-15B/D fighter aircraft teams which, like Space Fortress, comprise specialized two-person roles. The research design consisted of two between-subject partner conditions – human and intelligent agent.

5. Method

5.1. Participants

The final study sample (i.e., individuals for whom complete data were available on all variables) consisted of 92 volunteers (27 females) recruited from a large southwestern university using advertisements in the university newspaper, announcements in classrooms, and posted notices around campus. Participants were paid $7.50 per hour for 5 h of participation. A cash incentive of an extra $20 was also offered and awarded to individuals whose scores were in the top quartile. The mean age of the participants was 19.62 years ($SD = 2.30$).

5.2. Measures

Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). Because general mental ability has been shown to be predictive of Space Fortress performance (e.g., Day et al., 2005), the training conditions were matched on this variable. This was done on the basis of the Advanced Progressive Matrices scores. The Advanced Progressive Matrices is a measure of general mental ability consisting of 36 design problems arranged in an ascending order of difficulty. We used an administration time of 40 min and Advanced Progressive Matrices scores were obtained by summing the number of problems correctly solved. We obtained a Spearman–Brown odd–even split-half reliability of 0.82 for the Advanced Progressive Matrices scores.

Performance Task – Space Fortress. The trainee performance task was Space Fortress, a video game-based task (Arthur, Day, Bennett, McNelly, & Jordan, 1997; Day et al., 2001; Mane & Donchin, 1989). Trainees had their own task station which consisted of an IBM PC compatible computer, a high-resolution color monitor, a right-hand joystick, and a three-button mouse for the left hand. In Space Fortress, trainees control a space ship's flight path using a joystick and shoot missiles with a trigger on the joystick (pilot-gunner functions). A fortress was located center-screen with two concentric hexagons surrounding it. An information panel at the bottom of the screen indicated the fortress vulnerability which changed with each missile hit. Trainees used the three-button mouse to perform functions associated with mines and bonus opportunities (copilot mine-missile manager). Pressing the right-button twice with an interpress interval of 250–400 ms was necessary to "prime" foe mines before they could be destroyed. Bonus opportunities were available whenever the second of a pair of "$" symbols appeared on-screen. Points or missile bonus could be selected at this time by pressing either the left- or right-button, respectively.

Mines appeared every 4 s and remained on-screen for 10 s unless they were destroyed by the ship, collided with the ship, or until the ship was hit by the fortress. A single fortress hit damaged the trainees' ship with a concomitant loss of 50 points; four hits destroyed the ship with a concomitant loss of 100 points. While mines were on the screen, the fortress was invulnerable to the ship's missiles.

Before each 3 min game, trainees memorized three computer-generated letters which identified foe mines. When a mine appeared on-screen a letter always appeared in the identify-friend-or-foe (IFF) indicator. If a letter indicated a foe mine, trainees had to press the IFF button twice within an interpress interval of 250–400 ms. The information panel displayed this interval. If the interval was correct, one missile hit could destroy a mine; if incorrect, the ship's missiles were ineffective against mines. However, trainees could try again until a correct interval was achieved. Mines continually chased the ship attempting to destroy it by colliding with it. If the IFF button indicated a friendly mine, trainees were to avoid pressing the IFF button. If the IFF button was pressed, the friendly mine would become a deadly enemy mine incapable of being destroyed. If the IFF button was not pressed, the friendly mine pursued the ship waiting to be energized by a ship missile, subsequently scoring points and increasing the fortress vulnerability counter by one. A friendly mine could destroy the ship if it was not energized and collided with the ship.

When mines were not on the screen, the ship's missiles could damage and eventually destroy the fortress. Each of the first 10 missile hits on the fortress increased its vulnerability. After 10 hits, it could be destroyed by a double shot, which had to have an interpress interval of 250 ms or less. If a double shot hit the fortress before the 10th hit, the vulnerability was reset to zero.

The ship started with a main supply of 100 missiles. Trainees could fire more missiles after the main supply was depleted at a cost of three points per missile. The main missile supply could be replenished during bonus intervals, which were indicated by the second of two consecutive $ symbols appearing below the fortress. Symbols changed every 4 s and consisted of other symbols in addition to the $ symbol. The $ symbol always appeared in consecutive sets of two. If a bonus button was pressed during the first of the pair, no bonus was delivered, and the bonus buttons were deactivated during the second $ symbol. However, if the trainee waited until the second $ symbol appeared they could press the missile bonus button to receive extra missiles. If more than 50 missiles were remaining, their total would be restored to 100. If less than 50 missiles were remaining, 50 more would be added. Alternatively, the trainee could press the points bonus button to obtain 100 points during the bonus interval.

The spaceship flew in frictionless space so a thrust in one direction would move the ship at a constant velocity in that direction until another thrust was applied. Thrusts in the same direction accelerated the ship, in the opposite direction slowed the ship, and in other directions changed the ship's course. Pushing the joystick forward applied thrust, moving the joystick left or right rotated the ship, and pulling back on the joystick did nothing. If the ship left the screen, it wrapped and reappeared on the opposite side of the screen following the same trajectory.

As summary information, the information panel at the bottom of the screen showed the number of available missiles, a battle score, and component scores based on ship velocity, ship control, and the speed of dispatching mines. At the end of each game, the screen displayed a total score which was a composite of the subscores; this score was used as the measure of Space Fortress performance in the present study. Although participants trained (practiced) with either a human or intelligent agent partner using the AIM-dyad protocol (Arthur et al., 1997; Shebilske et al., 1992; also see *Procedure* section below), their test sessions were performed *alone*. Consequently, Space Fortress performance scores were operationalized as participants' total scores on their test sessions.

Intelligent Agent. Machine learning techniques were used to develop an agent that learned its Space Fortress game strategy from a human expert player. Consequently, using neural networks, this agent was trained to play the game like a human expert. Specifically, the agent used in the present study was designed based on a learning structure that uses the analogy of biological learning systems for finding the classification hypotheses within a training data set. Neural networks (Mitchell, 1997) use a series of interconnected units, called neurons, with each unit taking a number of real-valued inputs and outputting a single real-valued number. Each neuron contains a set of weight values, with a weight assigned to a particular input. After combining these neurons together into a network, the network begins receiving the training example's attributes as input into the network. As the network observes each example, the neuron weights tune themselves such that the network output maps to the proper classification. After several iterations through the data, the weights within the network represent the classification hypotheses for the training data. The network represents a function that, given any example, should produce the correct corresponding classification output. A detailed description of the development of the agent can be found in Whetzel (2005).

Self-Efficacy Measure. Participants' levels of efficacy for performing Space Fortress was assessed using a six-item self-efficacy measure that was developed (see; Arthur, Bell, & Edwards, in press; Arthur et al., 2006) following principles and guidelines recommended by Bandura (1997) for developing self-efficacy scales. In the present study, coefficient alphas of 0.94 and 0.95 were obtained for the self-efficacy scores for the Time 1 and Time 2 administrations, respectively.

Partner Reaction Measure. We developed a 10-item measure to assess trainees' reactions to their training partner. The measure consisted of five utility and five affect items and were responded to on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Coefficient alphas of 0.79 and 0.77 were obtained for the utility, and affect ratings, respectively. The alpha for the composite ratings (i.e., all 10 items) was 0.88.

5.3. Design and Procedure

The experimental design was a 1-way (i.e., human versus intelligent agent partner) repeated measures (multiple training sessions) design. Participation involved two consecutive days of 2-h training sessions. Table 1 provides an overview and summary of the research protocol. On the first day, trainees began with 20 min of video-taped instructions which explained the rules and optimal strategies to Space Fortress. Participants also received a reference packet (that was available to them throughout the study) which covered the same material presented in the video. Trainees then performed four 3-min baseline Space Fortress games followed by a 5-min review of the instructions. Following the review, trainees completed the Time 1 measure of task-specific self-efficacy. Trainees were then assigned to their specified training conditions ($n = 40$ and 52 for human and intelligent agent partner conditions, respectively) and completed training as presented in Table 1. Following the completion of training, trainees completed the second measure of task-specific self-efficacy followed by the reaction questionnaire. Because of well-documented sex-based performance differences in Space Fortress (Sanchez-Ku & Arthur, 2000), an attempt was also made to balance the training conditions on their sex composition. Furthermore, all human partner teams were homogeneous sex teams. Table 2 presents the ability and sex composition of the study sample.

Training was conducted using the AIM-dyad protocol (Arthur et al., 1997; Shebilske et al., 1992). For the human training partner condition, participants were seated next to their partner at an adjacent computer workstation. Both players saw the same game screen on their monitors. They also had their own computer along with their own joystick and mouse set for inputs. During a standard training session, trainees performed eight team practice and two individual test games. That is, practice games were performed as a team

Table 1
Overview of training and data collection procedures.

Day	Activity
Day 1 (3 h)	1. Informed consent and experiment overview 2. Video-taped Space Fortress instructions 3. Baseline (4) games 4. Video-taped Space Fortress summary instructions 5. Self-efficacy measure (Time 1) 6. Assignment into conditions 7. Practice (8) and test (2) games (Session 1)
Day 2 (2 h)	1. Practice (8) and test (2) games (Session 2) 2. Practice (8) and test (2) games (Session 3) 3. Practice (8) and test (2) games (Session 4) 4. Self-efficacy measure (Time 2) 5. Reaction measure

Practice games were team games and test games were individual games.

Table 2
General mental ability and sex composition of study sample.

Human partner				
		Male	Female	
	High ability	19	6	$n = 40$
	Low ability	9	6	APM mean = 27.90 (SD = 3.50)
Intelligent agent				
		Male	Female	
	High ability	22	7	$n = 52$
	Low ability	15	8	APM mean = 26.46 (SD = 5.05)

APM = Raven's Advanced Progressive Matrices. High ability = APM score is equal to or greater than sample median of 26.70 (SD = 4.70).

and test games were preformed individually. All games lasted 3 min. For each practice game one trainee, using his/her left hand, controlled all functions related to the mouse (mine-missile manager), and the other trainee, using his/her right hand, controlled all functions related to the joystick and trigger (pilot-gunner). Trainees alternated roles at the end of each practice game. Communication between trainees was encouraged. For test games, trainees played the two test games by themselves, simultaneously controlling all aspects of the game using the mouse and joystick. Participants' Space Fortress session performance was operationalized as the average of the total scores from the two test games.

The intelligent agent condition was identical to the human partner condition with the exception that during the practice sessions, one set of functions was controlled by the agent while the other was controlled by the trainee.

6. Results

Consistent with Sanchez-Ku and Arthur (2000) and Day et al. (2005) there were significant sex, $F(1, 88) = 23.28$, $p < 0.001$, $\eta^2 = 0.21$, and general mental ability effects, $F(1, 88) = 17.60$, $p < 0.001$, $\eta^2 = 0.16$, such that males performed better than females, and higher ability trainees performed better than lower ability trainees. However, because their interactions with partner conditions were not significant, $F(1, 88) = 0.54$, $p > 0.05$, $\eta^2 = 0.00$, and $F(1, 88) = 1.43$, $p > 0.05$, $\eta^2 = 0.01$, these variables were not included in subsequent analyses.

6.1. Comparative Effectiveness of Intelligent Agent and Human Partners

For descriptive purposes, the intercorrelations amongst all the study variables are presented in Table 3. Table 4 presents the Space Fortress test session descriptive statistics and

Table 3
Descriptive statistics and intercorrelations amongst study variables.

Variable		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	SF Baseline	—													
2.	SF Session 1	0.85	—												
3.	SF Session 2	0.67	0.81	—											
4.	SF Session 3	0.69	0.86	0.88	—										
5.	SF Session 4	0.65	0.83	0.85	0.95	—									
6.	Average score[a]	0.75	0.92	0.94	0.97	0.96	—								
7.	Self-efficacy T1	0.50	0.43	0.43	0.36	0.37	0.42	—							
8.	Self-efficacy T2	0.35	0.48	0.63	0.61	0.68	0.64	0.62	—						
9.	Affect	−0.19	−0.18	−0.15	−0.14	−0.10	−0.15	−0.06	0.01	—					
10.	Utility	−0.16	−0.13	−0.05	−0.04	0.03	−0.05	0.13	0.20	0.79	—				
11.	Overall reaction	−0.19	−0.17	−0.11	−0.10	−0.04	−0.11	0.04	0.11	0.95	0.95	—			
12.	Partner condition[b]	−0.06	−0.04	−0.05	−0.08	−0.03	−0.05	−0.05	0.13	−0.23	−0.16	−0.20	—		
13.	GMA[c]	0.44	0.43	0.36	0.42	0.42	0.43	0.10	0.19	0.00	0.05	0.03	−0.16	—	
14.	Sex[d]	0.47	0.48	0.41	0.39	0.36	0.43	0.43	0.25	−0.19	−0.09	−0.15	0.01	0.11	—
	Mean	−2367.00	−503.49	411.36	886.63	1304.00	524.75	3.11	3.71	3.81	3.80	3.81	—	27.09	—
	SD	1252.00	2120.00	2191.00	2219.00	2212.00	2070.00	1.02	0.99	0.80	0.80	0.76	—	4.48	—

SF = Space Fortress. All correlations larger than 0.34 are significant at $p < 0.001$.
[a] Average score = average of four Space Fortress test sessions.
[b] Partner condition, human partner = 0, intelligent agent = 1.
[c] GMA = general mental ability.
[d] For sex, females = 0, males = 1.

Table 4
Descriptive statistics and standardized mean differences between human partner and intelligent agent conditions on all Space Fortress Test Sessions.

Space Fortress Test Session	Human partner		Intelligent agent		
	Mean	SD	Mean	SD	d[b]
Baseline	−2286.22	1382.18	−2430.00	1152.18	−0.11
Session 1	−415.11	2096.72	−571.48	2155.08	−0.07
Session 2	547.48	2004.79	306.66	2338.41	−0.11
Session 3	1098.18	2161.88	723.89	2269.90	−0.17
Session 4	1378.98	2184.38	1247.19	2252.16	−0.06
Average score[a]	652.38	2013.38	426.57	2126.62	−0.11

[a] Average of four test session scores.
[b] In computing the ds, the intelligent agent condition ($n = 52$) was treated as the experimental condition and the human partner condition ($n = 40$) as the control. None of the ds were statistically significant.

standardized mean differences (ds) between the human partner and intelligent agent conditions. (These results are also illustrated in Figure 1.) Two-tailed univariate significance tests are also presented. Results at the baseline indicated that the initial Space Fortress performance of the intelligent agent and human partner trainees was not significantly

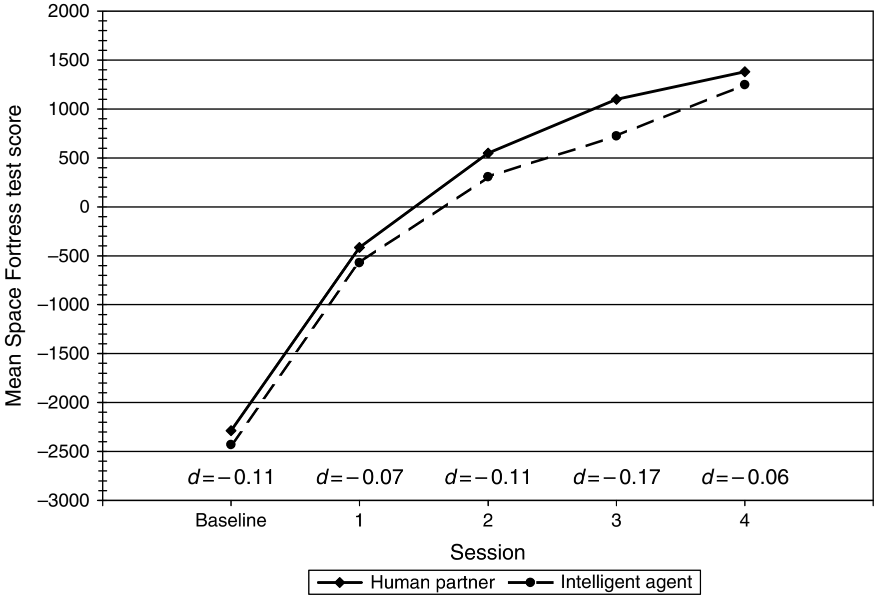

Figure 1. Mean Space Fortress test scores for human and intelligent agent partner conditions across training sessions along with standardized mean differences at each session.

different prior to their assignment into the training conditions, $F(1,\ 90) = 0.30$, $p > 0.05$, $d = 0.25$. In addition, using a 2×4 mixed analysis of covariance (ANCOVA), with baseline performance as the covariate, we obtained a nonsignificant between-subjects main effect indicating that participants in the human and intelligent agent conditions did not differ in their Space Fortress performance, $F(1,\ 89) = 0.03$, $p > 0.05$, $\eta^2 = 0.00$. Furthermore, we obtained a significant within-subjects effect for sessions, $F(3,\ 267) = 8.71$, $p < 0.001$, $\eta^2 = 0.09$, indicating a positive effect for training over time. Specifically, with a d of 1.82, the mean performance (i.e., 50th percentile) at Session 4 was equivalent to 97th percentile on the baseline. However, the partner conditions were not differentially effective in improving performance over sessions; that is, the Condition $\times$ Session interaction term was not significant, $F(3,\ 267) = 0.46$, $p > 0.05$, $\eta^2 = 0.00$.

6.2. Affective Reactions to Training

Table 5 presents the partner reactions descriptive statistics and ds between the human partner and intelligent agent conditions. Two-tailed univariate significance tests are also presented. The results indicate that although trainees in the human partner condition had more favorable affective reactions to their partners ($d = -0.46$, $p < 0.01$), the two conditions did not differ in their perceptions of utility ($d = -0.31$, *ns*). In addition, it is worth noting that even for the intelligent agent condition, the average ratings were all above the mid-point of the rating scale.

Table 5
Descriptive statistics and standardized mean differences for partner reaction ratings.

	Human partner		Intelligent agent		
	Mean	SD	Mean	SD	d^b
Affect	4.01	0.70	3.65	0.83	−0.46*
Utility	3.95	0.74	3.70	0.84	−0.31
Overall reaction[a]	3.98	0.69	3.67	0.78	−0.42*

[a] Average of affect and utility ratings.
[b] In computing the ds, the intelligent agent condition ($n = 52$) was treated as the experimental condition and the human partner condition ($n = 40$) as the control.
* $p < 0.01$ (two-tailed univariate test).

6.3. Effects of Human and Intelligent Agent Partners on Self-Efficacy

Table 6 presents the self-efficacy descriptive statistics and ds between the human partner and intelligent agent conditions. Two-tailed univariate significance tests are also presented. Using a 2×2 mixed analysis of variance (ANOVA) we obtained a nonsignificant between-subjects main effect indicating that participants in the human and intelligent

Table 6
Descriptive statistics and standardized mean differences for self-efficacy ratings.

	Human Partner		Intelligent agent		
Self-efficacy	Mean	SD	Mean	SD	d^a
Time 1	3.17	1.02	3.06	1.02	−0.11
Time 2	3.85	1.00	3.60	0.98	−0.25
d^b	0.67*		0.54*		

[a] In computing the between-groups ds, the intelligent agent condition ($n = 52$) was treated as the experimental condition and the human partner condition ($n = 40$) as the control.
[b] In computing the within-groups ds, Time 1 ratings were subtracted from Time 2 ratings.
* $p < 0.01$ (two-tailed univariate test).

agent conditions did not differ in their ratings of self-efficacy, $F(1,\ 90) = 0.91,\ p > 0.05,\ \eta^2 = 0.01$. Conversely, we observed a significant within-subjects effect for sessions, $F(1,\ 90) = 43.10,\ p < 0.001,\ \eta^2 = 0.32$, indicating an increase in self-efficacy from Time 1 to Time 2. However, self-efficacy did not increase differentially between partner conditions; that is, the Condition × Session interaction was not significant, $F(1,\ 90) = 0.59,\ p > 0.05,\ \eta^2 = 0.00$.

7. Discussion

The data presented here provide evidence for the potential utility of intelligent agents as team training partners. In spite of the absence of social contact and other facilitating factors, there were no performance differences between individuals who trained with a human partner and those who trained with an intelligent agent. Furthermore, both self-efficacy and perceived utility were not differentially affected by the partner condition. Specifically, individuals who trained with an intelligent agent reported levels of task-specific self-efficacy and perceived utility that were similar to those of individuals who trained with a human partner.

However, the absence of social contact and other facilitating factors seem to be reflected in the finding that affective ratings were higher in the human partner condition than the intelligent agent condition. Nevertheless, our data overall suggest that intelligent agents can be a potential substitute for human partners to serve as virtual teammates without any learning loss to the trainee or diminished levels of task-specific self-efficacy. Indeed, within the context of scaffolding and zone of proximal development, since our intelligent agent performed the task at the level of an expert, one could go a step further by manipulating the ability or performance level of the intelligent agent to further increase its potential effectiveness as a tutor and learning partner (Day et al., 2005). Specifically, one could conceivably vary the expertise level of the agent across the phases of skill acquisition to capitalize on the benefits of scaffolding that could possible result in comparatively higher acquisition with an intelligent agent partner compared to a human partner.

Our results suggest that individuals may be trained in a team context by replacing a human partner with an intelligent agent with no decrements in individual performance or self-efficacy. Although it appears that individuals may prefer to train with a human partner, decrements in affective reactions associated with training with an intelligent agent may not be large enough to warrant concern. That is, trainees may prefer training with a human, but training with an intelligent agent may still elicit acceptable trainee reactions.

7.1. Implications, Limitations, and Directions for Future Research

This study defined the skill acquisition phase as a fixed amount of training time (i.e., 4 h of training) as opposed to a specified level of performance (e.g., one errorless trial, three errorless trials, or asymptote). However, training to a specified level of performance could display differential effects not present in our data. Specifically, training to a specified level of performance would necessitate longer skill acquisition times. And with longer training times, it is plausible that differences not present in our data may emerge (cf. Edwards, Day, Arthur, & Bell, 2006). Thus, additional follow-up studies investigating whether the effects reported in this study are stable across different skill acquisition criteria and across lengthier training protocols are an important avenue of research. Relatedly, future research could also address a common limitation in the skill acquisition literature by investigating the comparative effectiveness of human and intelligent agent team training partners in terms of resilience to skill decay and enhancement of transfer (Arthur, Bennett, Stanush, & McNelly, 1998; Arthur et al., 1997; Schmidt & Björk, 1992).

A second point worth noting is our focus on taskwork as opposed to teamwork. Teamwork refers to the team's efforts to facilitate interaction among team members in the accomplishment of team tasks whereas taskwork refers to the team's efforts to understand and perform the requirements of the job, tasks, and equipment to be used. However, because we focused exclusively on taskwork, our data do not speak to the role of intelligent agents in the development of teamwork. It is conceivable that the effectiveness of one type of training partner (human versus intelligent agent) may be a function of whether the training focus is on taskwork, teamwork, or some combination of the two. Nevertheless, we have demonstrated that when taskwork is the focus of training, both intelligent agents and human partners are equally effective in terms of trainee learning and task-specific self-efficacy. Future research could focus on investigating the comparative effectiveness of intelligent agents used to train teamwork skills in a team training context.

Finally, the proportion of humans to intelligent agents may be an important consideration when opting to employ intelligent agents in team training. Again, it may be possible that the effectiveness of intelligent agents may be limited to a range of proportions within any one team. We focused on dyads, thus resulting in only two proportions (i.e., 0, and 50% intelligent agents). Future research may investigate the effectiveness of training teams with teams consisting of various proportions of intelligent agents.

8. Conclusions

A major administrative challenge or hurdle to team training is the scheduling of multiple individuals for team training session. Specifically, the inability to schedule or the absence of one team member may make it impossible to hold the training session. In addition, who one trains with plays a big role in the amount of learning and performance attained by the trainee (Arthur et al., 1996; Bowler et al., 2006; Day et al., 2005). Thus, these issues collectively pose a challenge not only in terms of administratively scheduling training, but also in situations where the performance of individuals in team-based training and learning contexts is used to make decisions pertaining to grades, selection, promotions, and other desirable rewards. A potential solution to these challenges is to substitute intelligent agents for human partners in team training. Consequently, our study investigated the viability of replacing a human partner with an intelligent agent by examining the comparative effects of human and intelligent agent team training partners on trainee complex skill acquisition, affective reactions to the training, and the trainee's level of task-specific self-efficacy. Our results indicated that the type of team training partner did not affect the level of skill acquisition. Furthermore, there was no difference in ratings of self-efficacy or perceived utility between trainees who trained with an intelligent agent and those who trained with a human partner. Finally, although trainees who were paired with human partners had more favorable affective reactions to the training than trainees paired with intelligent agent partners, the average affective ratings were above the mid-point of the rating scale. Our results suggest that intelligent agents may serve as effective surrogates for human partners in team training contexts with no loss in individual performance or self-efficacy. However, trainees may prefer to train with a human partner as reflected in their affective reactions to training.

Acknowledgment

Funding for this research was provided by DoD MURI grant #F49620-00-1-0326 and the National Physical Science Consortium Fellowship.

References

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1*, 270–303.

Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr, Traver, H., & Shotland, A. (1997). A meta-analysis of relations among training criteria. *Personnel Psychology, 50*, 341–358.

Anderson, J., Boyle, C., Corbett, A., & Lewis, M. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence, 42*, 7–49.

Arthur, W. Jr, Bell, S. T., & Edwards, B. D. (2007). A longitudinal examination of the comparative criterion-related validity of additive and referent-shift consensus operationalizations of team efficacy. *Organizational Research Methods*, *10*, 35–58.

Arthur, W., Jr, Bennett, W., Jr, Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*, 234–245.

Arthur, W., Jr, Bennett, W., Jr, Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*, 57–101.

Arthur, W., Jr, Day, E. A., Bennett, W., Jr, McNelly, T. L., & Jordan, J. A. (1997). Dyadic versus individual training protocols: Loss and reacquisition of a complex skill. *Journal of Applied Psychology, 82*, 783–791.

Arthur, W., Jr, Edwards, B. D., Bell, S. T., Villado, A. J., & Bennett, W., Jr. (2005). Team task analysis: Identifying tasks and jobs that are team-based. *Human Factors, 47*, 654–669.

Arthur, W., Jr, Tubre, T. C., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology, 23*, 275–285.

Arthur, W., Jr, Villado, A. J., & Bennett, W., Jr. (in press). Innovations in team task analysis: Identifying task elements, tasks, and jobs that are team-based. In W. Bennett, Jr (Ed.), *The future of job analysis*. Mahwah, NJ: LEA.

Arthur, W., Jr, Villado, A. J., Boatman, P. R., Bhupatkar, A., Kowollik, V., Day, E. A., & Bennett, W., Jr. (2006). *Standards for readiness assessment and tracking of warfighter training systems research*. The Defense Advanced Research Project Agency (DARPA), and Air Force Research Laboratory Human Effectiveness Directorate (AFRL/HE), Mesa, AZ. Contract No. F41624-97-D-5000, L-3 Communications Corporation, Link Simulation and Training Division, Mesa, AZ.

Arthur, W., Jr, Young, B., Jordan, J. A., & Shebilske, W. L. (1996). Effectiveness of individual and dyadic training protocols: The influence of trainee interaction anxiety. *Human Factors, 38*, 79–86.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). New York: Academic Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*, 87–99.

Bandura, A. & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology, 56*, 805–814.

Bowler, M. C., Woehr, D. J., & Rentsch, J. R. (2006, August). *The impact of aggressive individuals on team training*. Paper presented at the 66th Annual Meeting of the Academy of Management, Atlanta, Georgia.

Day, E. A., Arthur, W., Jr, Bell, S. T., Edwards, B. D., Bennett, W., Jr, Mendoza, J. L., & Tubre, T. C. (2005). Ability-based pairing strategies in the team-based training of a complex skill: Does the intelligence of your training partner matter? *Intelligence, 33*, 39–65.

Day, E. A., Arthur, W., Jr, & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*, 1022–1033.

Donchin, E. (1989). The learning strategies project: Introductory remarks. *Acta Psychologica, 71*, 1–15.

Edwards, B. D., Day, E. A., Arthur, W., Jr, & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology, 91*, 727–736.

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. E. Meyer & S. Kornblum (Eds). *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 299–322). Cambridge, MA: MIT Press.

Gopher, D., Weil, M., & Bareket, T. (1994). The transfer of skill from a computer game trainer to actual flight. *Human Factors, 36*, 387–405.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147–177.

Hart, S. G. & Battiste, V. (1992). Field test of a video game trainer. *Proceedings of the Human Factors Society 36th Annual Meeting*, (pp. 1291–1295). Santa Monica, CA: Human Factors Society.

Heggestad, E. D. & Kanfer, R. (2005). The predictive validity of self-efficacy in training performance: Little more than past performance. *Journal of Experimental Psychology: Applied, 11*, 84–97.

Ioerger, T. R., Sims, J., Volz, R. A., Workman, J., & Shebilske, W. L. (2003). *On the use of intelligent agents as partners in training systems for complex tasks*. Paper presented at the *25th Annual Meeting of the Cognitive Science Society, CogSci*, Boston, MA.

Johnston, A. N., Regian, J. W., & Shebilske, W. L. (1995). Observational learning and training of complex skills in laboratory and applied settings. In N. Johnston, R. Fuller, & N. McDonald (Eds), *Proceedings of the 21st Conference of the European Association of Aviation Psychology: Vol. 2. Aviation psychology: Training and selection* (pp. 316–323). Aldershot: Avebury Aviation.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (2nd ed., pp. 301–319). New York: McGraw-Hill.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311–328.

Locke, E. A. & Latham, G. P. (1994). Goal setting theory. In H. F. O'Neil & M. Drillings (Eds), *Motivation: Theory and research* (pp. 13–29). Hillsdale, NJ: Erlbaum.

Mane, A. M. & Donchin, E. (1989). The Space Fortress game. *Acta Psychologica, 71*, 17–22.

Mitchell, T. (1997). Artificial neural networks. In E. M. Munson (Ed.), *Machine learning* (pp. 81–105). Boston, MA: McGraw-Hill.

Noe, R. A. & Schmitt, N. M. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology, 39*, 497–523.

Raven, J. C., Raven, J., & Court, J. H. (1998). *A manual for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis.

Richard, E. M., Diefendorff, J. M., & Martin, J. H. (2006). Revisiting the within-person self-efficacy and performance relation. *Human Performance, 19*, 551–559.

Russell, S. & Norvig, P. (1995). *Artificial intelligence: A modern approach*. Upper Saddle River, NJ: Prentice-Hall.

Sadri, G. & Robertson, I. T. (1993). Self-efficacy and work-related behavior: A review and meta-analysis. *Applied Psychology: An International Review, 42*, 139–152.

Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471–499.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Sanchez-Ku, M. L. & Arthur, W., Jr. (2000). A dyadic protocol for training complex skills: A replication using female participants. *Human Factors, 42*, 512–520.

Schmidt, R. A. & Björk, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts in training. *Psychological Science, 3*, 207–217.

Shebilske, W. L., Goettl, B., & Regian, W. J. (1999). Individual and group protocols for training complex skills in laboratory and applied settings. In D. Gopher & A. Koriat (Eds), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 401–432). Hillsdale, NJ: Erlbaum.

Shebilske, W. L., Regian, W. J., Arthur, W., Jr, & Jordan, J. A. (1992). A dyadic protocol for training complex skills. *Human Factors, 34*, 369–374.

Stajkovic, A. D. & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*, 240–261.

Swezey, R. W. & Salas, E. (1992). Guidelines for use in team training development. In R. W. Swezey & E. Salas (Eds), *Teams: Their training and performance* (pp. 219–245). Norwood, NJ: Ablex.

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two studies examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology, 87*, 506–516.

Vancouver, J. B., Thompson, C. M., & Williams, A. A. (2001). The changing signs in the relationships among self-efficacy, personal goals, and performance. *Journal of Applied Psychology, 86*, 605–620.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology* (R. W. Rieber & A. S. Carton, Eds; N. Minick, Trans.), New York: Plenum Press.

Whetzel, J. H. (2005). *Developing intelligent agents for training systems that learn their strategies from expert players*. Unpublished masters thesis, Texas A& M University.

Wooldridge, M. & Jennings, N. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review, 10*, 115–152.

THE INFLUENCE OF TRAINEE GAMING EXPERIENCE ON AFFECTIVE AND MOTIVATIONAL LEARNER OUTCOMES OF VIDEOGAME-BASED TRAINING ENVIRONMENTS

Karin A. Orvis
Old Dominion University

Kara L. Orvis
Aptima, Inc.

James Belanich
U.S. Army Research Institute for the Behavioral and Social Sciences

Laura N. Mullin
Nuance Communications, Inc.

Abstract

The current research investigated the influence of trainee videogame experience on affective and motivational learner outcomes of a videogame-based training environment. In this research, 413 participants played a first-person-perspective videogame which included a single-player section followed by a multi-player section. Results indicated that greater prior videogame experience was predictive of less difficulty using the game interface and greater perceived team cohesion, training satisfaction, and time spent engaging in the training game. Further, a videogame genre-specific effect was demonstrated in that only specific prior game experiences that share similar characteristics with the current training game were predictive of the learner outcomes. These findings have implications for training game developers.

Videogames are emerging as an increasingly popular training tool (Hays, 2005). Some of these videogames are initially built for entertainment purposes, but then are incorporated into training. Other games are developed specifically for a particular training purpose while including many features that appear in entertainment games (for a review of the

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

range of instances that can be defined as training games, see Hays, 2005). Compared to the more traditional lecture-based form of training, training games usually allow for greater interactivity, as well as providing realistic feedback and multi-sensory stimulation (Garris, Ahlers, & Driskell, 2002; Tarr, Morris, & Singer, 2002). These games are found to be intrinsically motivating and can facilitate learning (Gee, 2003; O'Neil & Fisher, 2004; Prensky, 2001). Transfer of the skills learned in the game to real-life situations has also been demonstrated (Gopher, Weil, & Bareket, 1994). However, the research on videogame-based training is not all positive, with a fair share of research showing that instructional games do not always lead to instructional gains (Hays, 2005). This ambiguity in findings may be due to a variety of design issues that influence learning, or difficulty in identifying appropriate learning outcomes (O'Neil, Wainess, & Baker, 2005).

Given the increasing popularity of using videogames for training, it is important to investigate factors that maximize the effectiveness of this training medium. Research suggests that training effectiveness is influenced by three primary determinants: (a) the training program itself (e.g., the actual videogame), (b) the trainee (in terms of prerequisite skills, personal characteristics, and attitudes), and (c) the situational context in which the training takes place (Campbell & Kuncel, 2001; Colquitt, LePine, & Noe, 2000; Mathieu, Tannenbaum, & Salas, 1992; Salas & Cannon-Bowers, 2001). Most available research on training games has primarily focused on the first determinant of effectiveness: videogame features or characteristics. Game features such as challenge, realism, and interactivity have been found to influence trainee motivation and the length of time in which trainees are willing to invest in mastering the skills taught during game play (Belanich, Sibley, & Orvis, 2004; Corbeil, 1999; Garris et al., 2002; Malone & Lepper, 1987). Such research has enhanced our understanding of how to design a training game to improve its effectiveness. Yet, to date, little emphasis has been paid to the second determinant of training effectiveness; how trainee characteristics influence learner outcomes in videogame-based training environments.

There is reason to believe that individual characteristics of the trainee may play an important role in predicting learner outcomes. Research on e-learning environments has found that trainees' characteristics significantly influence learner outcomes such as learning and motivation. Both stable characteristics, such as personality traits, and more malleable characteristics, such as prior computer experience, have been found to predict e-learning outcomes (Brown, 2001; Orvis, Fisher, & Wasserman, 2003; Schmidt & Ford, 2003). While e-learning and videogame-based training are distinct types of training, these two training mediums share several common characteristics, as they both represent self-directed, technology-delivered training environments. As such, trainee characteristics may also play a role in various learner outcomes in videogame-based learning environments. Accordingly, the purpose of the current research is to investigate the influence of the trainee characteristic of prior videogame experience. For the purpose of this research, prior videogame experience was operationalized as any game experience with computer-based videogames or console-based videogames (e.g., Playstation®, Xbox®). This malleable trainee characteristic was chosen because instructors have the ability to influence trainee experience levels prior to training implementation in an effort to ensure maximum training effectiveness. In contrast, stable characteristics, such as personality variables, are not easily altered.

1. Learner Outcomes of Game-Based Training Environments

The fundamental goal of training is to positively impact learners in terms of their knowledge, skills, and/or attitudes; therefore, identifying training variables that influence learner outcomes will lead to enhanced training effectiveness. The current research focuses on four affective and motivational learner outcomes that are significant in videogame-based training environments: time on task, training satisfaction, ease in using the training game interface, and perceived cohesion with one's teammates while playing the game.

One of the principal perceived benefits of using videogames for training purposes is that they are motivating for individuals to play (O'Neil & Fisher, 2005; Prensky, 2001). Motivation can be defined as the direction, strength, and persistence of volitional behavior (Campbell & Kuncel, 2001); and in the context of instruction, the volitional behavior can be assessed in terms of the amount of time spent purposefully engaging in or interacting with the instructional media. The length of time in which a learner is engaged in learning the knowledge or skills to be mastered during instruction has consistently been demonstrated to be an important predictor of learning in both educational and organizational learning contexts (e.g., Bloom, 1976; Borg, 1980; Brown, 2001). For example, Brown (2001) found that amount of time spent engaging in an e-learning program was positively related to subsequent knowledge acquisition. Thus, identifying factors that positively influence the length of time in which a learner spends actively engaging in the content of a training program (i.e., learner's time on task) is of great concern to instructors and is an important outcome to investigate in videogame-based learning environments.

Another outcome of interest to the current research is the learner's satisfaction with the training game experience. Training satisfaction includes both emotionally-based opinions concerning the training (e.g., the trainee liking the training) and reactions regarding the utility of the training (e.g., the trainee believing the training enhanced his/her knowledge or skills). Trainees' level of satisfaction with the training has been found to be significantly related to learning in an e-learning environment (Wasserman, Orvis, Fisher, & Barry, 2002). This may be because when trainees are more satisfied with their training experiences, they are likely to remain actively engaged and put forth greater effort toward learning the training content; thus, resulting in greater levels of learning.

The perception of ease in using the training game interface is another criterion that may influence the level of engagement in a videogame-based training environment. If e-learning environments are frustrating and difficult to use, trainees may not fully engage in the training and may experience decreased motivation (Park & Tennyson, 1980; Tennyson, 1980); subsequently, they may choose to withdraw from the training prematurely (Bell & Kozlowski, 2002; Steinberg, 1989). Indeed, difficulty with the technology or interface in which the instructional content is delivered has been cited as a key frustration source and a reason for low completion rates in e-learning programs (Frankola, 2001). Prior research has also found that trainees' perceptions regarding the user interface of an e-learning program were positively related to their satisfaction with the overall training; which, in turn, was positively related to learning (Wasserman et al., 2002). This suggests that the ease in using a game user interface is a valuable learner outcome in game-based learning environments.

Finally, many training videogames are collaborative in nature, requiring the interaction and cooperation among trainee team members in order to be successful in the game and to learn the training content. Previous collaborative learning research has found that the quality of intra-team interactions is a key element in determining the extent and depth of learning in such environments (Gilbert & Moore, 1998; Northrup, 2001; Shute, Lajoie, & Gluck, 2000; Wagner, 1997; Van den Bossche, Gijselaers, Segers, & Kirschner, 2006). Quality team interactions originate, in part, from collaborative team states such as team cohesion. Indeed, team cohesion, most commonly defined as members liking for one another (Evans & Jarvis, 1980) and the extent to which team members are attracted to the idea of the group (Hogg, 1992), has been found to be directly related to group effectiveness (Chang & Bordia, 2001; Evans & Dion, 1991; Mullen, Anthony, Salas, & Driskell, 1993; Mullen & Copper, 1994). Accordingly, this suggests that the level of cohesion trainees perceive with their teammates is an important learner outcome to investigate in collaborative game-based learning environments.

2. Prior Experience

There is a paucity of research on the influence of prior experience on learner outcomes of videogame-based training environments. However, research has examined the influence of prior experience in PC-based learning environments (e.g., Shih, Munoz, & Sanchez, 2006; Sitzman, Kraiger, Stewart, & Wisher, 2006). As aforementioned, e-learning and game-based training both represent self-directed, technology-delivered training environments; thus, research findings from e-learning and other PC-based learning environments may be relevant in making inferences concerning the role of prior experience in videogame-based learning environments. Accordingly, we reviewed the literature on prior computer experience and PC-based learning.

In general, research on the relationship between prior experience and various computer-based attitudinal and learning outcomes is equivocal. Some research has found that previous experience is positively related to computer-related performance and attitudes, such as interest and comfort with electronic learning (Brinkerhoff & Korolghlanian, 2005; Dias, 2000; Dyck & Smither, 1994; Houle, 1996; Huang, 2002; Ivers, Lee, & Carter-Wells, 2005; Rozell & Gardner, 1999; Shashaani, 1994; Shih et al. 2006), and negatively related to anxiety in using computers (Brown, Fuller, & Vician, 2002; Dyck & Smither, 1994; Keeler & Anson, 1995). Additional research has found previous experience to have a positive influence on PC-based learning, such that computer experience positively predicts motivation to learn, time on task in a distributed learning class (Patterson, 1999), learning in a PC-based training course (Dyck & Smither, 1996; Martocchio & Webster, 1992), and willingness to participate in future training courses (Brinkerhoff & Korolghlanian, 2005). In contrast, other research has found no clear relationship between previous experience and computer-based outcomes (Kay, 1992). For instance, some research either failed to find or found mixed results regarding the relationships between computer experience and anxiety in using computers (Henderson, Deane, Barrelle, & Mahar, 1995; Houle, 1996; Rozell & Gardner, 2000), computer-related attitudes (Henderson et al., 1995; Woodrow, 1991), and time spent in an e-learning course (Brown, 2001).

Several researchers have suggested that these mixed findings may be due, in part, to the different approaches taken to operationalize prior computer experience (Hasan, 2003; Smith, Caputi, Crittenden, Jayasuriya, & Rawstone, 1999). The majority of studies examine experience with computers globally as a unidimensional construct, and address level of computer experience as the frequency or length of time of general computer use (e.g., Al-Khaldi & Al-Jabri, 1998; Loyd & Loyd, 1985; McInerney, McInerney, & Sinclair, 1994; Salzer & Burks, 2003). Operationalizing experience as a global measure does not account for experiences with specific types of computer applications, such as word processing, spreadsheets, data analysis, computer games, e-mail, or programming. As such, the implicit assumption is that one type of computer experience is equivalent to all others, and any computer experience should positively influence subsequent training in a completely different type of computer task.

Such an assumption may not be appropriate. Some research demonstrates that *specific* computer experiences are differentially related to learning outcomes in PC-based learning environments (Brinkerhoff & Korolghlanian, 2005; Polman & Fishman, 1995; Salanova, Grau, Cifre, & Llorens, 2000; Woodrow, 1991). For example, Woodrow (1991) found that prior programming experience significantly predicted learning (measured as final course grade) in a PC-based training course, while prior word processing experience was not predictive of learning. Woodrow explained that the final course grade was heavily skewed toward programming ability rather than other applications such as word processing. Likewise, Reed, Oughton, Ayersman, Ervin, and Geissler (2000) found that computer experience directly related to navigation through a hypermedia environment produced more efficient navigation than unassociated computer experience.

Such findings indicate that previous experiences with tasks and technology similar to the given computer-based training environment are most predictive of attitudes and performance in such learning environments (Arthur, Bennet, Edens, & Bell, 2003). As such, a growing number of researchers have begun to adopt a multi-dimensional view of computer-related experience, arguing that reducing experience to a unidimensional construct has resulted in an oversimplication of the construct (Smith et al., 1999; Szajna & MacKay, 1995; Tracey, Hinkin, Tannenbaum, & Mathieu, 2001). This line of research advocates that measurements of computer experience assess diversity of computer experience, in terms of familiarity or frequency of use of various computer technology applications (e.g., Anderson & Reed, 1998; Busch, 1995; Chu, 2003; Chua, Chen, & Wong, 1999; Hasan, 2003; Polman & Fishman, 1995). It also suggests that the validity of computer experience as a predictor of training outcomes depends upon the relatedness of the particular training criterion of interest (e.g., particular learner outcomes) and type of computer experience examined.

2.1. Videogame Experience

Limited research has investigated the influence of prior experience on learner outcomes of videogame-based training environments. Research that has investigated prior experience in gaming environments tends to mirror the majority of work on previous computer experience, operationalizing the experience construct as general experience in using computers. Other research has examined prior experience with videogames in particular and found that prior overall or *general* videogame experience (regardless of the type of videogame previously played) was related to future performance in videogame-based environments (Alvarez, Salas, & Garofano, 2004; Gagnon, 1985; Young, Broach, & Farmer, 1997).

We propose that trainees' previous videogame experience will predict the four learner outcomes of interest: time of task, training satisfaction, ease in using the game interface, and perceived team cohesion (Hypothesis 1). Specifically, we suggest that trainees with greater levels of experience (e.g., those who play videogames on a more frequent basis for personal enjoyment purposes) should be more motivated to engage in such environments for training purposes as well, and therefore will spend more time engaging in the training. Baldwin and Magjuka (1997) suggest that the motivating influence of any training design element is partially contingent on the trainees' accumulated experience with that design element in other settings. Thus if trainees have had frequent, positive experiences with videogames in the past, they should be more likely to find such environments engaging in the future. Further, because trainees with greater levels of videogame experience enjoy this type of environment, they should be more satisfied with gaining new knowledge or skills from a training environment with similar characteristics, as compared to trainees who have not had much exposure to this type of environment in the past. In support of this proposition, previous research on computer-based training contexts has found that frequency of computer use (i.e., prior computer experience) is positively related to favorable attitudes towards computers (Mitra, 1998).

Previous experience with a game interface should also lead to the development of strategies and heuristics that smooth a trainee's navigation in a similar environment. It follows that prior experience should reduce the cognitive load of learning a new game, thereby increasing the utility of the game as a learning tool (Sweller, in press). Additionally, training programs in which the learner is primarily responsible for his/her own learning, as is often the case in e-learning and game-based learning environments, are more successful for trainees who have greater levels of prior knowledge on the relevant topics (De Rouin, Fritzsche, & Salas, 2004; Gay, 1986; Kirschner, Sweller, & Clark, 2006; Lee & Lee, 1991). We suggest that a trainee's prior knowledge of videogames (e.g., knowledge of videogame interfaces obtained from prior experiences playing videogames) will enable the trainee to feel more at ease when presented with a similar training environment.

Correspondingly, we suggest that trainees with little prior experience working with others and building positive relationships in collaborative virtual learning environments are likely less equipped with the knowledge of how to interact effectively in such environments. Thus, trainees with less prior collaborative videogame experience will have more difficulty forming cohesive relationships when engaging in videogame-based training environments that require virtual collaboration.

Similar to the acknowledgement of the multi-dimensional nature of computer experience, we suggest that previous videogame experience should be examined at a more specific level of analysis rather than only a unidimensional, global view of videogame experience. It is important to consider specific videogame experiences (i.e., an individual's experiences with particular types or genres of videogames) as well. Based on prior research on computer experience (e.g., Polman & Fishman, 1995; Shih et al. 2006; Woodrow, 1991), we propose that the impact of prior videogame experience on subsequent learner outcomes in a given game-based training environment will depend on the type of prior game experience one has acquired. Specifically, we hypothesize that only prior game experiences that share similar game characteristics to the given training game environment will influence learner outcomes; while experience with unrelated games (i.e., games that do not share similar characteristics) will not predict learner outcomes (Hypothesis 2).

Demonstrating that the relationships between videogame experience and various learner outcomes are dependent on the specific types of experience amassed is valuable in furthering our understanding of game-based learning contexts. However, there have been several recent calls in the literature to go beyond the examination of bivariate correlations (e.g., Avis, Kudisch, & Fortunato, 2002; Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001; Cortina, Goldstein, Payne, Davison, & Gilliland, 2000). Accordingly, we suggest that additional information is gained by demonstrating incremental validity of the specific game experiences above and beyond more general videogame experience. The primary issue is whether relevant specific game experiences (e.g., first-person-perspective videogames) account for variance in learner outcomes *beyond* that accounted for by a general measure of videogame experience. If an individual's report of his/her specific game experiences do not significantly contribute beyond the report of his/her overall game experience, there would be less utility in operationalizing videogame experience as a multi-dimensional construct. We hypothesize that prior videogame experience with specific games that share similar characteristics with the training game will provide incremental validity over general videogame experience in the prediction of learner outcomes (Hypothesis 3).

3. Method

3.1. Participants

Participants were 413 first-year U.S. Military Academy cadets who took part in a videogame-based tactics training exercise. The mean age of participants was 18.89 years

(SD = 1.26 years). Following the 4-day training exercise, cadets were asked to complete an online questionnaire. Approximately 1100 cadets participated in the training exercise, of which 413 cadets voluntarily completed the research questionnaire on their own time.

3.2. Game

The game used as the training exercise was *America's Army*® (United States Army, 2004), an online, first-person-perspective game with both single-player and multi-player sections. *America's Army*, created by the Office of Economic and Manpower Analysis at the U.S. Military Academy, was originally developed to serve as a recruiting tool in order to inform potential recruits about what to expect during basic training and about Army core values, history, and Army background. The distribution of *America's Army* has been extensive, with over eight million registered players (http://www.americasarmy.com, retrieved June 1, 2007). This game was chosen for the tactics training exercise because of its ability to simulate small team environments that require decision making and collaboration skills.

3.3. Procedure

First, the cadets completed a "basic training" single-player section, where they learned how to play the game. This section contains four segments: (a) marksmanship training, (b) an obstacle course, c) weapons familiarization, and (d) a MOUT (military operations in urban terrain) training mission.

Once the basic training section had been completed, cadets were eligible to play the multi-player section. For this section, cadets were placed into small teams and engaged in several collaborative missions. For each mission, a team's goal was either to attack or to defend a radio tower. Regardless of the team's goal within a given mission, cadets took the perspective of a U.S. Soldier, while the opposing team was depicted as the enemy.

The multi-player section of *America's Army* represents a distributed, online environment because all team members engage in the same mission during "real" time. However, each team member plays the game on an individual computer, in a different physical location. Team members interacted in terms of observing each other's actions within the game environment and via typed communication using a chat feature built into the game interface. Team membership varied across missions depending on which cadets were currently online engaging in the videogame. Cadets were required to play a minimum of three collaborative missions over the course of the 4-day training exercise, with no maximum limit of missions set.

Upon completion of the training exercise, cadets were provided with a brief description of the purpose of this research, as well as the website address of the research questionnaire.

They were also assured of confidentiality of their responses. Note that all 413 cadets who completed the research questionnaire were eligible for the multi-player section and completed at least three collaborative team missions.

3.4. Measures

General game experience was assessed using one item, "Based on the past year, how frequently have you played videogames (e.g., PC-based, Nintendo, Playstation, arcade)?" Possible responses ranged from 1 (none) to 5 (much more than average).

Specific game experience was assessed using a seven-item scale. Using a yes/no response (e.g., 0 or 1), participants were asked to note whether they had frequently played a specific type of videogame. The seven types of specific game experience assessed were (a) first-person-perspective (e.g., Battlefield 1942®, James Bond 007®); (b) simulation (e.g., Falcon®, Microsoft® Flight Simulator, Lock On™: Modern Air); (c) online multi-player games (e.g., EverQuest®, Planetside®); (d) action (e.g., Grand Theft Auto®, NBA®, car racing); (e) command/strategy (e.g., Risk®, chess); (f) creative development (e.g., Sims™, Tycoon®, Civilization®); and (g) puzzle (e.g., Minesweeper®). In addition, participants were asked to indicate the extent to which they had previously played *America's Army*. Possible responses for this item ranged from 1 (none) to 5 (much more than average).

Satisfaction with training was assessed using a six-item scale. Sample items include "I was satisfied with the experience of using the America's Army game" and "Using the America's Army game allowed me to better understand combat-related cognitive skills and decision-making." Possible responses ranged from 1 (strongly disagree) to 5 (strongly agree). The coefficient alpha for this scale was 0.86.

Ease in using the game user interface was assessed using a three-item scale. Sample items include "How easy/difficult was it to learn how to use America's Army game?" and "How easy/difficult was it to use the menu system?" Possible responses ranged from 1 (very difficult) to 5 (very easy). The coefficient alpha for this scale was 0.81.

Perceived team cohesion was assessed using a nine-item scale adapted from Craig and Kelly (1999) in order to fit the game environment. Sample items include "To what extent was your team engaged in the multi-player missions of the America's Army game?" and "To what extent did members of your team like being a part of this team?" Because team membership varied across missions played, trainees were asked to respond to these items with respect to the most successful team in which they were a team member. Possible responses ranged from 1 (not at all) to 5 (great extent). The coefficient alpha for this scale was 0.95.

Time on task was assessed using one item which asked participants to indicate the total number of hours spent playing the game during the 4 days allotted for this training exercise. We believe this reflects a trainee's motivation to continue training, as this videogame-based training represents a self-regulated, voluntary training environment. Possible responses ranged from 1 (1–5 h) to 4 (more than 15 h).

4. Results

4.1. Descriptive Statistics

Means, standard deviations, and intercorrelations of the relevant variables are displayed in Table 1. Results indicate that there was a wide range of prior videogame experience across participants, with 17% of cadets reporting they had no experience playing videogames, 44% reporting they had limited videogame experience, 22% reporting average videogame experience, and 18% reporting greater than average experience.

4.2. Role of General and Specific Videogame Experience

In support of Hypothesis 1, a general measure of prior videogame experience significantly predicted the learner outcomes. This indicates that the more experience individuals had playing videogames, the more at ease and satisfied they were using the training game ($r = 0.37$ and 0.33, respectively, $p < 0.01$), the more cohesive they felt with their teammates ($r = 0.33,\ p < 0.01$), and the more time they spent engaging in the training game ($r = 0.20,\ p < 0.01$).

Hypotheses 2 and 3 examined the role of prior videogame experience with *specific* games on learner outcomes. To test Hypothesis 2, four regression analyses were conducted, where each of the four learner outcomes (team cohesion, training satisfaction, ease in using the game interface, and time on task) were regressed onto the eight specific game type measures. Results indicate that generally only prior experiences related to the videogame used in the training were significant predictors. Specifically, previous experience with the *America's Army* game was a unique predictor for all four learner outcome

Table 1
Means, standard deviations, and intercorrelations of variables.

Variable	*N*	*M*	SD	1	2	3	4	5	6
1. General videogame experience	413	2.44	1.05	–					
2. America's Army experience	413	1.59	0.89	0.35**	–				
3. First-person-perspective experience	412	0.63	0.48	0.45**	0.24**	–			
4. Training satisfaction	365	3.40	0.72	0.33**	0.32**	0.25**	–		
5. Team cohesion	246	2.96	0.98	0.33**	0.32**	0.23**	0.45**	–	
6. Ease in using interface	359	3.63	0.76	0.37**	0.26**	0.36**	0.38**	0.42**	–
7. Time on task	339	1.55	0.78	0.20**	0.19**	0.10+	0.21**	0.22**	0.11*

Note. $^{+}p < 0.10$. $^{*}p < 0.05$. $^{**}p < 0.01$ (two-tailed).

variables. Prior experience with other first-person-perspective games was also a significant predictor of training satisfaction and ease in using the game interface; while it approached significance for team cohesion. As expected, experience using other types of specific games which did not share several similar characteristics to the current training game, such as puzzles and creative development games, were not predictive of these four outcomes. An unexpected result was that prior simulation experience predicted time on task. We revisit this unexpected finding in the discussion section. Results are presented in Table 2.

To test Hypothesis 3, whether prior specific videogame experience provides incremental validity over general videogame experience in the prediction of the four learner outcomes, a separate hierarchical regression analysis was performed for each learner outcome. For each regression analysis, general videogame experience was entered first, followed by a block of variables in Step 2 consisting of the genre specific videogame experiences of prior *America's Army* experience and first-person-perspective videogame experience. Results indicate that prior *America's Army* experience and first-person-perspective experience predicted unique variance above and beyond general videogame experience for two of the four dependent variables, training satisfaction and ease in using the game interface. For the other two dependent variables, prior *America's Army* experience significantly contributed above and beyond general videogame experience for team cohesion and approached significance ($p = 0.06$) for time on task. In short, genre-specific game experience predicted the learner outcomes of interest beyond what was explained by globally reported videogame experience alone. As such, there was value to measuring trainees' prior specific game experiences as well as their general videogame experience. Results are presented in Table 3.

5. Discussion

The results of this research suggest that trainees' prior videogame experience has implications for several learner outcomes in a videogame-based training environment. Specifically, more experienced trainees reported greater ease using the training game interface, higher levels of training satisfaction, and a greater length of time spent engaging in the training game. This supports past research indicating that the exposure to computers and videogames is meaningful to a learner's experience in a game-based learning environment (Alvarez et al., 2004; Gagnon, 1985; Greenfield, deWinstanley, Kilpatrick, & Kaye, 1996; Sims & Mayer, 2002; Subrahmanyam & Greenfield, 1996; Young et al., 1997). The results also demonstrate that prior videogame experience has implications for videogame-based training environments which require trainee collaboration. In the current research, trainees with higher levels of experience more easily formed cohesive relationships with team members in the collaborative components of the training game, as compared to trainees who had less prior experience.

Further, the results suggest that the specificity of a trainee's prior experience predicts learner outcomes. Trainees with greater experience in playing games related to the current training game environment (i.e., *America's Army* and other first-person-perspective

Table 2
Regression analyses for specific videogame experiences predicting learner outcomes.

	Cohesion			Satisfaction			Ease using interface			Time on task		
Variable	*B*	SE *B*	β	*B*	SE *B*	β	*B*	SE *B*	β	*B*	SE *B*	β
America's Army	0.29	0.07	0.27**	0.21	0.04	0.26**	0.12	0.04	0.14**	0.11	0.05	0.12*
First-person-perspective	0.25	0.14	0.12+	0.24	0.09	0.16**	0.38	0.09	0.24**	−0.01	0.10	−0.01
Simulation	−0.03	0.18	−0.01	0.02	0.11	0.01	0.10	0.11	0.04	0.35	0.13	0.16**
Online multi-player	−0.02	0.18	−0.01	−0.01	0.11	−0.01	0.20	0.11	0.09+	0.20	0.13	0.09
Action	0.18	0.13	0.09	0.08	0.08	0.05	0.07	0.08	0.04	−0.03	0.09	−0.02
Command/strategy	0.09	0.14	0.05	0.13	0.08	0.09	0.13	0.09	0.09	0.04	0.10	0.02
Creative development	0.04	0.15	0.02	−0.12	0.09	−0.07	0.10	0.10	0.06	0.17	0.10	0.09
Puzzles	−0.02	0.12	−0.01	0.09	0.07	0.06	−0.06	0.08	−0.04	−0.06	0.09	−0.04

Note. $R^2 = 0.14^{**}$, 0.14^{**}, 0.20^{**}, and 0.09^{**}, for team cohesion, training satisfaction, ease in using interface, and time on task, respectively. $N = 244$, 362, 356, and 336, for each criterion, respectively. $^{+}p < 0.10$. $^{*}p < 0.05$. $^{**}p < 0.01$.

Table 3
Incremental validity results of specific videogame experiences over general videogame experience.

Regression step	Cohesion			Satisfaction			Ease using interface			Time on task		
	B	SE *B*	*β*	*B*	SE *B*	*β*	*B*	SE *B*	*β*	*B*	SE *B*	*β*
Step 1												
General game experience	0.30	0.06	0.33**	0.22	0.03	0.33**	0.27	0.04	0.37**	0.15	0.04	0.20**
Step 2												
General game experience	0.20	0.06	0.21**	0.14	0.04	0.20**	0.16	0.04	0.23**	0.11	0.05	0.15*
America's Army	0.24	0.07	0.23**	0.17	0.04	0.21**	0.10	0.04	0.11*	0.12	0.05	0.13*
First-person-perspective	0.19	0.13	0.09	0.16	0.08	0.11*	0.38	0.08	0.24**	0.01	0.10	0.01

Note. $R^2 = 0.11^{**}$ for Step 1; $\Delta R^2 = 0.06^{**}$ for Step 2, for team cohesion. $R^2 = 0.11^{**}$ for Step 1; $\Delta R^2 = 0.05^{**}$ for Step 2, for training satisfaction. $R^2 = 0.14^{**}$ for Step 1; $\Delta R^2 = 0.06^{**}$ for Step 2, for ease in using interface. $R^2 = 0.04^{**}$ for Step 1; $\Delta R^2 = 0.02^{+}$ for Step 2, for time on task. $N = 244$, 363, 357, and 337, for team cohesion, training satisfaction, ease in using interface, and time on task, respectively. $^{+}p < 0.10$. $^{*}p < 0.05$. $^{**}p < 0.01$.

games) reported greater ease in using the game interface, higher levels of training satisfaction, and greater team cohesion. Prior experience with *America's Army* also predicted time on task. In contrast, prior experiences with specific games that do not share similar characteristics with the current game-based training environment were generally not related to learner outcomes.

At first glance, one might have expected that online multi-player games would share several similarities with the multi-player component of *America's Army*; and thus should have significantly predicted learner outcomes, particularly team cohesion. While other multi-player videogames developed and popularized in the private sector (e.g., *Ultima Online*®, *Everquest*) do simultaneously engage multiple individuals in game play, such games represent a very different game playing experience than *America's Army*. First, the number of players interacting within the game at any one time is vastly different. For this research, participants playing *America's Army* were formed into teams composed of a maximum of 16 players. Then, each team participated in a scenario by interacting with one opposing team of a similar size. Although many scenarios could be run at once, from the player's point of view only a relatively small number of individuals were participating in game play at any one time. In contrast, games such as *Everquest* are termed massively multi-player online games because thousands of individuals are simultaneously involved in game play and the number of teams that can exist is not limited.

Another distinction was that in the current training exercise with *America's Army*, individuals were randomly assigned to a team for a given scenario and the scenarios in

America's Army tend to be relatively short, about 10 min long. Further, membership in the teams could vary across scenarios such that when a new scenario starts there may be some changes in particular players participating. In contrast, in most massively multi-player games, team membership is emergent and tends to be persistent, in that the game continues without a clear endpoint. Therefore, the groups and relationships develop over longer periods of time. We expect that individual's prior experience collaborating with specific players based on one's own selection, over longer periods of time, is likely to be a different experience than that of collaboration with other individuals assigned as team members for a relatively short time period.

One unexpected finding was that prior simulation game experience (e.g., with *Falcon, Lock On: Modern Air*) was positively related to time on task for the *America's Army* game. A possible explanation is that simulation games are more closely associated to training well-defined skills (versus solely providing entertainment) as compared to other types of videogames. Another explanation may be that *America's Army* and flight simulations share some common game features or characteristics (e.g., first-person-perspective in a virtual 3D world and pacing of game actions/events) that are critical to game success.

Finally, the results indicated that specific prior game experiences that share similar characteristics with the intended training environment provide incremental validity over general videogame experience in the prediction of learner outcomes. In other words, knowledge of both trainees' general and specific videogame experiences was valuable in predicting their future experiences with and reactions to a given game-based training environment.

5.1. Implications and Conclusions

On the basis of the current research findings, we suggest some practical implications for instructors and training games developers, as well as some future research directions. First, for instructors utilizing training games, we provide support for the value of assessing trainees' specific types of game experience. It seems intuitive that if one has prior general videogame experience, he/she should have more positive experiences in any videogame-based training environment. However, based on the current research, this is not the case. Experience with one type of gaming environment did not necessarily enhance the training outcomes in a different type of training game environment.

The good news is that learner experience is a malleable trainee characteristic that can be compensated for fairly easily. By assessing the amount and types of previous gaming experiences trainees possess, instructors will be able to identify those who lack the prerequisite game experience. In turn, instructors can then provide these trainees with targeted opportunities to gain such beneficial experiences prior to training. For example, if learners are to engage in a first-person-perspective game-based training program and some learners have little prior experience with this type of game, then the instructor would know to give them ample practice time before the learning segment of the training (i.e., when learners are acquiring the new skills or knowledge taught in the game). To facilitate instructors in providing the appropriate amount of preparatory practice for a given learner's needs, training game developers should incorporate a feature within the game that enables the instructor to select the desired amount and content of trainee orientation and practice.

Further, it may be assumed that most individuals who grew up in the digital age would have a great deal of experience with videogames. However, this assumption does not seem warranted given the experience levels of the students sampled, U.S. Military Academy cadets. In the current sample, 17% of students reported they had no experience playing videogames and 44% reported they had limited videogame experience. Given the selection process required to attend the U.S. Military Academy, this sample may not be indicative of typical college-age students or the general public. Nevertheless, the number of students with little to no experience was substantial; therefore, providing an orientation or additional practice with relevant games would likely be valuable. Doing so may improve a host of learner outcomes including training satisfaction, perceived team cohesion, ease in using the game interface, and the length of time spent engaging in the training.

We suggest that future research investigate the role of trainee *preference* for videogames on learner outcomes. In the real world, videogame preference and prior experience should be positively correlated, such that the more one enjoys a type of videogame, the more frequently he/she plays this kind of game. An important question is whether it is one's prior experience with videogames that influences subsequent learner outcomes or whether these relationships are a function of the individual's preference for playing these games. It is possible that actual enjoyment or preference for game playing may be driving the relationships between game experience and learner outcomes of the training.

Acknowledgments

We thank MAJ Carl Jacquet (Director) and Vincent Taijeron of the War Fighting Simulation Center at the U.S. Military Academy, for their aid in coordinating data collection. We also appreciate the technical support provided by Sharon Meyers, U.S. Army Research Institute (ARI). Portions of this chapter were presented at the E-Learn World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education Conference, Washington, DC, November 2004, and the American Psychological Association Division 21/19 Annual Symposium on Applied Experimental Research, Fairfax, VA, March 2005. The views, opinions, and/or findings contained in this chapter are solely those of the authors and should not be construed as an official Department of the Army or DOD position, policy, or decision, unless so designated by other documentation.

References

Al-Khaldi, M. A. & Al-Jabri, I. M. (1998). The relationship of attitudes to computer utilization: new evidence from a developing nation. *Computers in Human Behavior, 14*(1), 23–42.

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*, 385–416.

Anderson, D. K. & Reed, W. M. (1998). The effects of internet instruction, prior computer experience, and learning style on teachers' internet attitudes and knowledge. *Journal of Educational Computing Research, 19*(3), 227–246.

Arthur, W., Bennet, W., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*, 234–245.

Avis, J. M., Kudisch, J. D., & Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology, 17*(1), 87–105.

Baldwin, T. T. & Magjuka, R. J. (1997). Training as an organizational episode: pretraining influences on trainee motivation. In J. K. Ford & Associates (Eds), *Improving training effectiveness in work organizations* (pp. 99–127). Mahwah, NJ: Lawrence Erlbaum.

Belanich, J., Sibley, D., & Orvis, K. L. (2004). *Instructional characteristics and motivational features of a PC-based game* (Research Report, No. 1822). US Army Research Institute for the Behavioral and Social Sciences: Arlington, VA.

Bell, B. S. & Kozlowski, S. W. (2002). Adaptive guidance: enhancing self-regulation, knowledge and performance in technology-based training. *Personnel Psychology, 55*, 267–306.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Borg, W. R. (1980). Time and school learning. In C. Denham & A. Lieberman (Eds), *Time to learn* (pp. 33–72). Washington, DC: National Institute of Education.

Brinkerhoff, J. & Koroghlanian, C. M. (2005). Student computer skills and attitudes toward internet-delivered instruction: an assessment of stability over time and place. *Journal of Educational Computing Research, 32*, 27–56.

Brown, K. G. (2001). Using computers to deliver training: which employees learn and why? *Personnel Psychology, 54*, 271–296.

Brown, S. A., Fuller, R. M., & Vician, C. (2002). *Individual characteristics and e-learning: The role of computer anxiety and communication apprehension* (Technical Report 122-1). Indiana University Information Systems Department: Bloomington, Indiana.

Busch, T. (1995). Gender differences in self-efficacy and attitudes towards computers. *Journal of Educational Computing Research, 12*(2), 147–158.

Campbell, J. P. & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds), *Handbook of industrial, work and organizational psychology*. Thousand Oaks, CA: Sage Publications.

Change A. & Bordia P. (2001). A multidimensional approach to the group cohesion-group performance relationship. *Small Group Research, 32*(4), 379–405.

Chu, L. (2003). The effects of web page design instruction on computer self-efficacy of preservice teachers and correlates. *Journal of Educational Computing Research, 28*, 127–142.

Chua, S. L., Chen, D., & Wong, A. F. L. (1999). Computer anxiety and its correlates: A meta-analysis. *Computers in Human Behavior, 15*, 609–623.

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Harvey, V. S. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86*(3), 410–417.

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678–707.

Corbeil, P. (1999). Learning from the children: Practical and theoretical reflections on playing and learning. *Simulation & Gaming, 30*(2), 163–180.

Cortina, J. M., Goldstein, N. B., Payne, S. C., Davison, H. K., & Gilliland, S. W. (2000). The incremental validity of interview scores over and above cognitive ability and conscientiousness scores. *Personnel Psychology, 53*, 325–351.

Craig, T. Y. & Kelly, J. R. (1999). Group cohesiveness and creative performance. *Group Dynamics: Theory, Research, & Practice, 3*(4), 243–256.

De Rouin, R. E., Fritzsche, B. A., & Salas, E. (2004). Optimizing e-learning: Research-based guidelines for learner controlled training. *Human Resources Management, 43*, 147–162.

Dias, J. (2000). Learner autonomy in Japan: Transforming "help yourself" from threat to invitation. *Computer Assisted Language Learning, 13*(1), 49–64.

Dyck, J. L. & Smither, J. A. (1994). Age differences in computer anxiety: The role of computer experience, gender and education. *Journal of Educational Computing Research, 10*(3), 239–248.

Dyck, J. L. & Smither, J. A. (1996). Older adults' acquisition of word processing: The contribution of cognitive abilities and computer anxiety. *Computers in Human Behavior, 12*(1), 107–119.

Evans, C. R. & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. *Small Group Research, 22*, 175–186.

Evans, C. R. & Jarvis, P. A. (1980). Group cohesion: A review and re-evaluation. *Small Group Behavior, 11*, 359–370.

Frankola, K. (2001, October). Why online learners drop out. *Workforce, 80*(10), 52–60.

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *Educational Communication and Technology Journal, 33*, 263–275.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted video instruction. *Journal of Education Psychology, 78*, 225–227.

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

Gilbert, L. & Moore, D. R. (1998). Building interactivity into web courses: Tools for social and instructional interaction. *Educational Technology, 38*(3), 29–35.

Gopher, E., Weil, M., & Bareket, T. (1994). Transfer of a skill from a computer game training to flight. *Human Factors, 36*(3), 387–405.

Greenfield, P. M., deWinstanley, P., Kilpatrick, H., & Kaye, D. (1996). Action video games and informal education: Effects on strategies for dividing visual attention. In P. M. Greenfield & R. R. Cocking (Eds), *Interacting with video. Advances in applied developmental psychology* (Vol. 11, pp. 187–205). Westport, CT: Ablex Publishing.

Hasan, B. (2003). The influence of specific computer experiences on computer self-efficacy beliefs. *Computers in Human Behavior, 19*, 443–450.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion* (Technical Report 2005-004). Naval Air Warfare Center Training Systems Division: Orlando, FL.

Henderson, R., Deane, F., Barrelle, K., & Mahar, D. (1995). Computer anxiety: Correlates, norms and problem definition in health care and banking employees using the Computer Attitude Scale. *Interacting with Computers, 7*(2), 181–193.

Hogg, M. A. (1992). *The social psychology of group cohesiveness: For attraction to social identify*. London: Harvester Wheatsheaf.

Houle, P. A. (1996). Toward understanding student differences in a computer skills course. *Journal of Educational Computing Research, 14*(1), 25–48.

Huang, H. M. (2002). Student perceptions in an online mediated environment. *International Journal of Instructional Media, 29*, 405–422.

Ivers, K., Lee, J., & Carter-Wells, J. (2005). *Students' attitudes and perceptions of online instruction*. Paper presented at the National Education Computing Conference, Philadelphia, PA.

Kay, R. H. (1992). Charting pathways of conceptual change in the use of computer software: A formative analysis. *Journal of Research on Technology in Education, 26*(3), 403–417.

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*, 65–86.

Keeler, C. M. & Anson, R. (1995). An assessment of cooperative learning used for basic computer skills instruction in the college classroom. *Journal of Educational Computing Research, 12*(4), 379–393.

Lee, S. & Lee, Y. H. K. (1991). Effects of learner-control versus program-control strategies on computer-aided learning of chemistry problems: For acquisition or review? *Journal of Educational Psychology, 83*, 491–498.

Loyd, B. H. & Loyd, D. E. (1985). The reliability and validity of an instrument for the assessment of computer attitudes. *Educational and Psychological Measurement, 45*, 903–908.

Malone, T. W. & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow and M. J. Farr (Eds), *Aptitude, learning and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Martocchio, J. J. & Webster, J. (1992). Effects of feedback and cognitive playfulness on performance in microcomputer software training. *Personnel Psychology, 45*, 553–578.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 35*(4), 828–847.

McInerney, V., McInerney, D. M., & Sinclair, K. E. (1994). Student teachers, computer anxiety and computer experience. *Journal of Educational Computing Research, 11*(1), 27–50.

Mitra, A. (1998). Categories of computer use and their relationships with attitudes towards computers. *Journal of Research on Computing Education, 30*, 281–296.

Mullen, B., Anthony, T., Salas, E, & Driskell, J. E. (1994). Group cohesiveness and quality of decision making: An integration of tests of the groupthink hypothesis. *Small Group Research, 25*(2), 189–204.

Mullen, B. & Cooper, C. (1994). The relation between group cohesiveness and performance: An integration. *Psychological Bulletin, 115*(2), 210–227.

Northrup, P. (2001). A framework for designing interactivity into web-based instruction. *Educational Technology, 41*(2), 31–39.

O'Neil, H. F. & Fisher, Y. C. (2004). A technology to support leader development: Computer games. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leader development for transforming organizations* (pp. 99–121). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Wainess, R, & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474.

Orvis, K. A., Fisher, S. L., & Wasserman, M. E. (2003). *Am I e-nabled for e-learning? Individual differences and e-learning reactions*. Poster presented at the 18th annual conference for the Society for Industrial Organizational Psychology, Orlando, FL.

Park, O. & Tennyson, R. D. (1980). Adaptive design strategies for selecting number and presentation of examples in coordinate concept acquisition. *Journal of Educational Psychology, 72*, 362–370.

Patterson, N. J. H. (1999). *An evaluation of graduate class interaction in face-to-face and asynchronous computer groupware experiences: A collective case study*. Presented at the Association for the Study of Higher Education 23rd Annual Meeting.

Polman J. & Fishman, B. (1995). *Electronic communication tools in the classroom: Student and environmental characteristics as predictors of adoption*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

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

Reed, W. M., Oughton, J. M., Ayersman, D. J., Ervin, J. R., & Giessler, S. F. (2000). Computer experience, learning style, and hypermedia navigation. *Computers in Human Behavior, 16*, 609–628.

Rozell, E. J. & Gardner, W. L. (1999). Computer-related success and failure: A longitudinal field study of the factors influencing computer-related performance. *Computers in Human Behavior, 15*, 1–10.

Rozell, E. J. & Gardner, W. L. (2000). Cognitive, motivation, and affective processes associated with computer-related performance: A path analysis. *Computers in Human Behavior, 16*, 199–222.

Salanova, M., Grau, R. M., Cifre, E., & Llorens, S. (2000). Computer training, frequency of usage and burnout: The moderating role of computer self-efficacy. *Computers in Human Behavior, 16*, 575–590.

Salas, E. & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471–499.

Salzer, M. S. & Burks, V. (2003). A mediational study of computer attitudes, experience, and training interests among people with severe mental illnesses. *Computers in Human Behavior, 19*, 511–521.

Schmidt, A. M. & Ford, J. K. (2003). Learning within a learner control training environment: The interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology, 56*, 405–429.

Shashaani, L. (1994). Gender-differences in computer experience and its influence on computer attitudes. *Journal of Educational Computing Research, 11*(4), 347–367.

Shih, P. C., Munoz, D., & Sanchez, F (2006). The effect of previous experience with information and communication technologies on performance in a web-based learning program. *Computers in Human Behavior, 22*, 962–970.

Shute, V. J., Lajoie, S. P., & Gluck, K. (2000). Individual and group approaches to training. In S. Tobias & H. O'Neill (Eds), *Handbook on training* (pp. 171–207). Washington, DC: American Psychological Association.

Sims, V. K. & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*, 97–115.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*, 623–664.

Smith, B., Caputi, P., Crittenden, N., Jayasuriya, R., & Rawstone, P. (1999). A review of the construct of computer experience. *Computers in Human Behavior, 15*, 227–242.

Steinberg, E. R. (1989). Cognition and learner control: A literature review, 1977–1988. *Journal of Computer-Based Instruction, 16*(4), 117–121.

Subrahmanyam, K. & Greenfield, P. M. (1996). Effect of video game practice on spatial skills in girls and boys. In P. M. Greenfield & R. R. Cocking (Eds), *Interacting with video. Advances in applied developmental psychology* (Vol. 11, pp. 95–114). Westport, CT: Ablex Publishing.

Sweller, J. (in press). Human cognitive architecture and educational technology. In J. M. Spector, M. D. Merrill, J. J. G. Van Merriënboer, & M. P. Driscoll (Eds), *Handbook of Research on Educational Communications and Technology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Szajna, B. & MacKay, J. M. (1995). Predictors of computer-user training environment: A path analytic study. *International Journal of Human-Computer Interaction, 7*(2), 167–185.

Tarr, R. W., Morris, C. S., & Singer, M. J. (2002). *Low-cost PC gaming and simulation research: Doctrinal survey* (Research Note 2003-03). US Army Research Institute for the Behavioral and Social Sciences: Arlington, VA.

Tennyson, R. D. (1980). Instructional control strategies and content structure as design variables in concept acquisition using computer-based instruction. *Journal of Educational Psychology, 72*, 525–532.

Tracey, J. B., Hinkin, T. R., Tannenbaum, S., & Mathieu, J. E. (2001). The influence of individual characteristics and the work environment on varying levels of training outcomes. *Human Resource Development Quarterly, 12*, 5–23.

United States Army (2004). America's Army: Special Forces (Vanguard) (v2.2.0) [Computer software]. Retrieved from http://www.americasarmy.com.

Van den Bossche, Piet, Gijselaers W. H., Segers M., & Kirschner, P. A. (2006). Social and cognitive factors driving teamwork in collaborative learning environments. *Small Group Research, 37*(5), 490–521.

Wagner, R. K. (1997). Intelligence, training, and employment. *American Psychologist, 52*, 1059–1069.

Wasserman, M. E., Orvis, K. A., Fisher, S. L., & Barry, T. (2002). Trainee reactions to learner control features: An important link in the e-learning equation. In K. G. Brown (Chair), K. Kraiger (Discussant), & J. K. Ford (Discussant), *E-Learning participation and outcomes: Evidence from the field*. Symposium conducted at the 17th Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Woodrow, J. J. (1991). A comparison of four computer attitude scales. *Journal of Educational Computing Research, 7*(2), 165–187.

Young, W. C., Broach, D., & Farmer, W. L. (February, 1997). *The effects of video game experience on computer-based air traffic controller specialist, air traffic scenario test scores* (FAA-AM-97-04). FAA Office of Aviation Medicine Reports: Washington, DC.

This page intentionally left blank

UTILIZING MULTIPLAYER GAMES FOR TEAM TRAINING: SOME GUIDELINES

Michael T. Curtis, Michelle Harper-Sciarini, Deborah DiazGranados, Eduardo Salas, and Florian Jentsch

University of Central Florida

Abstract

This chapter examines the utility of multiplayer video games for the development of team skills. The focus of the chapter is to identify team skills that can be most saliently targeted in a multiplayer game. Characteristics of both commercial off-the-shelf and serious multiplayer games are then discussed in terms of how they can elicit these skills. From this we develop a set of guidelines that can be used to assess the strengths of a game in terms of team skills development. Finally, we suggest some things to consider for the future direction of multiplayer video games as training tools.

Each year, teams from all over the world gather to participate in the *Counter-Strike* World Championships (The 4th edition, 2006). *Counter-Strike* (Valve, 2004), a team-based first person shooter (FPS) video game with a modern warfare theme, has quickly become one of the faces of a new generation of video games. These video games allow players to go beyond the single-player limitations characteristic of the previous generation of games. Instead of single players only gaming with or against computer-generated characters, these new video games allow multiple players to work together and in competition with other teams. At the *Counter-Strike* World Championships, for example, each team participated with five players who, in order to succeed, were required to utilize effective tactical planning, efficient communication, and quick reaction on a dynamic playing field.

Video games are defined as games that require a screen for viewing and input devices, such as controllers, to interact in rule-governed, goal-focused, microcomputer-driven activities incorporating principles of gaming and computer-assisted instruction (Driskell & Dwyer, 1984; O'Neil & Fisher, 2004). There is no doubt that the teams competing in

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

the *Counter-Strike* World Championships had invested countless hours playing collaboratively and honing their skills in preparation for the games. Although perhaps unintentional, the activity of playing the game together had become a way for the players to build camaraderie, shared expectations, and procedures for cooperation and collaboration, all of which are teamwork skills that are essential for effective team performance.

For the reason that gaming formats, such as C*ounter-Strike* (Valve, 2004), involve interactions that make up teamwork behaviors, it can be assumed that video games have the tools for supporting the development of teamwork skills. In this chapter the purpose is twofold, the capability of utilizing video games for training teamwork skills will be identified along with theoretically driven guidelines that should be followed when selecting a video game that targets the development of teamwork skills. In addition, this chapter presents a platform for which empirical research on the use of video games for team training can be designed and executed.

The chapter is organized as follows. We begin by reviewing the nature of teams. We define teams, teamwork, and identify three critical constructs to the success of teams and the development of teamwork. We then discuss games and their utility in training. This discussion focuses on the use of commercial off-the-shelf video games as well as serious games (i.e., video games developed with a specific educational or training purpose). To accomplish our objective of identifying guidelines when selecting a video game for the purposes of training teams, we identify key components of multiplayer games in addition to game generic components. Finally, we close the discussion by suggesting future research efforts that will fill the gap found in the current literature on the use of video games for team training.

1. Teams and Teamwork

Teams and team behaviors have been studied extensively and as a result there is a cornucopia of works that provide theory on team functioning (Kozlowski & Bell, 2003; Levine & Moreland, 1990; Sundstrom, de Meuse, & Futrell, 1990; Sundstrom, McIntyre, Halfhill, & Richards, 2000). These works offer up many explanations on how teams operate based on processes, compositions, or functions specific to the team. Our intent is not to review these works or the theories presented therein but rather we identify three constructs found in teams that we designate as critical for successful team performance. These constructs or skills are described in terms of their impact on teamwork and are, in a later section of this chapter, discussed in relation to targeted features of video games that can be used to train teams.

In order to explain how multiplayer video games are relevant to the development of teamwork skill we have chosen to focus on the components of teamwork that appear to have the most salient relationship from video game to team performance. The three teamwork constructs discussed below are (1) communication, (2) coordination, and (3) team leadership. Each of these constructs is delineated after brief definition of team, teamwork, and team training is provided.

1.1. Definition of Teams and Teamwork

Teams are very complex, adaptive, and dynamic systems made up of two or more interdependent individuals whose goal is to meet a common and valued objective (Hollenbeck et al., 1995; McGrath, Arrow, & Berdahl, 2000; Salas, Dickinson, Converse, & Tannenbaum, 1992). What occurs within teams is teamwork. Teamwork is defined by Salas, Sims, and Burke (2005) as a set of interrelated thoughts, actions, and feelings that combine to facilitate coordinated and adaptive behaviors with the ultimate goal of completing taskwork objectives. Once team members have become proficient in teamwork skills a team's level of performance is greater than the combined efforts of its individual members. Teams are dynamic entities, thus it can be said that teams think, do, and feel. It is the synergy between these actions that allow teams to effectively complete their goals (Salas, Burke, & Stagl, 2004).

Nevertheless, simply putting a team together does not guarantee that these skills will develop. Team training is necessary. Team training is made up of a combination of various elements that target team processes and outcomes through practice and feedback in a dynamic environment (Cannon-Bowers & Salas, 1997; Salas et al., 2004). Video games are capable of fitting into this ensemble of elements that make up effective team training. As mentioned previously, we will address three team constructs that can be enhanced by video gameplay. These constructs will namely be communication, coordination, and team leadership. These constructs or team processes all influence team behaviors and thus team performance. As previously stated, these constructs were chosen because of the influence that video gameplay can have on the training of these team skills.

To elaborate, video games offer specific characteristics or features, such as communication modes, that will mimic the behaviors that make up the teamwork process. For example, some video games offer communication methods similar to those that are used in distributed team environments; therefore, video games that possess these communication features can provide trainee's with a medium for practicing their communication skills in a simulated environment. Coordination between team members can be enhanced by video game features that require teams to pre-plan their mission or to sequence those tasks that each team member is assigned. Also, team leadership can be enhanced by features that require team members to delegate tasks. These are only a few examples of how video gameplay can enhance team performance. Each of these three constructs identified are important teamwork processes and can be trainable through video gameplay. These three constructs chosen are the most saliently observable team process skills in video gameplay.

1.2. Teamwork Processes

Communication. Communication is seen as a means for enabling team processes such as coordination and cooperation. Research has shown that teams with exceptional communication skills perform better than teams with deficient communication skills (Seigel & Federman, 1973). Communication within a team can serve two very important functions that help team performance (Glickman et al., 1987); improvement of taskwork and in teamwork.

Effective taskwork communication behaviors include an exhibition of a clear and an accurate exchange of information and a development of team solutions to problems. Salas and colleagues (2005) proposed the use of closed-loop communication in order to minimize information exchange difficulties. Closed-loop communication involves three steps (1) the sender introduces the message, (2) the receiver then receives, interprets, and acknowledges receipt of the message, and (3) the sender then confirms that the intended message was actually received (McIntyre & Salas, 1995).

Teamwork communication focuses on establishing patterns of behaviors and enhancing their quality (Kozlowski & Bell, 2003). Research conducted on communication patterns between team members has found that team performance varies depending on these patterns (Ancona & Caldwell, 1992). What effective communication among team members does is, it affords shared understanding of the task at hand and ensures that all the members possess the information required to successfully accomplish a task. By considering communication as the means to develop other teamwork behaviors, as discussed later in this chapter (i.e., coordination), it is an invaluable skill that teams should foster in order to improve their effectiveness and performance.

Coordination. Coordination is the process by which team resources, activities, and responses are organized to ensure that tasks are integrated, synchronized, and completed within established temporal constraints (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995). The idea of integrating markedly different actions together in conjunction with temporal pacing is central to the conceptualization of coordination (Argote & McGrath, 1993). Task interaction and task organization are both examples of behaviors that make up coordination (Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986). Task interaction is best characterized as the method that team members use to interact and accomplish tasks. An example of task interaction behavior is when teammates obtain information about their tasks from other team members when necessary for the completion of tasks.

Other actions such as back-up behaviors and situational awareness add to the development of a well-coordinated team. Backing-up behaviors is defined as the helping actions provided to team members in order to perform their role (McIntyre & Salas, 1995; Morgan et al, 1986). In order for team members to be able to effectively engage in these behaviors they must have an understanding of their teammates' jobs but they must also be willing and able to provide and seek assistance when needed (Porter et al., 2003).

Situational awareness is the perception of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1995). Researchers have demonstrated that well-coordinated teams perform better than their less coordinated counterparts. In a study by Stout, Salas, and Carson (1994) it was found that coordination ratings positively predicted mission performance on a flight simulation task.

Leadership. The role of leader can vary from being an assigned position to one individual, an assigned position based on natural selection, or a role that is considered to be shared in which several team members occupy the role of leader (i.e., shared leadership). The traditional view of leadership is that there is a person that directs action, provides motivation, and has a final say in making decisions. A leader may face many challenges such as building a new team, socializing team members, or even developing team members. Development of team members may focus on establishing team orientation and team coherence.

Recent works have suggested that leadership in teams is not limited to one person who is designated as team leader, but instead that team leadership is dynamic or distributed within the team (Day, Gronn, & Salas, 2004). For shared leadership to be effective, teams must de-emphasize power differentials among members and work to balance the various types of power across members (Nichols, DeFriese, & Malone, 2002). Leadership within a self-managing team is dynamic in the sense that the leadership at any given time will depend on the expertise needed for a task. In order to accomplish tasks effectively the team should be aware of the goal of the team, know their respective roles and the roles of team members, and distribute work effectively (Salas, Sims, & Klein, 2004). In other words, leadership assignment is dependent on the member of the team that has the appropriate skill set required to attain the current goal of the mission. Within teams where task decisions may be distributed among team members there is evidence that teams with trained leaders still outperform those that do not (Tannenbaum, Smith-Jentsch, & Behson, 1998). The effectiveness of shared leadership among teams is dependent on each team member's knowledge about the role other team members have and their ability to play that role.

As mentioned before, the three team processes described above are the most saliently observable team process skills in video gameplay. Further, we believe video gameplay that focuses on improving these processes will contribute to the successful completion of the team's goals. In the next section we discuss how video games, specifically the features of a game, contribute to the improvement of communication, coordination, and leadership skills.

2. Video Games and Training

The video game industry accounted for $10.5 billion in U.S. sales in 2005 (NPD, 2006) which surpassed the motion picture industry ($9.2 billion for 2005) in sales. As a result of the growing gaming market, there is a growth in the online video gaming community. The most popular online games have up to 6 million subscribers worldwide per game (Maragos, 2006). This does not account for the countless millions of casual gamers that do not pay subscription fees, but utilize the online format. Of these online video games, a majority involve interaction with players both in competitive and cooperative forums. In total the online multiplayer game industry is booming and as a result presents the potential for using this popular format of interaction for education or training purposes.

In general, video games have been considered useful for improving skill-based performance in many domains (Lintern, Roscoe, Koonce, & Segal, 1990; Prensky, 2003). In addition to training technical knowledge in a specific domain, video games were quickly adapted by trainers to learn and practice psycho-motor skills, as well as perceptual and cognitive skills (Green & Bavelier, 2003; Griffith, Voloschin, Gibb, & Bailey, 1983). Findings such as these have helped to spur an increasing amount of industries to take a serious look at using video games for training purposes. Research involving the use of video games has improved recently with the quickly improving technologies in graphics and processing. These advancements allow for more realistic gameplay environments and software capabilities that provide researchers with more flexibility for training scenarios.

Whereas previously the majority of the video game training took place in aviation and military, now it is not uncommon to find video game training utilized for corporate teams (e.g., Cold Stone Creamery, Cisco, and Canon; see Reena, 2006), medical training (Ali, Mowery, Kaplan, & DeMaria, 2002; Barach, Satish, & Streufert, 2001), and even firefighting (Tate, Sibert, & King 1997). The benefit of increased interest from a variety of sources is that more and more skilled developers will be drawn to this area. The flood of new products that will follow should, however, be examined with a critical eye to make sure that targeted skills are in fact elicited through gameplay. For example, if a video game is chosen based on its superficial characteristics, such as the game's resemblance to "real" world situations or environments, it is likely that its effectiveness will be limited to only improving skills that are specific to the video game. When the goal, however, is to use video games to train teams to effectively complete a task or mission, certain video game characteristics may be irrelevant. For example, realism may be comparatively unimportant if the teamwork skills needed for completing a team's task or mission are developed by the video game.

In order to select an appropriate game for team skills development, it is important to understand the nature of the game and the game characteristics that facilitate learning these skills. The next few sections will describe the overall features of serious games, those games developed with a specific training purpose, in contrast to games which can be purchased off-the-shelf. To this end, a more specific description of video game features and how they relate or impact the team skills outlined earlier in this chapter will be addressed.

2.1. Serious Video Games vs. Commercial Off-the-Shelf Video Games

The distinction between serious games and commercial off-the-shelf (COTS) games is the impetus for developing the video game. Video games known as serious games are developed with a specific training goal in mind. On the contrary, COTS games are developed for entertainment purposes, but provide experiences that can elicit those specific to certain educational and training purposes (Woods, 2004).

A number of video games were developed for the specific goal of training target skills. An early, and still utilized, example of this is the video game *Space Fortress* (Donchin, 1995). This is a low-fidelity video game that was designed to simulate complex and dynamic flight environments. It was found to have a positive effect on pilot's attentional strategy development (Gopher, Weil, & Bareket, 1994). *Space Fortress* (Donchin, 1995) was an early example of how a game with low physical fidelity could provide training experience without matching the task specifically. The high production cost and restricted access to serious games have limited the advancement of these video games to a few industries that are able to provide the necessary resources.

In contrast to serious games, there is a potential for extracting a skills development program from COTS software (Jentsch & Bowers, 1998). An example of a COTS video game is the *Microsoft Flight Simulator* (Microsoft, 1996). There have been quite a few studies on the utility of the *Microsoft Flight Simulator* (Microsoft, 1996) in which crew resource management has been the focus (Brannick, Prince, & Salas, 2005; Prince & Jentsch, 2001). Previously the downside to COTS games was a lack of control over the simulated environment. As a result of advanced gaming technology, most games now come standard with editing software, reducing this problem. Another example of a COTS video game adapted for team training is the game *Doom* (Id Software, 1993) which was adapted to train four-person teams on small unit tactics for the U.S. Marine Corps (USMC; Gordon, 1996). The primary difficulty with using COTS video games today is isolating the team concepts for accurate measurement of performance.

Until recently the differentiation between the types of video games was much more dramatic. With the availability to improved video game development technology in conjunction with skilled programmers with COTS video games experience, the distinction has begun to blur. In addition, there is an increase in the number of serious games that are available off the shelf. For example, *America's Army* (US Army, 2002) was instituted by the military to provide a recruiting tool and to provide soldiers with another resource for learning concepts like rules of engagement, laws of war, and basic army values. In addition to providing these training benefits, the game, with over 5.5 million online subscribers, quickly became one of the most popular online action games available (America's Army, 2006).

2.2. Multiplayer Games Characteristics

Multiplayer games have features that allow more than one player to participate simultaneously in game scenarios. Regardless of whether a video game is developed for specific purposes or adapted from COTS games, there are several characteristics of multiplayer games which can, if effectively utilized, elicit the teamwork characteristics described earlier. The two most prominent features that not only distinguish them from single player video game but can also aid in team training are the mode of communication and the element of coordinated behavior to accomplish game tasks.

Games that have multiplayer capability have long been used to examine aspects of training especially in the domain of aviation. The recent increase of online multiplayer games provides a new possibility of using video games to train and practice teamwork skills (Macedonia, 2003). In addition, multiplayer games are capable of mimicking different team-oriented environments ranging from a co-located team with only two team members, to a worldwide distributed team with an almost infinite number of members.

Recent research (Weil, Hussain, Brunyé, Sidman, & Spahr, 2005) examined the use of a multiplayer game in training teamwork skills in a distributed environment. The results demonstrated instances of teamwork skills within the context of multiplayer games. Several other studies suggest that the use of video games in training can also improve skill development in coordination (Jentsch & Bowers, 1998; Stout et al., 1994; Stout, Salas, & Fowlkes, 1997), assertiveness (Brannick et al., 2005) and communication (Stout, Villegas, & Kim, 2001).

We will now briefly describe each of the two game features, communication and coordination, that can be utilized for team training. These are features unique to multiplayer games which help drive team interaction within the game. For each feature we will describe how they can promote the development of various team skills. In addition to this, we will describe one game feature, engagement, not unique to multiplayer games, but is a critical component to the successful implementation of a video game training program.

2.3. Communication Mode

Communication in multiplayer video games is dictated by the modes of communication available. There are a variety of ways in which current gaming technology can provide means of communication that would be valuable to the team. A video game may include a text-generated communication mode, a voice communication mode, or an alternate communication option.

Most multiplayer games provide a default text-generated communication mode that resembles the chat functions often available on the Internet. The primary driver of this form of communication is the typed messages that a player can provide. This is often supplemented by a set of preset commands that players can select in order to facilitate quick communication of commands that are highly relevant to the game task. For example, in *Battlefield 1942* (Electronic Arts, 2004) players can coordinate team movement through a series of single button presses that will communicate commands such as “Follow me,” “Hold fire,” or “Request reinforcements” that provide quick communication in a text presentation.

In addition to text communication many games are able to be equipped to support voice communication. This is generally embedded in the game design in newer games. For video games that do not support voice communication, downloading programs such as *Roger Wilco, Teamspeak, Team Sound, or Game Voice* will provide a voice communication medium that can run parallel with the game and enable players to talk within the group (Spohn, 2005). By using this feature, players are not restricted by having to stop to type in a request which can affect the flow of gameplay.

In addition to these types of communication there are several games that provide alternative communication options that are often unique to the game. *America’s Army* (US Army, 2002) is a good example of a game with an alternative communication mode. The game features a hand signal communication mode which is simulated by pressing designated buttons. What result is the avatar that other team members see, on the screen, displays a hand signal.

Team Communication. Communication channels have been shown to impact team interaction and information exchange (Barkhi, 2005; Hightower & Sayeed, 1996). Communication within a team is critical for information exchange. The communication modes provided in some games do not isolate the communication within teams, but instead broadcast all communication to all players involved. In an effort to better match the types of communication that would take place within an actual work team, it is critical to be able to isolate communication within teams. The tank simulation video game *Steel Beasts* (eSim, 2002) is a good example of this. In this video game, players are able to play cooperatively and even operate different aspects of the same tank (i.e., one player may act in the role of gunner and another player may act in the role of navigator). Players on each team are limited to text chat, but are only in contact with members of their team. This focuses the interactions within teams to be a streamlined set of interactions that are not distracted by the bedlam of talk that universal communication systems provide.

> *Guideline 1*
>
> *Ensure that the game has capability for team members to engage in information exchange (i.e., communicate).*

Coordination. In order for the communication mode to facilitate team coordination behaviors, the mode must facilitate interaction among teammates (Poole, Shannon, & DeSanctis, 1990). Communication modes frame the issues found in the video game such that the mode may encourage or discourage collaboration among team members. Generally the provision of communication on the game in itself is enough to help facilitate coordinating behaviors. That is, as long as the communication process is a two-way interaction, as opposed to a one-way receive and execute command method, coordination among team members should be facilitated.

> *Guideline 2*
>
> *The communication mode of the game must facilitate two-way (or more) interactions.*

Team Leadership. When discussing the utility of games for the purposes of team skills development, it is imperative to determine the structure of the team. Teams with a designated leader will experience a more structured flow of information as well as use of vertical channels between the member and the leader (Galbraith, 1973). There is evidence suggesting that hierarchical and non-hierarchical teams perform differently based on how much communication is taking place (Urban, Bowers, Monday, & Morgan, 1995). There are video games that provide a forced chain of command which could benefit teams with a hierarchical structure to them. In the video game *Battlefield 2* (Electronic Arts, 2005) the command mode properly sets up a network of communication between players in which the highest command player has only two-way communication with players who are directly below him/her in the hierarchy. Those players in turn have two-way communication with both their leader and subordinates within their own squad.

In contrast to hierarchical leadership structures, there are teams that share leadership amongst team members. In these scenarios, it is much more conducive for the development of leaders if all members are able to hone their leadership skills. In this case, the communication mode can facilitate these behaviors by providing all teammates with the ability to contact all other team members. Shared leadership skills can be improved if individuals are able to communicate with the entire group or with just specific members, specifically for situations where it would be disruptive to include all team members in the communication.

Guideline 3

Select a game that creates opportunities (i.e., task assignment or communication options) to execute team leadership behaviors or actions (shared and hierarchical).

2.4. Coordinated Gameplay

There are several methods to play video games that are considered to be multiplayer. There are competitive games in which individuals compete against opponents; these types of video games do not require teamwork to succeed. In order to elicit team behaviors, it is important to find games that provide a cooperative or interdependent multiplayer mode. There are several features that can affect the quality of coordinated gameplay. Such an example is progress monitoring. Progress monitoring features give in-mission updates on performance in order to help team members monitor the pace in which they are performing the mission. Mission planning is an additional feature that can help emphasize the organization of tasks within a team. Mission planning can address such issues as which decisions and what information are pertinent to team success. In addition to these features, games that provide dynamic gameplay or constantly changing events can also help recreate more realistic events that require team skills in order to accomplish the events successfully.

Team Communication. Communication can be facilitated by the presence of coordinated gameplay. In order to facilitate this, the game task should require interdependence between team members for success. In *Counter-Strike* (Valve, 2004), a counter-terrorism COTS video game, players can have similar roles on the same team, but the communication of locations and activity of opposing teams are crucial to the strategy and effectiveness of the team. In other words, the lone wolf strategy is not sufficient for success. A team that works interdependently and utilizes closed loop communication, in which players transmit, respond, and react to information that is exchanged, is crucial to the success of the game scenario.

Guideline 4

Select a game that supports task-interdependent team scenarios to accomplish game tasks.

Coordination. Coordination is a key component that relies on the efficient interaction between team members. The coordinated gameplay can aid this in several ways. In order to ensure that coordination behaviors are targeted, in addition to interdependency, a video game can require varying roles for team members. Although different roles can be assigned in video games where the skills of the players are generally the same, there are several video games, such as *A Tale in the Desert III* (eGenesis, 2006), that provide the option to take on different roles with different skill sets. For instance in *Tom Clancy's Rainbow Six: Raven Shield* (Ubisoft, 2003) video game players can assign themselves one of several skill sets for their character (i.e., assault skills, explosive skills, or stealth skills). Each skill set provides different strengths that, depending on the mission, can be crucial for success. In many cases, several or all of the different skill sets are required to complete missions. This presents a situation where team success is based on the efficiency of the interaction of the team and its coordinated actions.

In addition to providing differential roles for players, there are progress cues that can help the team to maintain situation awareness within gameplay. Often games will have maps or keys that help orient players on their surroundings. In the *Distributed Dynamic Decision-Making* (DDD) software developed by the Department of Defense, players are in charge of monitoring different sections of a map (Miller, Young, Kleinman, & Serfaty, 1998). Each team member can view the location of the team's assets on the map, but in addition to this, individual players on the team can view opposing assets in their area. Done efficiently, players who coordinate well should be able to synchronize the movements of their team assets to avoid attack from opponents.

> *Guideline 5*
>
> *Select a game that supports differential individual roles within team scenarios.*
>
> *Guideline 6*
>
> *Select a game that provides progress cues (i.e., maps, reports, etc.)*

Team Leadership. Part of team leadership is the development of team orientation where team members become familiar with what is required of them to complete the team task. Of the coordination features contained in video games, the mission planning feature can help to provide teams with exposure to the process of assigning responsibilities to a team. An example of this is the mission planning available on *Ghost Recon: Advanced Warfighter* (Ubisoft, 2006). Players are able to view satellite images within a couple of blocks of where their team members are located. From this, players are able to designate waypoints and action points which players can use as a guide for strategic action in the game. This helps players understand what is expected and allows them to monitor how closely they adhere to the orders.

In addition to this, adaptation is a key element of team leadership. Games can facilitate this with dynamic gameplay elements. Games that have multiple possible solutions enhance a team's ability to adapt to unexpected outcomes. An example of a video game that executes this well is the *Adaptive Thinking and Leadership Simulation* (Raybourn, Deagle, Mendini, & Heneghan, 2005) developed to help train US special forces troops. Players assume a role in a constantly changing environment in order to improve their adaptation skills while in scenarios that have ill-defined consequences for different courses of action (Sandia National Laboratories, 2005).

Guideline 7

Select a game that provides mission planning features.

Guideline 8

Select a game that has more than one solution to reach the team's goals.

2.5. Engagement

The features that distinguish a multiplayer game from the single player variety are critical to developing effective team training. Another critical feature of games that is not specific to multiplayer games is engagement. That is, does the game draw players in and have them not only actively involved in the storyline of the game, but also interested to continue play.

The serious games industry has struggled with this in the past. Arguments have been made stating that educational games, which comprise 7–9% of the computer and video game market (ESA, 2006), should be both fun and informative (Quinn, 1997). Developers find it difficult to incorporate the educational needs of the game and at the same time deliver a product that keeps players coming back for more.

Early research on the features that draw people to games suggest that external features of the game (i.e., visual appeal) are not enough to engage a user (Bowman, 1982), but that a combination of challenge, fantasy, and curiosity contribute to the level of engagement (Provenzo, 1991). Dickey (2005) outlined several specific features that help promote an engaged learning environment. By engaging a person to the task, that person becomes an active learner in the process.

By considering these features of engaged learning, Dickey (2005) suggested that the critical elements of game design are point of view, narrative arc, and interactive choice. Point of view research seems to suggest that playing first person point of view games seems to enhance feelings of involvement (Tamborini et al., 2001). A first person point of view game occurs when the player is viewing a screen as if the player were experiencing the game through the eyes of the avatar he/she is controlling. Discussions on narrative arc contend that games that have a story, but can be affected by choice are most effective in engaging users (Dickey, 2005). The level of interaction with the game environment is also said to improve engagement to the game. By providing hooks or choices the player must make, the game takes on a more personal feel for players, therefore the video game is engaging.

The cost of developing a game that provides a training or educational message, but does not engage the user is high. Not only are resources spent on development of the game, but then costs must be dispensed in order to motivate users to utilize the product. By developing educational or training video games that are engaging to the user, there is potential for not having to prompt or lure users to continue with their training.

Guideline 9

Select first person point of view games.

Guideline 10

Select games with engaging and realistic story lines.

Guideline 11

Select games that provide choices that directly affect the outcome of the game.

3. Conclusion and Future Research

The goal of this chapter is twofold: to provide guidelines for choosing games for team training purposes (see Table 1), and to provide a theoretical platform to drive future empirical research on the topic. As a result of the variety of games with a variety of

Table 1
Guidelines for video game selection.

Target team skill	Game feature	Guideline
Communication	Communication mode	Ensure that the game has capability for team members to engage in information exchange (i.e., communicate)
Coordination	Communication mode	The communication mode of the game must facilitate two-way (or more) interactions
Leadership	Communication mode	Select a game that creates opportunities to execute team leadership behaviors or actions (shared and hierarchical)
Communication	Coordinated gameplay	Select a game that supports task-interdependent team scenarios to accomplish game tasks
Coordination	Coordinated gameplay	Select a game that supports differential individual roles within team scenarios
Coordination	Coordinated gameplay	Select games that provide progress cues (i.e., maps, reports, etc.)
Leadership	Coordinated gameplay	Select games that provide mission planning features
Leadership	Coordinated gameplay	Select games that have more than one solution to reach the team's goals
	Engagement	Select first person point of view games
	Engagement	Select games with interesting story lines
	Engagement	Select games that provide choices that directly affect the outcome the game

features in both COTS and serious games, it is important to match the game to the training goals. This chapter served to initiate thought about the characteristics of video games that may impact the training of teams. That being said, one should consider the relatively sparse research in this area and consider the further examination of this to provide more solid evidence. For example, focusing on the examination of specific game features to identify by what methods these features can improve team skills is important for the advancement of using video games for training (see Table 2). In addition, it would be helpful to identify whether or not there are specific team skills that will transfer to the actual task from a video game that does not specifically match the task they would perform on the job. Identifying these skills can direct trainers to search games by skill instead of trying to match the domain in which the skill will be implemented.

Table 2
Potential video games that can be used for training.

Video game	Video game characteristic	Targeted team skills/behaviors
Counter Strike (Valve, 2004)	Communication mode – Voice – Chat – Open communication to all team members	Communication Coordination
	Coordinated gameplay – Interdependent gameplay – Player location cue – Multiple solutions	Communication Coordinated Leadership
	Engagement – First Person Shooter – Realistic equipment actions	
Microsoft Flight Simulator X (Microsoft, 2006)	Communication mode – Voice – Chat – Two way interaction with copilot	Communication Coordination
	Coordinated gameplay – Flight planner – Play as pilot, copilot or air traffic controller – Interdependent team scenarios – Maps, Radar and Realistic Instrument panels – Fly in varying conditions (i.e., weather, daylight, heavy traffic, etc.) – Multiple airports and runways to take off and land from	Communication Coordination Leadership
	Engagement – 3-D virtual cockpit – High fidelity – First person point of view – Pilot choices directly affect the performance of the scenario	

Table 2
(*Continued*)

Video game	Video game characteristic	Targeted team skills/behaviors
Doom (Id Software, 1993)	Communication mode – Chat – Open communication among players	Communication Coordination
	Coordinated gameplay – Multiplayer co-op mode – Scenario editing software	Coordination
	Engagement – First Person Shooter – Space marine story line	
America's Army (US Army, 2002)	Communication mode – chat – button press hand-signals – voice – open communication	Communication Coordination
	Coordinated gameplay – Interdependent gameplay scenarios – Maps – Multiple solutions to scenarios	Communication Coordination Leadership
	Engagement – First person shooter – Player choices affect outcome	
Battlefield 1942 (EA Games, 2004)	Communication mode – Chat – Open communication among team members	Communication Coordination
	Coordinated gameplay – Differential individual roles – Maps and player location indicator	Coordination
	Engagement – First Person Shooter	
Steel Beasts (eSim, 2002)	Communication mode – Chat – Open communication within team	Communication Coordination
	Coordination mode – Interdependent gameplay scenarios – Play as tank commander or tank gunner – Maps – Pre-mission objectives list – Extensive permission and during mission planning – Multiple solutions to accomplishing game task	Communication Coordination Leadership

(*continued on next page*)

Table 2
(*Continued*)

Video game	Video game characteristic	Targeted team skills/behaviors
	Engagement – First person point of view – Player choices have direct affect on outcomes	
Battlefield 2 (EA Games, 2005)	Communication mode – Voice – Chat – Two way communication depending on role of team member – Hierarchical team structure – Commander mode	Communication Coordination Leadership
	Coordinated gameplay – Interdependent gameplay scenarios – Differential individual roles – Real time map to monitor progress – Commander option to monitor and plan mission real time	Communication Coordination Leadership
	Engagement – First Person Shooter – Modern Military Shooter – Player choices directly affect outcomes	
Rainbow Six 3: Raven Shield (Ubisoft, 2003)	Communication mode – Chat – Open team interaction	Communication Coordination
	Coordinated gameplay – Interdependent gameplay scenarios – Differential roles within team – Maps – Pre-mission planning	Communication Coordination Leadership
	Engagement – First Person Shooter – Counter terror team story line – Player choices affect the outcome of the game	
DDD (Miller et al., 1998)	Communication mode – Voice – Chat – Open communication amongst team	Communication Coordination
	Coordinated gameplay – Interdependent gameplay scenarios – Differential roles within the team – Maps and real time progress reports – Multiple solutions to scenarios	Communication Coordination Leadership

Table 2
(*Continued*)

Video game	Video game characteristic	Targeted team skills/behaviors
Ghost Recon: Advanced Warfighter (Ubisoft, 2006)	Communication mode – voice – chat – Heads Up Display which links to other team members and drone Unmanned Aerial Vehicles – Open communication with team members	Communication Coordination
	Coordinated gameplay – Interdependent gameplay scenarios – Maps and Heads Up Display which links to other team members and drone Unmanned Aerial Vehicles – real time tactical map for designating waypoints and planning strategy	Communication Coordination Leadership
	Engagement – First Person Shooter – Modern Military themed story line – Player choices affect outcome of game	
Adaptive Thinking and Leadership Simulation (Sandia National Laboratories, 2005)	Communication mode – Communication training platform – Interaction between role player and spectators	Communication
	Coordinated gameplay – Differential individual roles within game	Coordination
	Engagement – First Person point of view – Modern military theme story line – Player choices directly affect outcomes	
A Tale in the Desert III (eGenesis, 2006)	Communication mode – Chat – Open communication with community – Hierarchical player system	Communication Coordination Leadership
	Coordination gameplay – Long term planning focus – Interdependent society simulation game – Differential roles within the society – Multiple options in open ended game world	Communication Coordination Leadership
	Engagement – Ancient Egyptian society story line – Players can create laws and take actions that affect the outcomes of all players in the society	

Note: The games above are not endorsed for training by the authors but are discussed in this chapter in order to demonstrate game features that may help team training.

One should note that this is only an initial theoretical attempt to discuss the use of multiplayer video games for team training, and the guidelines provided should be taken as a preliminary guide to select video games for team skills training. Further research into their implementation will help focus attention on game features that are beneficial for team training and could drive game development in both the COTS and more likely the serious games arena.

References

Ali, M. R., Mowery, Y., Kaplan, B., & DeMaria, E. J. (2002). Training the novice in laparoscopy. More challenge is better. *Surgical Endoscopy, 16*(12), 1732–1736.

America's Army (November, 2006). America's Army official home page. Retrieved November 2, 2006 from http://www.americasarmy.com/.

Ancona, D. G. & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science, 3*(3), 321–341.

Argote, L. & McGrath, J. E. (1993). Group processes in organizations: Continuity and change. In C. L. Cooper & I. T. Robertson (Eds), *International review of industrial and organizational psychology* (pp. 333–389). New York: John Wiley & Sons.

Barach, P., Satish, U., & Steuffert, S. (2001). Healthcare assessment and performance. *Simulation Gaming, 32*, 147–151.

Barkhi, R. (2005). Information exchange and induced cooperation in group decision support systems. *Communication Research, 32*(5), 646–678.

Bowman, R. F. (1982). A Pac-Man theory of motivation. Tactical implications for classroom instruction. *Educational Technology, 22*(9), 14–17.

Brannick, M. T., Prince, C., & Salas, E. (2005). Can PC-based systems enhance teamwork in the cockpit? *International Journal of Aviation Psychology, 15*(2), 173–187.

Cannon-Bowers, J. A. & Salas, E. (1997). Teamwork competencies: The interaction of team member knowledge, skills, and attitudes. In H. F. O'Neil (Ed.), *Workforce readiness: Competencies and assessment* (pp. 151–174). Mahwah, NJ: Lawrence Erlbaum.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds), *Team effectiveness and decision making in organizations* (pp. 333–375). San Francisco, CA: Jossey-Bass.

Day, D. V., Gronn, P., & Salas, E. (2004). Leadership capacity in teams. *Leadership Quarterly, 15*, 857–880.

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

Donchin, E. (1995). Video games as research tools: The space fortress game. *Behavior Research Methods, Instruments & Computers, 27*(2), 217–223.

Driskell, J. E. & Dwyer, D. J. (1984). Microcomputer videogame based training. *Educational Technology, 24*(2), 11–17.

eGenesis (2006). *A Tale in the Desert III [computer software]*. Pittsburgh, PA: eGenesis.

Electronic Arts Games (2004). *Battlefield 1942: World War II Anthology [computer software]*. Redwood City, CA: Electronic Arts Games.

Electronic Arts Games (2005). *Battlefield 2 [computer software]*. Redwood City, CA: Electronic Arts Games.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32–64.

eSim Games (2002). *Steel Beasts [computer software]*. Sunnyvale, CA: eSim Games.

Entertainment Software Association (2006). 2006 Sales, demographic, and usage data: Essential facts about the computer and video game industry. Retrieved November 1, 2006 from http://www.theesa.com/archives/files/Essential%20Facts%202006.pdf.

Galbraith, J. (1973). *Designing complex organizations*. Reading, MA: Addison-Wesley.

Glickman, A. S. et al. (1987). *The evolution of teamwork skills: An empirical assessment with implications for training* (ONR Tech. Report 87–016). Orlando, FL: Naval Training Systems Center, Human Factors Division.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*(3), 387–405.

Gordon, S. M. (1996). Doomed! *Marines, 25*(7), 20–21.

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selection attention. *Nature, 423*, 534–537.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye–hand motor coordination of video-game users and non-users. *Perceptual Motor Skills, 57*(1), 155–158.

Hightower, R. T. & Sayeed, L. (1996). Effects of communication mode and prediscussion information distribution characteristics on information exchange in groups. *Information Systems Research, 7*(4), 451–465.

Hollenbeck, J. R. et al. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology, 80*, 292–316.

Id Software (1993). *Doom [computer software]*. Mesquite, TX: Id Software.

Jentsch, F. & Bowers, C. A. (1998). Evidence for the validity of PC-based simulations in studying aircrew coordination. *The International Journal of Aviation Psychology, 8*(3), 243–260.

Kozlowski, S. W. J. & Bell, B. S. (2003). Work groups and teams in organizations. In W. Borman, D. R. Ilgen, & R. J. Klimoski (Vol. Eds), *Handbook of psychology: Vol. 12. Industrial and organizational psychology* (pp. 333–375). Hoboken, NJ: John Wiley.

Levine, J. M. & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41*, 585–634.

Lintern, G., Roscoe, S. N., Koonce, J. M., & Segal, L. D. (1990). Transfer of landing skill in beginning flight training. *Human Factors, 32*, 319–327.

Macedonia, M. (2003). Games soldiers play. *IEEE Spectrum*, 32–37.

Maragos, N. (2006). World of warcraft reaches 6 million subscribers total. Retrieved December 9, 2006 from http://www.gamasutra.com/php-bin/news_index.php?story=8352.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review, 4*, 95–105.

McIntyre, R. M. & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds), *Team effectiveness and decision making in organizations* (pp. 9–45). San Francisco, CA: Jossey-Bass.

Microsoft (1996). *Microsoft Flight Simulator [computer software]*. Redmond, WA: Microsoft.

Microsoft (2006). *Microsoft Flight Simulator X [computer software]*. Redmond, WA: Microsoft.

Miller, D. L., Young, P., Kleinman, D., & Serfaty, D. (1998). *Distributed dynamic decision-making simulation phase I: Release notes and user's manual*. Woburn, MA: Aptima.

Morgan, B. B., Glickman, A. S., Woodard, E. A., Blaiwes, A. S., & Salas, E. (1986). *Measurement of team behavior in navy training environment* (Technical Report TR-86-014). Orlando, FL: Navel Training Systems Center, Human Factors Division.

Nichols, L., DeFriese, A. M., & Malone, C. C. (2002). Team process. In G. D. Heinemann & A. M. Zeiss (Eds), *Team performance in health care: Assessment and development* (pp. 71–88). New York: Kluwer Academic/Plenum.

NPD Group (2006, January 17). The NPD group reports annual 2005 U.S. video game industry retail sales. Retrieved November 2, 2006, from http://www.npd.com/dynamic/releases/press_060117.html.

O'Neil, H. & Fisher, Y. C. (2004). A technology to support leader development: Computer games. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leader development for transforming organizations: Growing leaders for tomorrow*. Mahwah, NJ: Lawrence Erlbaum.

Poole, M., Shannon, D., & DeSanctis, G. (1992). Communication media and negotiation processes. In L. Putnam & M. Roloff (Eds), *Framing, reframing, and issue development* (pp. 46–66). Newbury Park, CA: Sage.

Porter, C. et al. (2003). Towards understanding backing up behaviors in teams: The role of personality and legitimacy of need. *Journal of Applied Psychology, 88*, 391–403.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE), 1*(1), 21.

Prince, C. & Jentsch, F. (2001). Aviation crew resource management training with low-fidelity devices. In E. Salas, C. A. Bowers, & E. Edens (Eds), *Improving teamwork in organizations: Applications of resource management training* (pp. 147–164). Mahwah, NJ: Lawrence Erlbaum.

Provenzo, E. F. (1991). *Video kids: Making sense of nintendo*. Cambridge, MA: Harvard University Press.

Quinn, C. N. (1997). Engaging learning. *Instructional Technology Forum*, Retrieved December 8, 2006 from http://it.coe.uga.edu/itforum/paper18/paper18.html.

Raybourn, E. M., Deagle, E., Mendini, K., & Heneghan, J. (2005). Adaptive thinking and leadership simulation game training for special forces officers. *Proceedings of the Interservice/Industry Training, Simulation, & Education Conference*, Orlando, FL.

Reena, J. (2006). http://www.businessweek.com/magazine/content/06_13/b3977062.htm On-The-Job Video Gaming.

Salas, E., Burke, C. S., & Stagl, K. C. (2004). Developing teams and team leaders: Strategies and principles. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leader development for transforming organizations* (pp. 325–355). Mahwah, NJ: Lawrence Erlbaum.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*(5), 555–599.

Salas, E., Sims, D. E., & Klein, C. (2004). Cooperation at work. In C. D. Speilberger (Ed.), *Encyclopedia of applied psychology* (Vol. 1, pp. 497–505). San Diego, CA: Academic Press.

Sandia National Laboratories (2005, July 26). Sandia and U.S. Army's America's Army simulation helps special forces learn adaptive thinking and cultural awareness. Retrieved November 22, 2006 from http://www.sandia.gov/news-center/news-releases/2005/def-nonprolif-sec/americaarmy.html.

Siegel, A. I. & Federman, P. J. (1973). Communications content training as an ingredient in effective team performance. *Ergonomics, 16*(4), 403–416.

Spohn, D. (2005). Making yourself heard: Headsets becoming essential game hardware. Retrieved March 24, 2005, from the about.com website http://internetgames.about.com/cs/hardware/a/aavoice.htm.

Stout, R. J., Salas, E., & Carson, R. (1994). Individual task proficiency and team process: What's important for team functioning. *Military Psychology, 6*, 177—192.

Stout, R. J., Salas, E., & Fowlkes, J. E. (1997). Enhancing teamwork in complex environments through team training. *Group Dynamics: Theory, Research, and Practice, 1*(2), 169–182.

Stout, R. J., Villegas, J., & Kim, H. (2001). Enhancing learning through use of interactive tools on health-related websites. *Health Education Research, 16*(6), 721–733.

Sundstrom, E., de Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45*, 120–133.

Sundstrom, E., McIntyre, M., Halfhill, T., & Richards, H. (2000). Work groups: From the Hawthorne studies to work teams of the 1990s and beyond. *Group Dynamics: Theory, Research, and Practice, 4*, 44–67.

Tamborini, R., Eastin, M., Lachlin, K., Skalski, P., Fediuk, T., & Brady, R. (2001). *Hostile thoughts, presence and violent virtual video games*. Paper presented at the *51st Annual Convention of the International Communication Association*, Washington, DC.

Tannenbaum, S. I., Smith-Jentsch, K. A., & Behson, S. J. (1998). Training team leaders to facilitate team learning and performance. In J. A. Cannon-Bowers & E. Salas (Eds), *Making Decisions Under Stress*, Washington, DC: American Psychological Association.

Tate, D., Sibert, L., & King, T. (1997). Using virtual environments to train firefighters. *IEEE Computer Graphics & Applications, 17*(6), 23–29.

The 4th edition of the electronic sports world cup from 27th June to 2nd July 2006 in Paris-Bercy. Retrieved December 2, 2006 from http://www.esworldcup.com/2006/?act=document_info& id_document= 1.

UbiSoft (2003). *Tom Clancy's Rainbow Six 3: Raven Shield [computer software]*. San Francisco, CA: UbiSoft.

UbiSoft (2006). *Tom Clancy's Ghost Recon: Advanced Warfighter [computer software]*. San Francisco, CA: UbiSoft.

United States Army (2002). *America's Army [computer software]*. Monterey, CA.

Urban, J. M., Bowers, C. A., Monday, S. D., & Morgan, B. B. (1995). Workload, team structure, and communication in team performance. *Military Psychology, 7*(2), 123–139.

Valve (2004). *Counter Strike: Source [computer software]*. Kirkland, WA: Valve.

Weil, S. A., Hussain, T. S., Brunyé, T., Sidman, J., & Spahr, L. (2005). The use of massive multi-player gaming technology for military training: A preliminary evaluation. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Santa Monica, CA: HFES.

Woods, S. (2004). Loading the dice: The challenge of serious videogames. *The International Journal of Computer Game Research*, 4(1), Retrieved October 4, 2006 from http://www.gamestudies.org/0401/woods/.

This page intentionally left blank

PART 3

INDIVIDUAL LEARNING

This page intentionally left blank

THE EFFECTS OF CHANGING RESOURCES ON GAME PERFORMANCE, SUBJECTIVE WORKLOAD, AND STRATEGIES

Adrienne Y. Lee and Aaron M. Perez

New Mexico State University

Abstract

This chapter describes research which examines subjective workload and performance in the context of a gaming environment. Actual workload is varied by changing the level of resources available to players. A transfer methodology was used where participants were either given a lot of resources initially (high resource) and then transferred to a low resource condition or they were given fewer resources (low resource) and kept at the same low resource level. For Test 1, workload judgments by high resources participants matched their performance, while performance by low resource participants was generally poor (and did not match workload judgments). However, for Task 2, although some participants who transferred from high to low resources still thought that the workload was low, the workload judgments for both transfer groups generally matched performance. These results indicate that dissociations between subjective workload and actual workload may exist but can be overcome with further training trials. Other interpretations, implications and further studies are discussed.

Tasks in complex, changing environments often require us to make quick decisions with limited resources and, in some cases, these decision may affect other people's lives. The mental effort that it takes to make such decisions under stress can be called mental workload (Eggemeier, 1988). Much of the literature in mental workload (Annett, 2002; Pickup, Wilson, Sharpies, Norris, Clarke, & Young, 2005; Stanton & Young, 2005; Young & Stanton, 2002, 2004), has focused on vehicle use; but, even traditional treatises on workload, such as Jex (1988), emphasize that the general principles should be applicable to any dynamic system including gaming simulations.

A number of concepts such as attentional capacity or resources, performance, and expertise are linked with mental workload (Brown & Boltz, 2002). Attentional capacity is

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

the idea that we only have a certain limited amount of attentional resources (Kahneman, 1973). If one thinks about a tank filled with water, the water represents the amount of attention available (attentional capacity) with different tasks requiring different amounts of the water (attention). Although attentional capacity for most tasks has been assumed to be fixed (Wickens, 1984, 1992), in certain circumstances, one can increase this pool. For example, a person can become an expert car driver and be able to converse with a passenger while driving. Thus, when a person becomes an expert in an area, fewer attentional resources are often needed due to automaticity (Schneider & Shiffrin, 1977; see also Jex, 1988). However, if something were to happen, such as rocks started falling onto the road as one is driving, performance on primary tasks (driving) may not measurably decrease, but a secondary task, such as chatting, may cease entirely. Thus, an additional task can result in an increase in mental workload that is reflected in decreased performance (Eggemeier & Wilson, 1991).

On the other hand, the person may react to a high level of workload or a change in workload by using heuristics to lower the cognitive load (Payne, Bettman, & Johnson, 1988). For example, if a mathematician is asked to solve complicated math problems and a time constraint is added, time-accuracy trade-off would predict that the person would have reduced performance. However, an expert mathematician may have many math short-cuts or rules of thumb (heuristics) that could help to reduce the amount of time without reducing performance. Such heuristics may allow performance for automatic tasks to remain virtually the same. Therefore, although some perceptual measures may not change, attentional resources could vary directly with task demands (Young & Stanton, 2002). However, a difference exists between actual workload, as indicated by increases in task demands, versus subjective mental workload, where an individual may or may not believe that an increase in workload has occurred.

This chapter describes research in which we illustrate methodology for testing factors such as subjective workload and performance in the context of a gaming environment. Actual workload is varied through providing different levels of actual resources to participants and both subjective workload and performance are measured. These differing levels of actual resources provide increased levels of task demand which presumably affect subjective judgments and performance. The focus is not merely on performance at specific (different) levels but what happens when levels are changed. When levels are changed, we expect that people will be able to use their knowledge from previous situations, transfer, and therefore they should experience higher levels of performance and reduced levels of subjective workload (Lee, 1998). Thus, a transfer study methodology is good for studying all types of mental functioning, particularly for complex, gaming environments.

1. Subjective Workload

Subjective workload is an individual's personal view as to how much effort he/she is expending or encountering at a particular point in time. In instances where we have finite resources, we may either view the situation in a neutral or positive way with no effect on performance or we could view the situation in a negative way and performance itself might be affected. Moreover, researchers have found both dissociation and a concordance between subjective ratings of workload and performance (Vidulich & Wickens, 1986; see Zhang & Luximon, 2006, for a recent review that still supports subjective measures).

Although performance can improve with greater resource supply in resource limited tasks, Yeh and Wickens (1988) claim that performance and subjective workload can be dissociated because both performance and workload are multidimensional. Dissociations are most likely to occur when people are underloaded or overloaded.[1] For example, a simple dual task may cause higher subjective workload measures and better performance as compared to a difficult single task which causes lower subjective workload measures and poorer performance. In some ways, this discussion is similar to the Yerkes-Dodson Law (1908) in which optimal performance would be expected when a certain level of simulation (not too much and not too little) is given. Essentially, an individual must perceive that he/she can perform the task and the task must not be too difficult.

However, the discussion of workload needs to separate out the actual workload from the perceived workload and actual performance. In this study, varying levels of resources (actual workload) should cause participants some stress and thereby result in differences in subjective workload.

2. NASA-TLX

The most common method to measure workload is through subjective workload tests. Annett (2002) argues that subjective measures may be useful in areas such as comfort, annoyance and workload and these measures may be effective in specific situations. For this study, we chose the NASA-TLX. Although other measures are available, the NASA-TLX measure has been widely validated (Eggemeier, 1988; Hart & Staveland, 1988; Lysaght et al., 1989; Nygren, 1991; Wierwille & Eggemeier, 1993) and the degree of intrusion on the participant is low. (See Rubio, Diaz, Martin, & Puente (2004) for review of NASA-TLX and comparison with other measures. They recommend the NASA-TLX for prediction of performance for individuals.)

3. Workload, Simulations and Gaming

Within experimental settings, actual workload can be manipulated in various ways. A dual task methodology in which a participant does two tasks at once is often used because the change in workload by adding the second task can be measured (Brown & Boltz, 2002). Many studies use physical devices (e.g., tapping certain keys, following a cursor

[1] Underload or low workload can be as much of a problem as high workload due to boredom factors (Gregoriades & Sutcliffe (2006)).

and entering information.) Requiring physical action is common in workload studies. For example, some experimenters have measured drivers' ability by manipulating changes in traffic (Collet, Petit, Champely, & Dittmar, 2003; Stanton & Young, 2005; Verwey & Veltman, 1996; Young & Stanton, 2002, 2004), driving simulators (Collet, Petit, Champely, & Dittmar, 2003; Lansdown, Brook-Carter, & Kersloot, 2004), and cell phone use while driving (Fairclough, Ashby, Ross, & Parkes, 1991; Matthews, Legg, & Charlton, 2003; Parkes, Fairclough, & Ashby, 1993; Recarte & Nunes, 2003). Although these studies utilize simulations, they often do not require purely cognitive manipulations such as those found in air traffic control (Ahlstrom, 2005; Averty, Collet, Dittmar, Athenes, & Vernet-Maury, 2004; Hopp, Smith, Clegg, & Hyeggestad, 2005; Lamoureaux, 1999; Metzger & Parasuraman, 2005), computer use (Hjortskov et al. 2004) or management simulations (Dumblekar, 2004). These simulations can produce measurable changes in workload without adding a dual task.

Simulations are related to gaming environments, especially with the latest technology available, because they place a participant in a situation where they need to use their knowledge and skills to perform a task in a realistic environment. In the case of driving simulators and cell phone use, the goal is to continue driving well. Gaming environments differ slightly because they often have a goal of "winning" (e.g., taking out an enemy, amassing more resources than opponents, making it to a goal before others do). Thus, a gaming environment may produce different results from a straight simulation because participants are asked to perform with cognitive goals ("winning") added.

The current research uses a computer simulation or gaming environment that mimics a military personnel's actual resource allocation decisions and reflects conditions that shift under changing world and environmental conditions. This computer gaming environment's main strength is its dynamic and somewhat unpredictable nature which challenges the participant to constantly monitor the environment in order to correctly adjust the resources allocated. For this study, participants experienced high and low workload conditions, similar to Metzer and Parasuraman (2005), in what could be considered more cognitive workload conditions (decision making) without the focus on physical device performance. In this case, workload was determined by the amount of resources available for allocation by participants.

However, this study also differs from previous work because the main focus of the study was on measuring changes in performance when participants either received a change in the amount of resources (from high to low or low to high) or no change at all (high for both tests, low for both tests). Thus, this study was designed to measure transfer between high and low, rather than a single measure of workload, and the effects on performance.

4. Summary

This experiment simulates an environment where people have different amounts of resources, and tests whether changing resources (actual workload) affects performance, subjective workload, and strategy. Based on previous findings (Urban, Weaver, Bowers, & Rhodenizer, 1996), we hypothesized that participants assigned to conditions in which there are fewer resources have higher subjective ratings of workload than those participants in higher resource conditions and yield lower performance scores and visa versa. However, when participants with high amounts of resources are switched to a lower resource condition, their performance may be worse than those who started in a lower resource condition and their subjective workload ratings will be even lower than those who experienced the low resource condition to start. In addition, because maximal performance requires an optimal allocation of resources (Ball, Langholtz, Auble, & Sopchak, 1998), and no optimal allocation is specified for participants before using the simulation, participants should have no a priori thoughts about what their own level performance should be.

5. Method

5.1. Participants

Forty-one undergraduate students, 18 males and 23 females, enrolled in Introductory Psychology, were voluntary participants in this experiment and were compensated with class credit. Participants were required to have computer experience and to be fluent in English. Two participants were not included because of computer failure.

5.2. Design

This was a between-subjects design with two conditions. The practice mission for both conditions was exactly the same. Each participant first completed a practice trial and two test trials. For the test trials, participants either stayed in the same condition (control: low to low) or transferred to a different resource condition (experimental: high to low). The allotment of financial resources for each mission was 900 for practice missions, 200 for low resource missions, and 1600 for the high resource missions.

5.3. Materials

Participants used the simulation, "Real War: Rogue States" developed by Rival Interactive Inc., a Cornerstone Industry, Inc., and Simon and Schuster Interactive publication operation of Viacom International Inc. This game is based on the Joint Chiefs of Staff training simulation for the United States military. With this simulation, the operator has the flexibility to allot resources in $25 increments from $0 to $2600. These financial resources can be used to build buildings (e.g., barracks and vehicle yards) and to order troops and vehicles (such as tanks) from these buildings. The main objectives of this simulation were for the operator to build up an army with the allotted financial resources, maintain that army, and find and eliminate the enemy targets. The simulations were recorded using Camtasia Studio Version 2.1©, by Tech Smith Corp, www.techsmith.com. This software package allows users to record screen action and was utilized to record the mission scenarios and replay them for later analysis.

5.3.1. Measures

In order to ensure that participants had equivalent backgrounds and cognitive skills before testing, we administered a short-term memory (STM) task and a working memory (WM) task to every participant. The STM task was the Wechsler Digit Span Forward task (Wechsler, 1981). Series of 3–9 number combinations were read to each participant. Each participant was instructed to repeat a series of numbers in the same order read to them. Each series of words were presented in two trials. The task ended when the participant could no longer recall any one of the two number combinations. Scoring was determined by the last recalled series of numbers.

The WM task used was the Operation Span (O-Span) (Turner & Engle, 1989). This task required participants to memorize a series of words while also solving math problems. For example, a participant would see the following, "Is $(4 \times 2) + 1 = 9$? Cow." At that point, the participant would read the math problem aloud, verify the answer as correct or incorrect by saying "Yes" or "No," and then say the word, "Cow." The experimenter would then press a key, which triggered the appearance of another math problem/word combination. After a series of two to six of these combinations, three questions marks appeared on the screen (???), prompting, the participants to write the memorized words on a sheet of paper, in the order that they remember seeing them. If the participants could not remember any word, they were given instructions to leave that blank empty. Participants were not allowed to pause between each set of words and were not allowed more than nine math errors. Scoring for this task was done by the all-or-none method. For example, if the participants wrote all of the correct words in a series of blanks for number one they would receive a score for the number of words in that series, and given a score of zero for each series of words in which there was a blank, a misplaced word, or a misspelled word, to the extent that it spelled another coherent word.

5.3.2. Workload

Workload was measured by using the NASA-TLX. Inquiries in the NASA-TLX include perceived frustration level, temporal demand, physical demand, mental demand, effort, and performance level. For the first section, participants made ratings on a 20-point scale from low to high. For the second section, participants were asked to compare and weigh each factor. For example, which factor "Performance" or "Temporal Demand," was the "most important contributor to workload" for the previous task? The overall workload score is obtained by two measures. Ratings were computed by taking the rating provided by the user, 12 for example, and multiplying it by 5 (this simple multiplication allows for a possible rating of 100 on a 20-point scale). The formula is $(R \times 5) = \mathrm{WL}$, where R equals the rating given by each user. Weights are computed by totaling the number of times each factor was chosen and dividing that number by 15, the number of comparisons made. This formulas is $(T/15)$, where T equals the number of times each rating was selected. The total score is derived by multiplying each rating by its weight and dividing by 15. The overall workload formula is $(R \times T)/15$.

5.4. Procedure

After signing the consent form, participants answered a biographical questionnaire to assess computer experience and fluency in the English language. Next, participants took the STM task. Participants started with two simulation-based training sessions. This training consisted of a description on the nature of the computer simulation; the objectives and essential resources; and a short introduction explaining how to construct buildings, selecting units and moving these selected units to enemy locations, defending their own base and attacking and ultimately eliminating the enemy targets. Participants then watched a 7-min video to understand better the advanced process of constructing buildings, where to find and how to attack their enemies, and most importantly how to acquire and maintain their supplies.

Participants performed a 20-min practice trial to apply the skills learned during training. Initially, the participants started this trial with 900 financial resources, and the experimenter reminded them about the primary objectives of the simulation. After 10 min, the simulation was paused to assess subjective workload. After the questionnaire was complete, participants notified the experimenter. The experimenter again reminded each participant about the primary objectives of the simulation and then resumed the simulation. After 20 min, the simulation was again paused, and the participants were given a final subjective workload assessment of the simulation, followed by a 5 min break. (During the training, participants were allowed to ask any questions about the simulation, but during testing no questions about the simulation were allowed.)

Test 1. At the start of each test, the experimenter told the participants about the amount of allotted resources. If they were allotted a low amount of financial resources for this scenario, they were told "because so many resources were in the last mission, you will be given a lesser amount for this one." The timer was started when the mission began. After 10 min, the simulation was paused to assess subjective workload. The simulation was resumed after completion of the assessment, and participants continued interacting with the simulation for an additional ten minutes. After another 10 min, the participants were again given an assessment of subjective workload. Once the participants finished this questionnaire, they were asked to exit the room briefly so that we had the chance to write down their performance scores and set up the next testing scenario.

Test 2. The second testing scenario used the same procedure as the first testing scenario. However, when the participants finished the second subjective workload measure, they remained in the experiment room. Next, the O-Span task was administered. When participants complete this task, they were debriefed on the nature of the experiment and thanked for their participation.

6. Results

Participants were given a series of questionnaires to assess previous computer and gaming experience. ANOVAs were performed on each of these experience variables to see if there were any differences before starting but no differences between the two conditions (high resources first versus low resources first) were found.

During the practice session, all participants regardless of condition received the same amount of resources. An ANOVA was performed with perceived workload difference and condition as independent variables and performance as the dependent variable. No differences between groups were found and workload did not predict performance.

For the first test, half of the participants received a high allocation of assets (high resource) and half received a low allocation (low resource). A repeated measures ANOVA was performed using the perceived workload for both time 2 and time 1 as the dependent variable and condition as the independent variable. Participants in the high resource condition reduced their workload score, while participants in the low resource condition increased their workload score, $F(1, 39) = 4.7, p < 0.04$. Note that the initial rating is the same but that the change comes at time 2. This would be expected because after time has passed participants in the high resource condition should not feel as stressed (Table 1).

An ANOVA was performed with perceived workload and condition as independent variables and performance as the dependent variable. (Participants were given an initial store of assets and the performance measure is the difference between the number of assets at the end of test minus the amount in the beginning.) A main effect of condition was found, where participants in the low resources condition performed worse than participants in the high resources condition, $F(1, 68) = 8.5, p < 0.01$ (Mean of low assets $= 32.0$, ste $= 5.0$; mean of high assets $= 23.7$, ste $= 3.2$). An interaction between condition and subjective workload indicated that good performance (fewer assets lost) was measured by those in the high resources condition who perceived the workload as easy, $F(1, 68) = 7.2, p < 0.01$.

Figure 1 shows this interaction. The median workload judgment at time $2 = 61$. Participants who scored below 61 were grouped into the "easy" perceived workload condition and those who scored above 61 were grouped in the "hard" perceived workload condition. Note that there is no difference between the easy/hard perceived workload conditions for those participants who received low assets (post-hoc t-test, $t(9) = 0.66$,

Table 1
Test 1 Change in perceived workload

	Perceived workload	
	Mean time 1 rating	Mean time 2 rating
High resource	62.0 (3.5)	58.9 (3.2)
Low resource	62.7 (3.2)	66.8 (2.9)

Note. Standard error is in parenthesis.

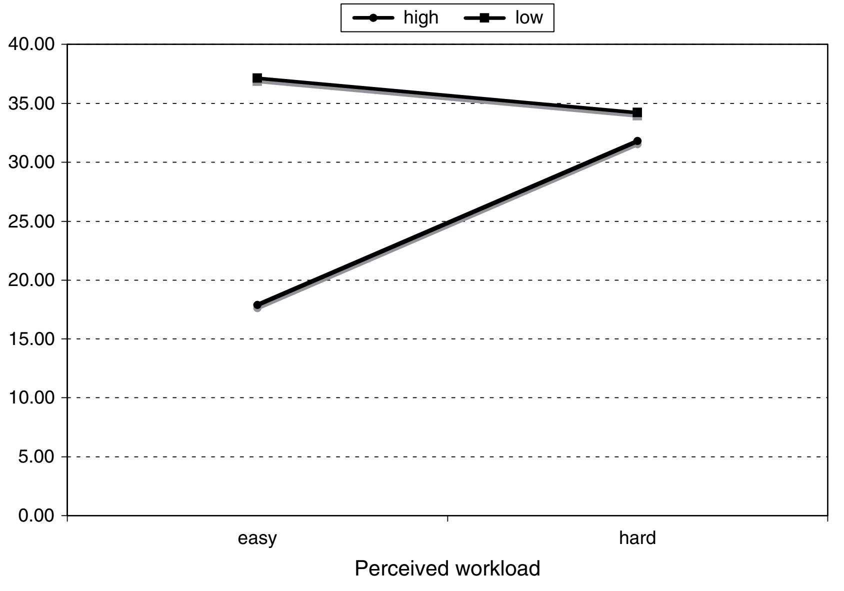

Figure 1. Test 1 results showing high resources condition versus low resources condition, perceived workload and performance.

mean low resources, easy = 37.1, mean low resources, hard = 34.2). Thus, the high resources condition participants were better able to judge the difficulty of the task and the performance matched the judgment.

For Test 2, a repeated measures ANOVA was performed with perceived workload for both time 2 and time 1 as the dependent variable and condition as the independent variable. Participant's perceived workload increased between time 1 and time 2 for all participants, $F(1, 39) = 6.2,\ p < 0.05$ (Table 2).

The performance data for the transfer between Tests 1 and 2 is shown in Figure 2. This figure shows the predicted decrease in performance between Tests 1 and 2 by the experimental group. A repeated measures ANOVA was performed with test as the

Table 2
Test 2 change in perceived workload

Test 1	Perceived workload	
	Mean time 1 rating	Mean time 2 rating
High resource	59.5 (4.1)	65.1 (4.2)
Low resource	61.5 (4.3)	64.5 (4.8)

Note. Standard error is in parenthesis.

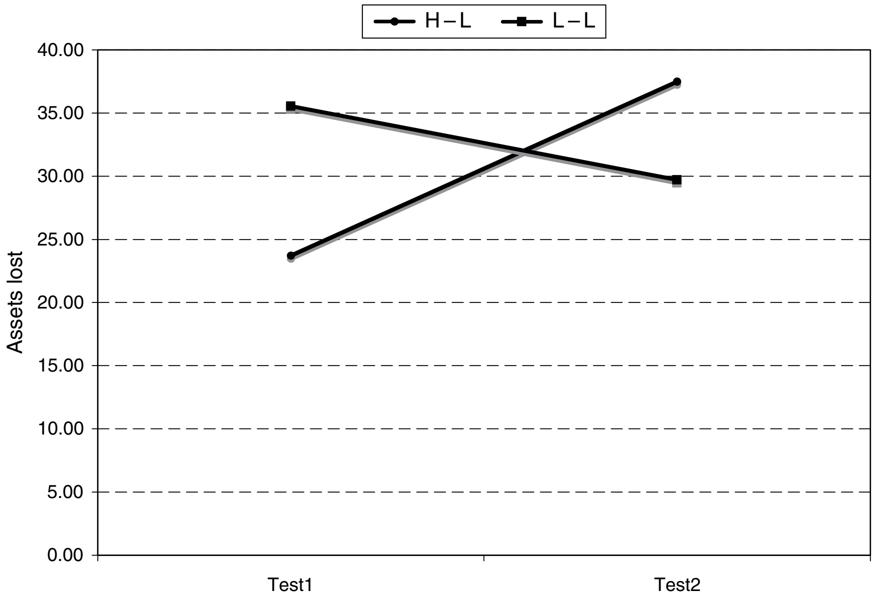

Figure 2. Experimental (H–L) and Control (L–L) conditions performance from Tests 1 to 2.

dependent variable and condition (experimental versus control) as the independent variable. No main effects were found but there was a significant interaction, $F(1,\ 39) = 3.8,\ p < 0.05$. Thus, transfer to lower resources was more difficult for the experimental condition, while the control condition participants improved performance between tests.

Means for performance by median split for workload are shown in Table 3. The median workload score for time $2 = 61$. Participants who scored below 61 were placed in the "easy" group and those who scored above 61 were placed in the "hard" group. A repeated measures ANOVA was performed with Tests 1 and 2 asset loss as the dependent variable and condition as the independent variable. When the analysis is split by workload (easy or hard), no differences are found for participants who rated the workload as hard but a significant difference is found for the interaction between test and condition, $F(1,\ 18) = 5.97,\ p < 0.03$. Thus, participants in the high resource condition for those who viewed the workload as easy for Test 2 performed worse, whereas, participants in the low resource condition who viewed it as easy, did better. Since the trend was for low resource condition participants to improve, their judgments were workload matching the practice/learning effects measured.

Figure 3 shows just the performance data for Test 2 by workload. Notice that this figure is quite different from Figure 1. The results from Test 2 alone indicates that participants in whatever condition they were in produced workload ratings that matched performance. These graphical results are consistent with the analysis depicted in Table 3 because even though experimental participants who rated the workload as easy lost more assets than

Table 3
Test 2 change in test performance by workload

Test 1	Mean Test 1	Mean Test 2
Workload = easy		
High resource	17.70 (4.1)	34.89 (8.9)
Low resource	37.65 (10.1)	27.22 (8.2)
Workload = hard		
High resource	30.41 (8.3)	40.36 (10.6)
Low resource	33.77 (7.6)	31.8 (8.73)

Note. Standard error is in parenthesis.

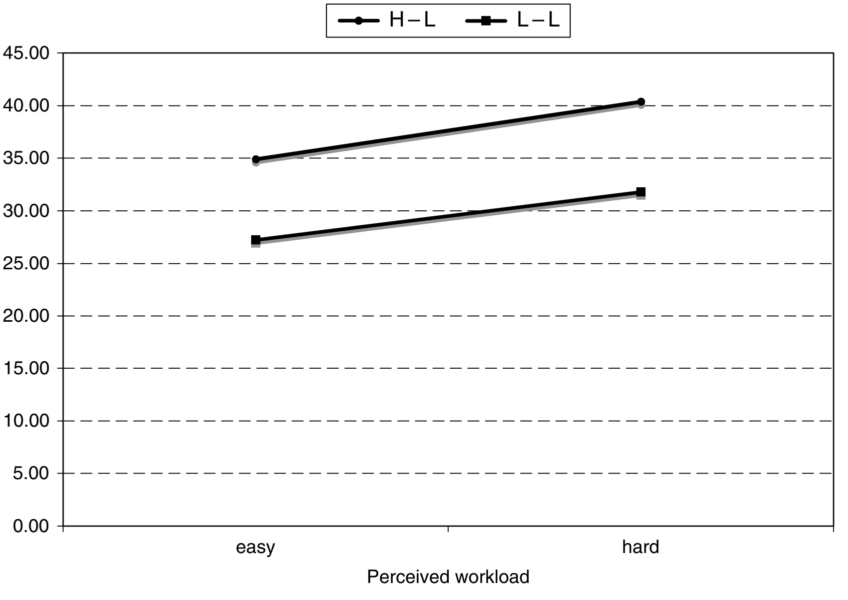

Figure 3. Test 2 results showing high resources condition versus low resources condition, perceived workload and performance.

those in the control who rated the workload as easy, they still outperformed (lost fewer assets) than the experimental participants who rated the workload as hard.

7. Discussion

This experiment was designed to test the relationship between perceived workload and performance in a gaming environment where actual workload could be easily manipulated.

For Test 1, high resource participants' performance matched their workload judgments but low resource participants' performance who rated the workload as easy performed worse than high resource participants. However, a dissociation may not exist for the low resource condition because no significant difference was found in performance for the two levels (easy/hard) of workload judgment.

For Test 2, as expected, in the experimental conditions where participants switched from a high resource situation to a low resource situation, performance suffered. Similar to Test 1, experimental participants who rated the workload as easy performed less well than control participants who rated the workload as easy. However, these experimental participants performed better than their fellow experimental participants who rated the workload as hard. Thus, by Test 2, all participants' workload ratings were matching their performance.

Previous research has documented dissociations between workload and performance (see Yeh & Wickens, 1988, for summary). On the one hand, a quick examination of the data indicates some dissociation amongst the workload ratings and performance measures in both tests. However, for both tests alternative explanations are also possible. For example in Test 1, low resource condition participants who judged the workload to be easy did perform worse but all the ratings for the low resource condition indicated a high workload and the performance was worse than the high resource condition. In the same way, for Test 2, changing between Tests 1 and 2, some high resource condition participants thought the workload was lower (and performed worse); but, focusing on Test 2 alone, participants in both conditions appear to have perceived the workload to performance relationship fairly accurately. One difference between this study and others is that we included training trials that also included the workload assessment (thereby allowing practice on the workload assessment itself). Training and familiarity with test may make participants more accurate in their judgments of perceived workload in the long term and thereby allow for more accurate measurements.

7.1. Future Research

A weakness in this research is the fact that the other transfer (low to high) and control (high to high) conditions were not tested. Research on mental workload has tended to focus on reducing workload for improved task performance; however, in some cases, if few demands are placed, individuals may become complacent and performance may actually suffer (underload). This would be consistent with the Yerkes and Dodson (1908) where a certain level of anxiety is needed for peak performance. Some researchers believe that underload can cause serious problems, while some recent research may indicate that underload in some situations may not be a problem; therefore, future research may need to examine subjective workload in relation to underload (see Billings, 1997; Young & Stanton, 2002).

Another issue is that although changing actual workload (by decreasing the amount of available resources) provides an objective measure of physical change for the participant, other aspects of the situation may change imperceptively (or psychologically) with such a change. Boag, Neal, Loft, and Halford (2006) found that many other aspects of situations other than merely number of aircraft may play a role in subjective workload for air traffic controllers. Further, cognitive abilities may play a role. For example, Gonzalez (2005) measured cognitive abilities using the Raven's test and found that high ability individuals were not necessarily as affect by changes in workload. Thus, further research is needed on imperceptible factors of the situation and individual differences.

Research presented in this paper used a single standard, common measure of subjective workload; however, recent research has focused on physiological measures (Nickel & Nachreiner, 2003; Verwey & Veltman, 1996; Wilson & Russell, 2003). Physiological measure may indicate an actual reaction but may not provide all the additional psychological aspects that a subjective workload measure can give. These studies also focused on a different tasks than the one described in this paper. Verwey and Veltman (1996) found no connection between subjective workload using one measure (SWAT, Subjective Workload Assessment Technique) versus a second measure (RSME, Rating Scale Mental Effort) and Levin et al. (2006) found that three different types of measures all gave inconsistent results.[2] On the other hand, Hjortskov et al. (2004) found that subjective workload measures and electrocardiogram measures were consistent with each other and Murata (2005) found that subjective workload and neocortical activity increased with the difficulty of the task. In another related study, Collet et al. (2003) used physiological measures alone and interpreted increased activity as increased mental workload; however, no other confirming measures were used. Therefore, further research is needed to examine the connection between physiological and subjective measures (see also Dussault, Jouanin, Phippe, & Guezennec, 2005).

Recent research has also indicated that the NASA-TLX may measure different properties compared to a task complexity measure developed by Braarud (2001). His initial studies indicate that complexity may not be completely covered by the NASA-TLX. Therefore, future research should include other subjective measures such as Braarud's complexity measures, along with the physiological measures discussed earlier.

7.2. Interpretations and Implications

One potential interpretation of this research is that perceived workload is easier to judge when there are fewer things to keep track of. High resource participants were given so many resources that they merely had to play the game and assess how they were doing in relationship to the game; however, low resource participants had to both play the game and pay attention to their resources. Over time, the low resource participants demonstrated learning (between Test 1 and Test 2) and this may have allowed them to better judge their workload in Test 2. Although performance was affected in Test 2 for experimental participants, their judgment was still accurate. Experimental participants

[2] SWAT, Subjective Workload Assessment Technique, was developed by Reid & Nygren (1988) and RSME, Rating Scale Mental Effort, was developed by Zijlstra (1993). SWAT is comparable to NASA-XLT used in this research (Verwey & Veltman, 1996).

may have learned to judge their workload based upon the lower workload in Test 1 and this was transferred to Test 2. Thus, an implication of this research is that when learning in a complex gaming environment, initially reducing workload along some dimension may result in better learning than when an individual is immersed in the full environment. It is better not to stress a learner initially because they are better able to detect when they are unable to handle the situation.

An additional implication is that, as a methodology, transfer studies could be a useful way to study workload and further research needs to be performed using this design (in addition to the usual dual task study). In the current study, reduction of workload was studied through transfer but the effects of both overload and underload at the same time.

8. Conclusion

The rise in the use of gaming and simulations in training situations reflects a shift toward providing more realistic learning environments. These gaming environments are characterized by fast information flow and quick, accurate decisions are needed at a moments notice. The advantage of such training/gaming environments is the ability to repeat a task many times and measure the changes of the processing that may affect decision making. Although more research is needed, this study's use of multiple tests (measuring transfer) and workload judgments over time allowed us to see how the relationship between performance and perceived workload changes. From this study, it is clear that a relationship does exist between perceived workload, actual workload and performance, and that training and previous experience may play an important role.

References

Ahlstrom, U. (2005). Subjective workload ratings and eye movement activity measures. Technical Report DOT/FAA/CT-05/32.

Annett, J. (2002). Subjective rating scales: Science or art? *Ergonomics, 45*, 966–987.

Averty, P., Collet, C., Dittmar, A., Athenes, S., & Vernet-Maury, E. (2004). Mental workload in air traffic control: An index constructed from field tests. *Aviation, Space, and Environmental Medicine, 75*, 333–341.

Ball, C., Langholtz, H. J., Auble, J., & Sopchak, B. (1998). Resource-allocation strategies: A verbal protocol analysis. *Organizational Behavior and Human Decision Processes, 76*, 70–88.

Billings, C. E. (1997). Some questions about advanced automation. In M. Moulaoua & J. M. Koonce (Eds), *Human-automation interaction* (pp. 314–320). Mahwah, NJ: Erlbaum.

Boag, C., Neal, A., Loft, S., & Halford, G. S. (2006). An analysis of relational complexity in an air traffic control conflict detection task. *Ergonomics, 49*, 1508–1526.

Braarud, P. O. (2001). Subjective task complexity and subjective workload: Criterion validity for complex team tasks. *International Journal of Cognitive Ergonomics, 5*, 261–273.

Brown, S. W. & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology, Human Perception & Performance, 28*, 600–615.

Collet, C., Petit, C., Champely, S., & Dittmar, A. (2003). Assessing workload through physiological measurements in bus drivers using an automated system during docking. *Human Factors, 45*, 539–548.

Dumblekar, V. (2004). Management simulations: Tests of effectiveness. In *Changing trends in management – Challenges and opportunities*. (pp. 104–115). Delhi: Excel Books.

Dussault, C., Jouanin, J.-C., Phippe, M., & Guezennec, C.-Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation, Space & Environmental Medicine, 76*, 344–350.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds), *Human mental workload. Series, Advances in Psychology*. Volume 52. (pp. 41–62). Amsterdam: North-Holland.

Eggemeier, F. T. & Wilson, G. F. (1991). Performance-based and subjective assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 217–278). London: Taylor & Francis.

Fairclough, S. H., Ashby, M. C., Ross, T., & Parkes, A. M. (1991). Effects of hands free telephone use on driving behavior. In *Proceedings of the ISATA Conference*. Florence, Italy.

Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors, 47*, 92–102.

Gregoriades, A. & Sutcliffe, A. G. (2006). Automated assistance for human factors analysis in complex systems. *Ergonomics, 49*, 1265–1287.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds), *Human mental workload*. Series, Advances in Psychology. Volume 52. (pp. 139–183). Amsterdam: North-Holland.

Hjortskov, N., Rissen, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Sagaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology, 92*, 84–89.

Hopp, P. J., Smith, C. A. P., Clegg, B. A., & Hyeggestad, E. D. (2005). Interruption management: The use of attention-directing tactile cues. *Human Factors, 47*, 1–11.

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. In P. A. Hancock & N. Meshkati (Eds), *Human mental workload. Series, Advances in Psychology*. Volume 52. (pp. 5–39). Amserdam: North-Holland.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Lamoureaux, T. (1999). The influence of aircraft proximity data on the subjective mental workload of controllers in the air traffic control task. *Ergonomics, 42*, 1482–1491.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: Vehicle performance and mental workload implications. *Ergonomics, 47*, 91–104.

Lee, A. Y. (1998). Transfer as a measure of intellectual functioning. In S. Soraci & W. J. McIlvane (Eds), *Perspectives on fundamental processes in intellectual functioning*. (pp. 351–366). London: Ablex.

Levin, S., France, D. J., Hemphill, R., Jones, I., Chen, K. Y., Rickard, D., Makowski, R., & Aronsky, D. (2006). Tracking workload in the emergency department. *Human Factors, 48*, 526–539.

Lysaght, R. J., Hill, S. G., Dick, A. O., Aamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklas, A. L., Bittner, A. C., & Wherry, R. J. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (Report NO. TR-851). Alexandria, VA: US Army Research Institute for Behavioral and Social Sciences.

Matthews, R., Legg, S., & Charlton, S. (2003). The effect of cell phone type on drivers subjective workload during concurrent driving and conversing. *Accident analysis and prevention, 35*, 451–457.

Metzer, U. & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors, 47*, 35–49.

Murata, A. (2005). An attempt to evaluate mental workload using wavelet transform of EEG. *Human Factors, 47*, 498–508.

Nickel, P. & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human Factors, 45*, 575–590.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in assessment of perceived mental workload. *Human Factors, 33*, 17–33.

Parkes, A. M., Fairclough, S. H., & Ashby, M. C. (1993). Carphone use and motorway driving. In E. J. Lovesey (Ed.), *Contemporary Ergonomics*. London: Taylor & Francis.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection indecision making. *Journal of Experimental Psychology: Learning, Memory & Cognition, 14*, 534–552.

Pickup, L., Wilson, J. R., Sharpies, S., Norris, B., Clarke, T., & Young, M. S. (2005). Fundamental examination of mental workload in the rail industry. *Theoretical Issues in Ergonomics Science, 6*, 463–582.

Recarte, M. A. & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied, 9*, 119–137.

Reid, G. B. & Nygren, T. E. (1988). The Subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Neshkati (Eds), *Human mental workload* (pp. 185–218). Amsterdam: North Holland.

Rubio, S., Diza, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review, 53*, 61–86.

Schneider, W. & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Stanton, N. A. & Young, M. S. (2005). Driver behavior with adaptive cruise control. *Ergonomics, 48*, 1294–1313.

Turner, M. L. & Engle, R. W. (1989). Working memory capacity: an individual differences approach. *Journal of Memory and Language*, *28*, 1–28.

Urban, J. M., Weaver, J. L., Bowers, C. A., & Rhodenizer, L. (1996). Effects of workload and structure on team processes and performance: Implications for complex team decision making. *Human Factors, 38*, 300–310.

Verwey, W. B. & Veltman, H. A. (1996). Detecting short periods of elevated workload: A comparison of nine workload assessment techniques. *Journal of Experimental Psychology: Applied, 2*, 270–285.

Vidulich, M. A. & Wickens, C. D. (1986). Causes of dissociation between subjective workload measures and performance. *Applied Ergonomics, 17*, 291–296.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale-Revised. New York: Psychological Corporation.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds), *Varieties of attention* (pp. 63–101). NY: Academic.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). NY: Harper Collins.

Wierwille, W. W. & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors, 35*, 263–281.

Wilson, G. F. & Russell, C. A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45*, 635–643.

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative and Neurological Psychology, 18*, 459–482.

Yeh, Y. & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors, 30*, 111–120.

Young, M. S. & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A new explanation for the effects of mental underload on performance. *Human Factors, 44*, 364–375.

Young, M. S. & Stanton, N. A. (2004). Taking the load off: Investigations of how adaptive cruise control affects mental workload. *Ergonomics, 47*, 1014–1035.

Zhang, Y. & Luximon, A. (2006). Subjective mental workload measures. *Ergonomia IJE& HF, 27*, 199–206.

Zijlstra, F. R. H. (1993). Efficiency in work behavior. Unpublished doctoral dissertation, Technical University, Delft, The Netherlands.

ROLE OF WORKED EXAMPLES TO STIMULATE LEARNING IN A GAME

Chun-Yi Shen
Tamkang University, Taiwan

Harold F. O'Neil
University of Southern California

Abstract

Training by computer games is one of the important activities in many environments. Problem solving may also be effectively improved by computer games. Researchers investigated the efficacy of using worked examples in classroom instruction and provided evidence in the effectiveness of worked examples instruction. The purpose of this study is to examine the effectiveness of worked examples on problem solving in a game-based environment. Seventy-two adults were randomly assigned into the worked example group or the control group. The results showed that the worked example group improved significantly more than the control group in content understanding and problem-solving strategies.

1. Introduction

1.1. Background

Computer games have been used for training in many different environments, such as academic (Adams, 1998), business (Faria, 1998; Lane, 1995), military (Chambers, Sherlock, & Kucik, 2002), and medical (Ruben, 1999). Researchers pointed out that games are widely accepted as a powerful alternative to traditional ways of teaching and learning, with the merits of facilitating learning by doing (Mayer, Mouton, & Prothero, 2002; Schank, 1997). In addition, problem solving may be effectively improved by computer games (Mayer, 2002). According to O'Neil and Fisher (2002), the effects of computer games can be generally divided into four

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

categories: (a) promotion of motivation, (b) enhancement of thinking skills and metacognition, (c) improvement of knowledge, and (d) improvement of attitudes.

According to Gagné (1977), the purpose of educational programs is to teach students to solve problems such as mathematical and physical problems, health problems, social problems, and problems of personal adjustment. Some psychologists concluded that most human learning engages problem-solving activities (Anderson, 1993). Problem solving is also one of the most significant competencies whether in job settings or in schools and, as a result, teaching and assessing problem solving becomes one of the most significant educational objectives (Mayer, 2002). Mayer (2002) stated teaching problem-solving transfer has become one of the most critical educational objectives. As O'Neil (1999) pointed out, problem solving is a critical competency requirement of college students and employees.

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has conducted studies on problem solving (Baker & O'Neil, 2002; Baker & Mayer, 1999; O'Neil & Herl, 1998; O'Neil, Baker, & Fisher, 2002). CRESST adapted the problem-solving models of Mayer and Wittrock (1996). The CRESST model of problem solving consists of three components: (a) content understanding, (b) problem-solving strategies, and (c) self-regulation (Baker & Mayer, 1999; O'Neil, 1999).

The two major alternative techniques to teach problem solving are the use of worked examples and goal-free problems (Sweller, 1990; Ward & Sweller, 1990). Sweller (1989) defined worked examples as a procedure that focuses on problem states and associated operators (i.e., solution steps), enabling students to induce generalized solutions or schemas. According to Sweller (1993), a goal-free problem is a problem without specific goals that encourages learners to find the value of as many variables as possible. In the last 15 years, many investigators paid a considerable amount of attention to worked examples and concluded that worked examples instruction is superior to the conventional problem-solving instruction, especially in the field of mathematics, computer programming, and physics (Carroll, 1994; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Renkl, Atkinson, Maier, & Staley, 2002; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). A number of researchers investigated the efficacy of using worked examples in classroom instruction and provided evidence in favor of worked examples instruction rather than problem-solving instruction (Cooper & Sweller, 1987; Cooper, Tindall-Ford, Chandler, & Sweller, 2001; Ginns, Chandler, & Sweller, 2003; Paas, Renkl, & Sweller, 2003; Pawley, Ayres, Cooper, & Sweller, 2005). However, worked examples have not been applied to improve learning in game-based environments.

1.2. Purpose of the Study

The main purpose of this study was to examine the effectiveness of worked examples on problem solving in a game-based environment. The researcher used the problem-solving assessment model developed by the CRESST to measure the three components of problem solving, which are content understanding, problem-solving strategy, and self-regulation (Baker & Mayer, 1999; Herl, O'Neil, Chung, & Schacter, 1999; Mayer, 2002). The data and conclusion presented in this study was based on the senior author's doctoral dissertation.

2. Worked Examples

2.1. Introduction

A number of researchers investigated the efficacy of using worked examples in classroom instruction and provided evidence of the effectiveness of worked examples instruction (Carroll, 1994; Cooper & Sweller, 1987; Ward & Sweller, 1990; Zhu & Simon, 1987). As Atkinson, Derry, Renkl, and Wortham (2000) have defined, worked examples are instructional devices that provide an expert's solution for a learner to study. According to Atkinson et al. (2000), as instructional devices, typical worked examples include a problem statement and a step-by-step procedure for problem solving. Both of these elements are meant to show how similar problems might be solved. In addition, worked examples provide an expert's problem-solving model for the learner to study and follow (Atkinson et al., 2000).

According to Sweller (1990), a worked example is a procedure that focuses on problem states and associated operators (i.e., solution steps), enabling students to induce generalized solutions or schemas. The definition made by Sweller (1990) was used in this study.

In last 15 years, many researchers (e.g., Carroll, 1994; Chi et al., 1989; Renkl et al., 2002; Sweller, Chandler, Tierney, & Cooper, 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990) concluded that worked examples instruction is superior to the conventional problem-solving instruction, especially in the field of mathematics, computer programming, and physics.

Sweller and his colleagues (e.g., Mawer & Sweller, 1982; Owen & Sweller, 1985; Sweller, Mawer, & Howe, 1982; Sweller, Mawer, & Ward, 1983) conducted many studies to examine how students learn schemas and patterns that facilitate problem solving via conventional practice-oriented instruction. Sweller's study found empirical evidence showing that traditional, practice-oriented problem solving was not an ideal method for improving problem solving when compared to instruction that paired practice with worked examples (Cooper & Sweller, 1987; Sweller & Cooper, 1985).

Van Gerven, Paas, Van Merriënboer, and Schmidt (2002) suggested that worked examples could promote acquisition of complex cognitive skills for adults by reducing their cognitive load and irrelevant information. As Zhu and Simon (1987) pointed out, worked examples can be an appropriate and acceptable instructional substitute when compared to conventional classroom activity. In their study, the students completed a 3-year course in 2 years by using worked examples as instructional material (Zhu & Simon, 1987). In addition, in Carroll's (1994) study, students were asked to learn how to make mathematical equations with worked examples. The results of this study indicated that worked examples could promote students' confidence in mathematics and illustrated mathematical principles and classes of problem situations.

2.2. Theories of Worked Examples

In many studies of worked examples (e.g., Sweller, 1990), there are three major theories applied, which are (a) scaffolding, (b) schema theory, and (c) cognitive load theory. These three theories are described as the following:

2.2.1. Scaffolding

The term "scaffolding" was initially introduced by Vygotsky (1989) as tutoring or other assistance provided in a learning setting to assist students with attaining levels of understanding impossible for them to achieve without assistance (Brush & Saye, 2001). Scaffolds are tools, strategies, and guides that support students in attaining a higher level of understanding (Hannafin, Land, Oliver, & Reigeluth, 1999; Vygotsky, 1989).

According to Vygotsky (1989), scaffolding is helping learners develop their own evolving knowledge through interactions with other people. Vygotsky (1989) viewed scaffolds as personal aid provided by a teacher or peer to help with the learning process. Vygotsky believed that learners construct their knowledge by interacting with other members of society and cannot be understood apart from the cultural settings. This theory suggested that teachers need to do more than just arrange the environment for students in order to learn on their own. This meant that teachers should assist or guide students' learning rather than transmit knowledge to them. Assisted learning requires scaffolding, which is the support for learning and problem solving.

According to Vygotsky (1989), there is a zone between what learners can do by themselves and with assistance. The zone of proximal development (ZPD) is a phase at which a learner can master a task if given appropriate help or support (Woolfolk, 2001). Wertsch and Resnick (1991) defined the zone of proximal development as an area where the learner cannot solve problems alone but can be successful under guidance. This is the area where instruction can succeed because real learning is possible.

Recently, the concept of scaffolding has been broadened to include a multitude of different tools and resources that can be used by students to assist them with instructional activities (Brush & Saye, 2001). According to Rosenshine and Meister (1992), assisted learning includes: (a) adapting materials or problems to students' current levels, (b) demonstrating skills or thought process, (c) walking students through the steps of a complex problem, (d) doing part of the problem, (e) giving detailed feedback and allowing revisions, and (f) asking questions that focus on students' attention. According to Rosenshine and Meister (1992), worked examples could be one type of scaffolding. The effective worked examples instruction should provide appropriate help and involvement in the zone of proximal development (Rosenshine & Meister, 1992).

2.2.2. Schema Theory

The notion of schema was developed by psychologists in order to deal with the fact that much of our knowledge seems integrated (Woolfolk, 2001). Schema is a cognitive construct that categorizes the elements of information by the manner with which they will be dealt (Sweller, 1990). Furthermore, Sweller defined schema as a cognitive construct that allows problem solvers to recognize problems and problem states as belonging to a particular category that requires particular moves for solution (Sweller, 1990). According to schema theory, knowledge is stored in long-term memory in the form of schemas (Sweller, Van Merriënboer, & Paas, 1998). Schema construction occurs when learners determine the elements that need to be merged into a schema (Ginns et al., 2003).

Schema automation is a critical factor in schema construction. Schema automation occurs when learners practice sufficiently (Sweller et al., 1998). Many studies have confirmed the importance of schemas in algebraic problem solving (e.g., Larkin, McDermott, Simon, & Simon, 1980; Sweller, 1990), algebraic word problem solving (Low & Over, 1990, 1992), and geometry problem solving (Koedinger & Anderson, 1990). Van Merriënboer, Clark, and de Croock (2002) have suggested that cognitive schemata enable problem solvers to solve a new problem by serving as an analogy.

In addition, studying instructional material, such as worked examples, promotes the incorporation of procedures and elements into schemas held in long-term memory. That means learners can retrieve the schema as a single element so that they can process it in working memory without the assistance of instruction (Ginns et al., 2003). In a series of worked example studies, for example, Cooper and Sweller (1987), Sweller and Cooper (1985), Cooper et al. (2001), Ginns et al. (2003), Paas et al. (2003), Pawley et al. (2005), the researchers concluded that worked examples can enhance learning by schema construction, reducing extraneous cognitive load, and focusing attention properly.

2.2.3. Cognitive Load Theory

Sweller and Cooper (1985) and Mwangi and Sweller (1998) conducted a series of studies examining the effects of worked example instruction on mathematical learning. According to the results of the studies, Sweller (1990, 1993, 1994, 2004), Sweller and Chandler (1994), and Paas et al. (2003) developed the cognitive load theory to explain the limitation of cognitive resources during problem solving.

Sweller (1989) indicated that people are only able to work with about two or three items of information at a time when required to process rather than merely hold information. This limitation has been considered as one of the most important factors in instructional design (Carlson, Chandler, & Sweller, 2003; Chandler & Sweller, 1991, 1992, 1996; Jeung, Chandler, & Sweller, 1997; Mayer, 2001; Pass & Van Merriënboer, 1994; Sweller et al., 1990, 1998; Tindall-Ford, Chandler, & Sweller, 1997).

Cognitive load theory (Paas et al., 2003; Sweller 1989, 1994) assumes that all learning uses a very limited working memory and an unlimited long-term memory. This theory suggests that the limitation of working memory is the most critical factor when students are studying new instructional material. According to this theory, cognitive load can be divided into three categories: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Sweller et al., 1998; Carlson et al., 2003; Gerjets & Scheiter, 2003; Renkl & Atkinson, 2003).

Intrinsic cognitive load is determined by the degree of interaction among the elements in working memory during learning (Sweller, 1994; Sweller & Chandler, 1994). If there are too many interactive elements presented to the learners at the same time, the cognitive load might exceed the limits of working memory and the effectiveness of learning would be restricted (Carlson et al., 2003). On the other hand, extraneous cognitive load is determined by the organization of the instructional material. Instructional materials can be presented in different ways, such as diagrams, text, and worked examples, and different ways cause different amounts of extraneous cognitive load (Carlson et al., 2003). For example, the integration of diagrams and text enhance learning compared to diagrams-only or text-only instructional materials (Mayer & Moreno, 2003; Mayer & Sims, 1994; Mayer, Moutone, & Prothero, 2002).

Instructional design needs to reduce extraneous cognitive load or ineffective cognitive load by using the appropriate instructional methods and formats (Sweller, 1990; Sweller et al., 1990). In addition, according to Wulf and Shea (2002), learners may perform worse or even stop learning under the conditions of excessively low cognitive load or high cognitive load. If the instructional materials are too easy for learners, they may benefit from practice that increases the cognitive load and challenges them. On the other hand, if the instructional materials are too difficult, students may benefit from practice that reduces their cognitive load to an appropriate level.

According to Gerjets and Scheiter (2003), germane or effective cognitive load may enhance schema construction by implementing higher level cognitive processes due to beneficial cognitive processes, for example, abstractions, comparisons, elaboration, and inferences that are encouraged by the instructional presentation. As pointed out by Paas, Renkl, and Sweller (2003), germane cognitive load is the load induced by learners' efforts related to the processes that contribute to the construction and automation of schemas. According to Renkl and Atkinson (2003), in the case of learning from worked examples, self-explanation would be considered germane load while a learner's effort in gaining an understanding of a solution rationale.

According to cognitive load theory, a large number of elements cannot be manipulated in working memory without assistance. Cognitive load can be reduced effectively after schema automation (Sweller et al., 1998). Ward and Sweller (1990) suggested that worked examples could facilitate schema construction and automation.

Cognitive load study has shown that learning from worked-out examples, in comparison to problem solving, is very effective during the initial stages of cognitive skill acquisition. In later stages however, solving problems is superior (Renkl & Atkinson, 2003). In Paas and Van Merriënboer's (1994) study, students who studied worked examples gained most from high-variability examples, invested less time and mental effort in practice, and attained better and less effort-demanding transfer performance than students who first attempted to solve conventional problems and then studied work examples.

According to Cooper et al. (2001), students engaged in studying worked examples that emphasized understanding and remembering procedures and concepts were compared with students who were engaged in imagining worked examples that emphasized imagining procedures and concepts. It was hypothesized that students who held prerequisite schemas would find imagining to have a beneficial effect on learning, compared with students studying the material. Whereas students who were less knowledgeable would find imagining to have a negative effect on learning when compared to studying worked examples. Experimental results were in accord with our hypotheses. It was concluded that, under specific circumstances, encouraging students to imagine procedures and concepts could substantially facilitate learning. In addition, the study conducted by Ginns et al. (2003) indicated that worked example was superior to imagining when subjects had low prior knowledge combined with complex tasks. But the superiority was reserved when subjects had high prior knowledge or the tasks were less complex.

Although there are many studies showing that worked examples are superior to traditional problem solving instruction, not all worked examples are effective (Sweller, 1990). Instruction needs to be formatted to make cognitive resources focus on facilitating schema acquisition rather than directed to other irrelevant activities (Sweller, 1990). Many researchers (e.g., Sweller et al., 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990) suggested that the efficacy of worked examples as an instructional technique was limited by the quantity of information processing required. This meant that worked examples were only effective under certain cognitive conditions. If the worked examples involved a heavy cognitive load, cognitive resources directed to problem states and the related steps to the resolution might be reduced to the detriment of schema acquisition (Lim & Moore, 2002). Thus design of worked examples became critical.

2.3. Design of Worked Examples

The design or structure of worked examples plays an important role in the effectiveness of worked examples (Catrambone, 1994; Mwangi & Sweller, 1998; Ward & Sweller, 1990; Zhu & Simon, 1987). In many cases, worked examples consist of aids, such as diagrams, verbal instruction, and paired problems. Although the worked examples used by many researchers were not similar, they shared the same fundamental purpose: to demonstrate a pattern or principle (Atkinson et al., 2000). These principles of designing worked examples are (a) before vs. after, (b) complete vs. incomplete, (c) backward fading vs. forward fading, (d) text vs. diagrams, (e) visual vs. verbal (f) steps vs. subgoals. Each principle is briefly discussed, and its implication for this study is noted as the following:

2.3.1. Before vs. After

Laboratory studies indicated that when presented with traditional practice exercises, students tended to use typical novice strategies, such as trial and error, while students presented with worked examples before solving often used more efficient problem-solving strategies and focused on structural aspects of problems (Atkinson et al., 2000). However, Stark, Gruber, Renkl, and Mandl (2000) concluded that presenting problems first followed by similar worked examples (problem-example) is significantly more effective than exposing learners to either problems only or worked examples only instructions.

In the methodology of this study, the participants were asked to study the worked examples before the problems. During the problem solving, the participants could review the worked examples at any time to reduce their cognitive load.

2.3.2. Complete vs. Incomplete

There are empirical studies (e.g., Renkl et al., 2002) showing that in some conditions complete worked example instructions are not as effective as incomplete worked example instructions. Many researchers (e.g., Pass, 1992; Van Merriënboer, 1990) argued that incomplete worked examples effectively support the acquisition of cognitive skills.

Stark, Mandl, Gruber, and Renkl (1999) found that studying incomplete examples fostered the quality of self-explanations and the near and medium transfer of learned solution methods significantly more than studying complete worked examples. Stark et al. (1999) defined near transfer problems as problems that are similar in structure (same solution rationale) to the worked examples but contain different surface features (cover story, objects, numbers). Medium transfer problems are problems with a different structure (a modified solution procedure), but similar surface features. Far transfer problems differed with respect to both structure and surface features (Stark et al., 1999).

In general, early studies on worked examples used complete examples to establish their effectiveness. Then the studies on incomplete worked examples were conducted. For this study, the author used complete worked examples as this is the first study of its kind and because the purpose of this study is to investigate the effectiveness of worked examples by comparing worked example groups with nonworked example groups in a game-based environment.

2.3.3. Backward Fading vs. Forward Fading

Renkl et al. (2002) introduced a new feature for worked example instruction called fading, which integrates worked examples and problem solving and builds a bridge between example study in early phases of cognitive skill acquisition and problem solving in later stages. They conducted a successive integration of problem-solving elements into worked examples until the learners solved problems on their own (i.e. complete worked examples only to increasingly more incomplete examples to problem only). They found that the fading procedure can foster the performance of near and medium transfer problems only, and it is more favorable to fade out the solution steps of the worked examples in a backward manner (leaving out the last solution step first) when compared with a forward manner (omitting the first solution step first) (Renkl et al., 2002).

Because the previous studies could not provide enough evidence of the effect of fading on far transfer problems, Atkinson, Renkl, and Merrill (2003) conducted another study combining fading with self-explanation prompts designed to encourage learners to identify the underlying principle demonstrated in each worked example solution step. This study successfully demonstrated using backward fading procedure combined with self-explanation prompting significantly fosters not only near and medium transfer learning but also far transfer learning (Atkinson et al., 2003). According to Renkl, Atkinson, and Grobe (2004), students would learn most about those principles that were faded, which suggested that specific self-explanation activities are triggered by faded steps.

In this dissertation, the authors did not use fading in the worked examples because the purpose of this study was to investigate the effectiveness of worked examples by comparing worked example groups with nonworked example groups in a game-based environment.

2.3.4. Text vs. Diagrams

According to Sweller et al. (1998), the purpose of worked example instruction is to reduce extraneous cognitive load. However, studies provide evidence for the split-attention effect that might reduce the effectiveness of worked examples (Sweller et al., 1990; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). The split-attention effect occurs when learners are required to divide their attention and mentally integrate multiple sources of information (Mousavi, Low, & Sweller, 1995). According to Mwangi and Sweller (1998), Tarmizi and Sweller (1988), and Ward and Sweller (1990), the participants in the split-attention group who had to attend to the information from separate diagrams and text showed decreased effectiveness of learning in a series of worked example studies. They also found that restructuring the worked examples by integrating the diagrams and textual explanations of problems facilitates learning. Sweller et al. (1998) concluded that the split-attention effect occurs when learners' attention is distracted from separate text and diagrams. Therefore, the text and diagrams need to be integrated appropriately.

In summary, learners benefited most from the instructional materials with integration of diagrams and text compared to diagrams-only or text-only instructional materials. In this study, the author integrated the text and diagrams of the worked examples in order to avoid the split-attention effect.

2.3.5. Visual vs. Verbal

The integration of diagram and text can enhance learning. Integration of visual and verbal can also promote the effectiveness of worked example instruction (Tarmizi & Sweller, 1988; Mayer & Moreno, 2003; Moreno & Mayer, 2002; Mousavi et al., 1995; Ward & Sweller, 1990). For example, Mousavi et al. (1995) conducted a series of worked example studies to compare the effectiveness of using worked examples to teach eighth-grade geometry with different formats: (a) visual–visual, where the integration of diagrams and text were presented, (b) visual–auditory, where diagram was presented with aural associated statements, and (c) simultaneous, where diagrams was presented with both text and aural statements. The results show that visual–auditory and simultaneous methods are superior to visual–visual method.

Mayer and his colleagues (Mayer, Moreno, & Boire, 1999) consistently found that integration of dual-presentation, which was visual–verbal, could enhance learning in a series of multimedia studies. In addition, in Atkinson, Mayer, and Merrill's (2005) study, college students were asked to learn how to solve proportional reasoning word problems using a computer-based multimedia program. Some of their studies involved a set of worked examples with proportional reasoning narrated by an animated agent with a human voice, and some of their studies involved the same set of worked examples but narrated by an agent with a machine synthesized voice. The results showed that the students with a human voice performed better on near and far transfer test than those with a machine synthesized voice.

However, Jeung et al. (1997) found that using visual–visual worked examples is superior to using visual–verbal worked examples in complex conditions. They conducted a series of studies to compare visual–verbal and visual–visual worked example instructions in complex and simple problem settings. They concluded that visual–verbal instruction appears to be superior to visual–visual instruction only in simple problem tasks (Jeung et al., 1997).

Based on previous studies, further study is needed to examine the effectiveness of the integration of diagrams, text, and audition on simple and complex worked examples. Since this study is the first study of its kind and because the purpose of this study is to investigate the effectiveness of worked examples by comparing worked example groups with nonworked example groups in a game-based task in a complex task. The authors chose the visual–visual worked examples as instructional intervention in this study.

2.3.6. Steps vs. Subgoals

Catrambone (1994) and Catrambone and Holyoak (1990) suggested that subgoals appear to enhance learning by modifying old steps rather than applying it without adaptation. That might come from (a) emphasizing meaningful chunks of problems' solutions, (b) adding labels to the solutions, (c) inducing the worked examples' underlying goal structures, and (d) leading learners to discover meaningful generalizations (Catrambone & Holyoak, 1990; Atkinson et al., 2000). Catrambone (1994) indicated that learners are encouraged to distinguish the function of the subgoals and then explain why the steps are grouped together.

In addition, these cognitive activities help learners induct the principles and schemas used in the worked examples (Atkinson et al., 2000). The researchers (Catrambone, 1994; Catrambone & Holyoak, 1990) found that the performances of the participants receiving the subgoals were significantly better than the participants without subgoals. According to Sweller (1990), the subgoals or general strategies in worked examples can facilitate learners' transfer of knowledge while the specific steps can enhance learners' retention.

According to Gerjets, Scheiter, and Catrambone (2004), studying the worked examples that focus on steps, problem-type schemas, structural task features, and category-specific solution procedures might be cognitively demanding because it requires learners to simultaneously hold a substantial amount of information in working memory. Therefore, the researchers suggested to break down the worked examples with complex steps into smaller subgoals with meaningful solution elements that can be conveyed separately. Based on their results, the learners who had the subgoals performed better than those who had the steps. In addition, according to Van Gog, Paas, and Van Merriënboer (2004), worked examples should focus on the principle information ("why") and strategic information ("how") that experts use when solving problems using subgoals. The results of their study confirmed that process-oriented information of worked examples could enhance transfer performance, especially for complex cognitive skills with multiple possible solution paths.

In this study, the authors chose the simple steps in the worked examples because the purpose of this study is to investigate the effectiveness of worked examples by comparing worked example group with non-worked example group in a game-based environment.

2.4. Methodology and Results

The present study consisted of a pilot and a main study. The purpose of the pilot study was to investigate the feasibility of procedures and measurements, the format of worked examples and the assessment tools for problem solving. The purpose of the main study was to evaluate the effectiveness of worked examples in a game-based problem-solving task. The design of the main study was the randomized pretest posttest design with control group, which was a true-experiment design. The research question in this study is "Will participants in the worked example group increase their problem solving in a game-based task (i.e. SafeCracker) after studying worked examples compared to the control group."

In the main study, 72 undergraduate and/or graduate students were randomly assigned into two groups, which were the worked example and the control group. The age of participants ranged from 18 to 38. The average age of the participants was about 26 years old. About half of them seldom played video games and the other half played at least 1 h per week. Each group was asked to play the computer puzzle game, that is, SafeCracker. SafeCracker is a computer puzzle game that requires the players to find the clues, apply the tools and knowledge, and solve the puzzles in order to open the safes in a mansion. For example, in order to open the Brown safe (Pascal Triangle puzzle) in Room 2 (Small Showroom), the participants need to provide the correct values for the elements of Pascal's triangle; each value is determined by adding together the two numbers immediately above. Or use the clue found in the brown book (Mathematical Puzzles & Numerical Figures) on the desk in Room 6. In the procedure, the participants were asked to fill out self-regulation questionnaires first. Then, they were asked to complete the knowledge maps and problem-solving strategy questions after the first game playing session as the pretests. Next, the worked example group studied the worked examples, and the control group did not. Then, they were asked to complete the knowledge maps and problem-solving strategy questions after the second game playing session as the posttests. Finally, the debriefing completed the main study.

The results obtained in this study provided the evidence that worked examples could enhance problem solving in a game-based environment. First, the participants who received worked examples improved significantly more than those who did not receive worked examples on the knowledge map. The average improvement of knowledge map score of the worked example group ($M = 2.21$ out of 81, or 2.73%) was significantly ($p < 0.01$) more than that of the control group ($M = 0.62$ out of 81, or 0.77%).

Second, the participants who received worked examples improved more than those who did not on the problem-solving retention question. However, a significant ($p < 0.05$) difference between the two groups was found in the pretest. Thus an analysis of covariance (ANCOVA) was run while assigning the mean score of pretest as the covariate. The adjusted mean of the posttest of retention score of the worked example group ($M = 4.18$ out of 28, or 14.93%) was significantly ($p < 0.01$) greater than that of the control group ($M = 3.54$ out of 28, or 12.64%).

Third, the worked example group improved more than the control group on the problem-solving transfer question. However, a significant ($p < 0.05$) difference between the two groups was also found in pretest. Thus an ANCOVA was run while assigning the mean score of pretest as the covariate. The adjusted mean of the posttest of transfer score of the worked example group ($M = 2.37$ out of 22, or 10.77%) was significantly ($p < 0.05$) more than that of the control group ($M = 1.97$ out of 22, or 8.95%).

In addition, as predicted, no significant difference between the two groups was found in the scores of trait self-regulation questionnaire. Significant ($p < 0.05$) correlations were found only between the scores of trait self-regulation questionnaire and knowledge map scores for the total sample. For the total sample, the participants with higher planning performed better in knowledge map pretest and posttest; the participants with higher self-monitoring got higher scores on knowledge map posttest and improvement; the participants with higher effort performed better on knowledge map improvement; and the participants with higher self-efficacy got higher scores on knowledge map pretest, posttest, and improvement.

3. Discussion and Implications

The purpose of this study was to examine the effectiveness of worked examples in a game-based problem-solving task. The CRESST model of problem solving consists of three components: (a) content understanding, (b) problem-solving strategies, and (c) self-regulation (Baker & Mayer, 1999; O'Neil, 1999). Therefore, this study examined the effectiveness of worked examples on these three components.

Hypothesis 1: Worked examples instruction will produce a significant improvement in content understanding compared to control group. This hypothesis was supported.

The results were consistent with several studies. For example, Ward and Sweller (1990) suggested that worked examples could facilitate schema construction and automation. In Van Gog, Paas, and Van Merriënboer's (2004) study, the participants who received the process-oriented worked examples performed better on content understanding test and transfer test than those who did not. Kalyuga, Chandler, Tuovinen, and Sweller (2001) conducted a study examining the interaction between levels of learners' prior knowledge and levels of instructional guidance. They concluded that (a) the participants with low prior knowledge benefited most from the worked examples condition, and (b) the effectiveness of worked examples or problem solving depends heavily on levels of learners' prior knowledge. In Kalyuga, Chandler, and Sweller's (2001) study, they investigated interaction between levels of learner knowledge, levels of tasks' complexity, and levels of instructional guidance. The results indicated that there was no significant difference between worked example group and control group in simple tasks. As expected, they found inexperienced subjects clearly benefited most from the worked examples procedure in complex tasks.

In this study, the participants had no experience in playing SafeCracker, so their level of prior knowledge was low. In addition, the problems in this study required several high-level cognitive skills and had more than seven steps, so the problems were considered complex. As expected, the participants who received worked examples performed significantly better than those who did not on the knowledge map test. The results supported the previous studies on worked examples.

In addition, in Ginns et al. (2003) study, the average percentage of correct answers was 77.39% among 20 inexperienced college students who were given worked examples to learn difficult HTML coding. In the study conducted by Kalyuga et al. (2001), the average percentage of correct answers on a programming performance test was 54.17%, of the worked example group for 24 novice trade apprentices learning Programmable logic controller (PLC) programming. Further, the study conducted by Tuovinen and Sweller (1999) investigated the effectiveness of worked examples in learning complicate computer skills. The results indicated that the average percentage of correct answers was 30.25%. Thus the average percentage of correct answers in the previous studies by Sweller and his colleagues was 58.79%.

In this study, the achievement test could be the alternative problem-solving test which was the task of opening the safes. Therefore, the score of the alternative problem-solving test in this study could be the percentage of the opened safes. For the three safes that had worked examples (the Liberty Safe in Room 5, the Diapicture Safe in Room 6, and the Big Switch Safe in Room 27), there was a significant difference in the percentage of opened safes between the two groups. The worked example group ($M = 84.26\%$) performed significantly better than the control group ($M = 34.26\%$). Furthermore, the average percentage of the five safes opened by the worked example group in the second game-playing was 57.22%, which was close to the average percentage of correct answers in previous studies (e.g., Ginns et al., 2003; Kalyuga et al., 2001; Tuovinen & Sweller, 1999). In addition, according to the results of t-test, there was a significant difference in the improvement of the percentage of the opened safes between the two groups. The worked example group improved significantly more than the control group (20.19% vs. −6.85%). As expected, there was no significant difference in the percentage of the opened safes in the pretest (first round game-playing) between the two groups.

However, the result of this study was not consistent with Chen's (2005) study. The purpose of Chen's (2005) study was to examine the effectiveness of games. Thus the knowledge map pretest was before the first round game playing session and the posttest was after the second game playing session. In Chen's (2005) study, the difference between the knowledge map pretest and posttest was statistically significant in favor of the performance after the game playing session. There was only one group in Chen's (2005) study. The procedure used in Chen's (2005) study was similar to the procedure of the control group in this study. However, the purpose of this study is to examine the effectiveness of worked examples. Thus the knowledge map pretest was after the first round game playing session and the posttest was after the second game playing session. Due to the different procedures from Chen's (2005) study, the results in this study showed there was no significant difference between the knowledge map pretest and posttest of the control group.

Although the results from this study confirmed the superiority for these worked example treatment in a game-based environment, the improvement of content understanding was small (from 4.08% on the pretest to 6.81% on the posttest). The rational of this small improvement is discussed in the next section.

Hypothesis 2: Worked examples instruction will produce a significant improvement in problem-solving strategies compared to the control group. This hypothesis was confirmed.

The results in this study were consistent with the results of the following studies. For example, Van Gerven, Paas, and Schmidt (2000) suggested that worked examples could promote acquisition of complex cognitive skills for adults by reducing their cognitive load. According to Cary and Carlson (1999), worked examples can facilitate problem solving in a complex retention arithmetic task. In addition, the results of the study conducted by Paas (1992) concluded that worked example is an efficient knowledge base for solving transfer problems. In Van Gog, Paas, and Van Merriënboer's (2004) study, the participants who received the process-oriented worked examples performed better than those who did not on a transfer test.

In addition, the results of problem-solving strategy in this study were consistent with Chen's (2005) study. In her study, there were significant differences between the scores of pretests and posttests of problem-solving strategy questions of retention and transfer. According to the results of paired-sample t-test, there were significant differences between scores of pretests and posttests of problem-solving strategy questions of retention and transfer of the control group in this study. Unlike the knowledge map tests, the procedures of problem-solving measures in this study were the same as the procedure in Chen's (2005) study. The pretests of problem-solving strategy question of retention and transfer were presented after the first round of game playing session, and the posttests were presented after the second round of game playing session in both studies. In this study, as expected, the worked example group performed significantly better than the control group in problem-solving questions of retention and transfer.

However, the improvement of the worked example group was small. For retention questions, the improvement was from 2.17 on the pretest to 4.18 on the posttest; and the improvement was from 1.75 on the pretest to 2.37 on the posttest for transfer question. The reasons could be

1. There were only three worked examples (60%) out of a total of five safes in the second round game playing session.
2. The problems in this game were complicated and there was more than one solution for each safe. The worked examples in this study provided only one of the solutions for each of the three safes.
3. The worked example instruction in this study lacked practice problems (Renkl et al., 2002), fading procedure (Atkinson et al., 2003; Renkl et al., 2004), and self-explanations prompting (Renkl & Atkinson, 2003).

Therefore, in the subsequent study, the worked example instruction needs to be modified to improve problem solving in a game-based environment. In addition, the interaction among levels of instruction, levels of complexity, levels of prior knowledge, fading procedure, and self-explanations prompting also needs to be investigated.

Hypothesis 3: There will be no significant difference in self-regulation between worked example groups and control groups. However, higher self-regulation will be associated with higher content understanding and problem-solving strategies. This study confirmed that self-regulation was associated with content understanding, but not associated with problem-solving strategies.

As expected, the results confirmed that there was no significant difference in self-regulation between worked example groups and control groups. For the total sample, the participants with higher planning performed better in knowledge map pretest and posttest; the participants with higher self-monitoring received higher scores on knowledge map posttest and improvement; the participants with higher effort performed better on knowledge map improvement; and the participants with higher self-efficacy had higher scores on knowledge map pretest, posttest, and improvement. Thus, the scores of self-regulation were associated only with content understanding, but not related to problem-solving strategies.

The results were consistent with Chen's (2005) study. In Chen's (2005) study, significant correlations were found between self-monitoring and the learning of knowledge map, self-efficacy and knowledge map pretest and posttest. However, the results were not consistent with the following studies. According to Howard, McGee, Hong, and Shia (2000), self-regulation could be a significant predictor of both content understanding and problem solving. In Phye's (1998) study, the results indicated that self-regulation could significantly influence both domain-specific and domain-independent problem-solving performance. In addition, according to Theodorou and Bonnie (2001), self-regulation could predict transfer of problem solving. Thus the relationships among self-regulation, problem solving, and worked examples need to be examined in future study.

There are several implications resulting from this study in the fields of worked example instruction, problem solving, and game-based learning environments. The present study confirmed that worked example is effective in a game-based problem-solving task. In this study, worked example instruction produced a significant increase in content understanding and problem-solving strategies compared to the control group. The utility of a knowledge mapping system to assess content understanding was also indicated. In addition, a new scoring system for problem-solving strategy was developed in this study. However, the validity of this new system needs to be evaluated in future studies. Perhaps the most important result is that providing effective instructional strategies, that is, worked examples, could enhance the training effectiveness of commercial off-the-shelf computer games.

According to Moreno (2004), more study needs to be conducted to determine which instructional methods are most effective for which educational objectives and which learners, based on how an educational game should be designed. The results in this study provided evidence that using worked examples could be one of the good instructional methods to facilitate adults problem solving with a commercial off-the-shelf computer game.

According to Sweller (1994, 1998, 2000, 2004) and Sweller and Chandler (1994), cognitive load theory suggests that the limitation of working memory is the most critical factor when students are studying instructional material. Many studies confirmed that worked example could enhance learning by reducing learners' cognitive load (e.g., Carroll, 1994; Tarmizi & Sweller, 1988; Chi et al., 1989; Renkl et al., 2002; Sweller et al., 1990; Ward & Sweller, 1990). The results of this study were consistent with these studies. The use of worked examples in a game-based problem-solving task was supported. However, the improvement was unsatisfactory. Further study is needed to evaluate the effective design of worked examples in a game-based environment.

For example, in Renkl et al.'s (2002) study, the procedure of fading, which integrated worked examples and problem solving, could build a bridge between worked example study in early phases of cognitive skill acquisition and problem solving in later stages. They found the fading procedure could enhance the performance of transfer problems. In addition, it was more favorable to backward fading procedure (omitting the last solution step first). In the study conducted by Renkl and Atkinson (2003), the conclusion showed that prompting for self-explanations in the worked examples could facilitate fading procedure to be more effective. In addition, according to Atkinson, Renkl, and Merrill (2003), combining backward fading with self-explanation prompts could foster transfer learning by encouraging learners to identify the underlying principle demonstrated in each worked example solution step. Since this study was the first study to investigate the effectiveness of worked example in a game-based problem-solving task, the worked examples used in this study were simple and did not consist of fading or self-explanations procedure. Thus further study is needed to examine the effectiveness of fading and self-explanations on worked examples in a game-based environment.

In summary, based on the relevant studies, the merits of the worked example used in this study were (a) integration of diagrams and text (Sweller et al., 1998), (b) multiple examples (Sweller, 1990), and (c) high variety (Atkinson et al., 2000). Although the results suggested that worked example is superior to problem solving on enhancement of problem solving in a game-based environment, the improvement was small. In order to obtain greater improvement, the worked example instruction could add: (a) practice problems (Pass, 1992; Van Merriënboer, 1990; Stark et al., 1999), (b) fading procedure

(Renkl et al., 2002), (c) self-explanations (Atkinson et al., 2003; Renkl et al., 2004), (d) verbal instruction (Mayer, Moreno, & Boire, 1999), and (e) subgoals (Catrambone, 1994; Catrambone & Holyoak, 1990) in future study.

References

Adams, P. C. (1998). Teaching and learning with SimCity 2000 [Electronic Version]. Journal of Geography, 97(2), 47–55.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35–44.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research, 70*(2), 181–214.

Atkinson, R. K., Mayer, R., & Merrill, M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent s voice. *Contemporary Educational Psychology, 30*, 117–139.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783.

Baker, E. L. & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*, 269–282.

Baker, E. L. & O'Neil, H. F., Jr. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior, 18*(6), 609–622.

Brush, T. & Saye, J. (2001). The use of embedded scaffolds with hypermedia-supported student-centered learning. *Journal of Educational Multimedia and Hypermedia, 10*(4), 333–356.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology, 95*(3), 629–640

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*(3), 360–367.

Cary, M. & Carlson, R. A. (1999). External support and the development of problem-solving routines. *Journal of Experimental Psychology. Learning, Memory and Cognition, 25*(4), 1053–1070.

Catrambone, R. (1994). Improving examples to improve transfer to novel problems. *Memory & Cognition, 22*, 606–615.

Catrambone, R. & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18*(6), 593–603

Chambers, C., Sherlock, T. D., & Kucik, P. (2002). The Army Game Project. *Army, 52*(6), 59–62.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition & Instruction, 8*(4), 293–332.

Chandler, P. & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233–246.

Chandler, P. & Sweller, J. (1996) Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*(2), 151–170.

Chen, H.-H. (2005). *A formative evaluation of the training effectiveness of a computer game*. Unpublished doctoral dissertation, University of Southern California.

Chi, M., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

Cooper, G. & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347–362.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied, 7*(1), 68–82.

Faria, A. J. (1998). Business simulation games: current usage levels-an update. *Simulation & Gaming, 29*, 295–308.

Gagné, E. (1977). Educating delinquents: A review of research. *Journal of Special Education, 11*(1), 13–27.

Gerjets, P. & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*(1), 33–41.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*(1–2), 33–58

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*(2), 229–251.

Hannafin, M., Land, S., Oliver, K., & Reigeluth, C. M. (1999). Open learning environments: Foundation, methods, and models. In *Instructional-design theories and models: A new paradigm of instructional theory, Vol. II* (pp. 115–140). New Jersey: Lawrence Erlbaum Associates.

Herl, H. E., O'Neil, H. F., Jr, Chung, G., & Schacter, J. (1999) Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computer in Human Behavior, 15*, 315–333.

Howard, B. C., McGee, S., Hong, N. S., & Shia, R. (2000, April). *The influence of Metacognitive self-regulation on problem-solving in computer-based science inquiry*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Jeung, H.-J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*(3), 329–343.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology*, *21*(1), 5–23.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579–588.

Koedinger, K. R. & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science, 14*(4), 511–550.

Lane, D. C. (1995). On a resurgence of management simulations and games. *Journal of the Operational Research Society, 46*, 604–625.

Larkin, J., McDermott, J., Simon, D., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science, 208*(4450), 1335–1342.

Lim, E. L. & Moore, D. W. (2002). Problem solving in geometry: Comparing the effects of non-goal specific instruction and conventional worked examples. *Educational Psychology, 22*(5), 591–612.

Low, R. & Over, R. (1990). Text editing of algebraic word problems. *Australian Journal of Psychology, 42*(1), 63–73.

Low, R. & Over, R. (1992). Hierarchical ordering of schematic knowledge relating to area-of-rectangle problems. *Journal of Educational Psychology, 84*(1), 62–69.

Mawer, R. F. & Sweller, J. (1982). Effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 8*(3), 252–259.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem-solving. *Computer in Human Behavior, 18*, 623–632.

Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E. & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist, 38*(1), 3–52.

Mayer, R. E., Moreno, R., & Boire, M. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638–643.

Mayer, R. E., Moutone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171–185.

Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401.

Mayer, R. E. & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Ed.), *Handbook of educational psychology* (pp. 47–62). New York, NJ: Macmillian Library Reference USA, Simon & Schuster Macmillan.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*(1–2), 99–113.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.

Mwangi, W. & Sweller, J. (1998). Learning to solve compare word problems: the effect of example format and generating self-explanations. *Cognition and Instruction, 16*(2), 173–199.

O'Neil, H. F., Jr. (1999). Perspectives on computer-based performance assessment of problem-solving. *Computers in Human Behavior, 15*, 225–268.

O'Neil, H. F., Jr, Baker, E. L., & Fisher, J. Y.-C. (2002). *A formative evaluation of ICT games*. Los Angeles: University of Southern California; UCLA/National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

O'Neil, H. F., Jr & Fisher, J. Y.-C. (2002). A technology to support leader development: Computer games. In V. D. Day & S. J. Zaccaro (Eds), *Leadership development for transforming organization*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr & Herl, H. E. (1998). *Reliability and validity of a trait measure of self-regulation*. Los Angeles, University if California, Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Owen, E. & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology, 77*(3), 272–284.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory. *Educational Psychologist, 38*(1), 5–71.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434.

Pawley, D., Ayres, P., Cooper, M., & Sweller, J. (2005). Translating words into equations: A cognitive load theory approach. *Educational Psychology, 25*(1), 75–97.

Phye, G. D. (1998, April). *Components of shelf-regulation during within- and between-domain problem-solving performance*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Renkl, A. & Atkinson, R. K. (2003). Structuring the Transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15–22.

Renkl, A., Atkinson, R., & Grobe, C. S. (2004). How fading worked solution steps works – A cognitive load perspective. *Instructional Science, 32*(1–2), 59–82.

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: smooth transitions help learning. *The Journal of Experimental Education, 70*(4), 293–315.

Rosenshine, B. & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership, 49*, 26–33.

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.

Schank, R. C. (1997). *Virtual learning: A revolutionary approach to build a highly skilled workforce*. New York: McGraw-Hill Trade.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (2000). Does the combination of worked-out examples and problem-solving tasks pay off? *Psychologie, 14*(4), 206–218.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (1999). Instructional means to overcome transfer problems in the domain of economics: Empirical studies. *International Journal of Educational Research, 31*(7), 591–609

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457–466.

Sweller, J. (1990). Cognitive processes and instruction procedures. *Australian Journal of Education, 34*(2), 125–130.

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. *Australian Journal of Psychology, 45*(1), 1–8.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning & Instruction, 4*(4), 295–312.

Sweller, J. (2004, April). *Why understanding instructional design principles requires an understanding of the evolution of human cognitive architecture*. Paper presented at the meeting of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angles, CA.

Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176–192.

Sweller, J. & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition & Instruction, 2*(1), 59–89.

Sweller, J., Mawer, R., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *American Journal of Psychology, 95*(3), 455–483.

Sweller, J., Mawer, R., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology, 112*(4), 639–661.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 1*0(3), 251–296.

Tarmizi, R. A. & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424–436.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology, 3*(4), 257–287.

Theodorou, E. S. & Bonnie, J. F. (2001, April). *Can self-regulated learning predict transfer of problem-solving and text structure?* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Tuovinen, J. E. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334–341.

Van Gerven, P. W., Paas, F. G. W. C., Van Merriënboer, J. J. G., & Schmidt, H. G. (2002). Cognitive load theory and aging: effects of worked examples on training efficiency. *Learning and Instruction, 12*(1), 87–105.

Van Gerven, P. W., Paas, F. G. W. C., & Schmidt, H. G. (2000). Cognitive load theory and the acquisition of complex cognitive skills in the elderly: Towards an integrative framework. *Educational Gerontology, 26*(6), 503–521.

Van Gog, T., Paas, F., & Van Merriënboer, J. (2004). Process-oriented worked examples: improving transfer performance through enhanced understanding. *Instructional Science, 32*(1–2), 83–98.

Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6*(3), 265–285.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-Model. *Educational Technology Research & Development, 50*(2), 39–64.

Vygotsky, L. S. (1989). Concrete human psychology. *Soviet Psychology, 27*(2), 53–77.

Ward, M. & Sweller, J. (1990). Structuring effective worked examples. *Cognition & Instruction, 7*(1), 1–39.

Wertsch, J. V. & Resnick, L. B. (Ed.) (1991). A sociocultural approach to socially shared cognition. *Perspectives on socially shared cognition* (pp. 85–100). Washington: American Psychological Association.

Woolfolk, A. E. (2001). *Educational Psychology* (8th ed.). Needham Heights, MA: Allyn and Bacon.

Wulf, G. & Shea, C. (2002), Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review, 9*(2), 185–211.

Zhu, X. & Simon, H. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–66.

TRAINING VISUAL ATTENTION WITH VIDEO GAMES: NOT ALL GAMES ARE CREATED EQUAL

J. E. Cohen, C. S. Green and D. Bavelier

University of Minnesota

Abstract

Playing action video games enhances visual selective attention; however, little is known about the facets of video game play that contribute to such enhancements. To address this issue, participants with little to no previous gaming experience were trained on one of several video games. Each training game was selected to emphasize different aspects of typical game play that may contribute to learning. Participants were tested on two paradigms designed to explore dynamic aspects of visual selective attention. The first, the attentional blink task, measures the temporal resolution of visual attention; the second, the multiple object tracking task, measures the number of objects that can be simultaneously attended over a period of several seconds. Participants trained for 12 h on an action video game showed improvement on these two measures of visual attention. Individuals trained on a variety of other games showed comparatively less or no improvement. These results point out the combined requirements of monitoring several objects at once and being highly engaged in a very fast paced game as key factors in producing changes in visual selective attention.

1. Introduction

There is a substantial literature that demonstrates the positive effects that video game play can have on cognitive and perceptual abilities (see Green & Bavelier, 2006a for a review). For instance, previous studies of video games have shown that game play can lead to faster reaction times in healthy adults (Bialystok, 2006; Orosy-Fildes & Allan, 1989), young children (Yuji, 1996), and in the elderly (Clark, Lanphear, & Riddick, 1987; Goldstein et al., 1997). There are studies that have found evidence of enhanced

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

motor coordination related to video game experience (Griffith, Voloschin, Gibb, & Bailey, 1983) and produced by game training in the elderly (Drew & Waters, 1986). A number of video game training regimens have produced enhancements in skills such as spatial visualization, mental rotation, and distinguishing between trajectories of moving objects (Dorval & Pepin, 1986; Gagnon, 1985; McClurg & Chaille, 1987; Subrahmanyam & Greenfield, 1994). Several studies have demonstrated enhancements in various of aspects of visual attention, such as the number of objects that can be attended at once, the spatial and temporal distribution of visual selective attention, and the ability to divide attention (Green & Bavelier, 2003, 2006b,c, 2007; Greenfield, DeWinstanley, Kilpatrick, & Kaye, 1994). Finally, there is work that demonstrates that these enhancements may have real-world benefits, such as in the training of pilots or laparoscopic surgeons (Gopher, Weil, & Bareket, 1994; Rosenberg, Landsittel, & Averch, 2005).

The variety of different skills that can be modified by video game experience, and the degree to which they can be modified, is certainly striking. However, it is of great consequence to recognize that most of the studies comparing players and non-players have lumped together a wide variety of game types when determining which participants qualify as "video game players." This fact is of critical importance when attempting to interpret the results of these studies or make predictions based upon them. One could argue that considering the effect of "video game play" on a certain skill is roughly equivalent to considering the effect of "playing a sport" on a skill. The literature on the effect of sports on perceptual and motor skills (and in fact, the great majority of the literature on perceptual learning in general) has demonstrated that the nature of the training greatly influences the types of effects that are observed. For instance, baseball experience decreases Go/NoGo reaction times but tennis does not (Kida, Oda, & Matsumura, 2005), while soccer and volleyball players show enhancements in voluntary orienting that swimmers and track athletes do not (Lum, Enns, & Pratt, 2002). In this chapter, we ask whether the varied environments and demands offered by different types of games will lead to distinct cognitive effects on their players. For this chapter, the term "video games" refers to a game wherein players interact with objects on a screen (displayed by either a computer or game console) for entertainment purposes. All of these games involve a human player responding to images and/or sounds controlled by either the computer/console or other human players, with the ultimate goal of winning points or achieving a given mission.

The hypothesis that the characteristics of a game are directly related to the types of processes that are modified is implicitly acknowledged in the types of games that experimenters have chosen when training participants. For example, when testing mental rotation skills, McClurg and Chaille (1987) used The Factory (Kosel & Fish, 1983) as one of their training games because it explicitly requires the mental rotation of a figure as you plan the figure's progress through a series of virtual stamping and cutting machines to achieve a desired final shape. When looking to improve the speed of selection processing, Clark et al. (1987) chose Pac Man (Namco, 1979) and Donkey Kong (Nintendo, 1981) as their training games, because successfully playing these games requires the rapid selection of responses to sudden events. In an attempt to improve pilot performance, Gopher et al. (1994) made use of a modified version of the game Space Fortress (Donchin, Fabiani,

& Sanders, 1989), which was purposely altered to tap the types of attentional skills present during flight. In our work on visual attention (Green & Bavelier, 2003, 2006b,c), we have specifically looked at the effect of "action games," as these games make heavy demands upon the types of processes we have studied (efficient monitoring of the periphery for the occurrence of unpredictable events, tracking of many fast-moving objects, effective distractor rejection, etc.).

The hypothesis that games with different requirements should have different effects on cognitive processing is also represented by the video game we have chosen as a "control" video game in our studies. The game Tetris (Alexey Pajitnov, 1985) was chosen as a control because it places great demands on visuo-motor coordination, but otherwise makes very different demands on visual attention than the experimental action game. For instance, Tetris requires participants to pay attention to just one block at a time, only a few spatial locations have to be attended during the game, their locations are always predictable (the piece falling and the game board), and the trajectory of the attended piece is entirely under the control of the user (thus, unlike in the action game, Tetris players know at all times where their attention should be directed). Although Tetris was chosen as our control because it is unlikely to alter the specific visual attentional processes we have examined, were we to measure the effect of video game experience on mental rotation Tetris would be a good candidate for the experimental game (see Sims & Mayer, 2002).

In the only study (to our knowledge) that specifically attempted to measure the effects of different categories of games, Gagnon (1985) found a relationship between performance on a test of spatial orientation and successful training on the three-dimensional game Battlezone (Atari, 1980) but not the two-dimensional game Targ (Exidy, 1980) while measures of spatial visualization and visual pursuit were improved by playing either game. This result therefore supports the suggestion that game content or format determines the nature of cognitive and perceptual skills that can be acquired through video game experience.

While categories of video games available in 1985 could be broadly categorized by whether they were two- or three-dimensional, commercially available games today are often extremely complex and contain a variety of traits that may contribute to learning. Today's games range from those that require indiscriminate killing of enemies, to virtual versions of popular sports, to slow-paced turn-based strategy games, just to describe a few standard formats. Critically, these formats vary widely in the extent and nature of spatial and temporal attentional demands (as characterized, for example, by the number of attackers or the rate at which they appear), the players' emotional and cognitive engagement in the game, the predictability of events, and the visual complexity of the virtual environment. Games also vary in regard to whether the play is individual or team-based and interactive. Video game journals employ a rough classification scheme that divides games into a number of broad categories, including first-person shooter, third-person shooter, fighting, racing, role-playing, simulation, sports, and puzzle games; however, no system of categorization can cleanly partition all games as many incorporate more than one form of play. Furthermore, there is much variability to be found within these semi-distinct genres as even games with similar basic premises differ widely in terms of speed, difficulty, violence, and optimal strategies of play.

The goal of the present study is to further characterize the components of the gaming experience that enhance visual selective attention. Visual selective attention allows the enhanced processing of elements of the visual scene that are relevant to a specific task and can minimize interference from task-irrelevant distractors. Visual attention can be seen as an interaction between an external visual stimulus activating low-level perceptual processes and top-down control based on experience and task-specific knowledge. These attentional mechanisms play an important role in perceptual learning. For example, Ahissar and Hochstein (1993) found that adult participants who practiced one task did not improve on an alternate task, despite the fact that both tasks presented the exact same visual stimuli. This lack of transfer, they argued, was in part due to requiring attention to different stimulus attributes across tasks. Thus, exposure to the stimuli was insufficient to produce learning; rather learning appears limited by the way visual attention is distributed over the scene. While there is some evidence that low-level perceptual learning can occur in the absence of attention or even awareness (Fiser & Aslin, 2001; Watanabe et al., 2001), most evidence in the literature points to highly enhanced perceptual learning for attended aspects of visual input (e.g., Ahissar & Hochstein, 1993; Ball & Sekuler, 1987; Fiorentini & Berardi, 1980; Ramachandran & Braddick, 1973; Shui & Pashler, 1992). The allocation of attention is greatly influenced by task demands including the spatial distribution of targets, speed of target presentation, and overall task difficulty. These changes in visual attention are then reflected in the extent of perceptual learning that can occur (Ahissar & Hochstein, 1997, 2000). Thus, the nature of the visual environment and the task demands presented by a specific video game are likely to be highly relevant to the aspects of visual processing that may be enhanced.

With so many game-related factors in play, it is highly impractical to design and carry out a study that tests every possible combination of game criteria, content, and level of challenge. Thus, for the present study, we selected games that are representative of four popular genres, as well as one clinical training paradigm. Training games were selected so as to place different demands on temporal and capacity-related aspects of visual selective attention, that is, they differ from one another in terms of game speed, the number of objects that must be kept track of during the course of the game, and the frequency, speed, and precision of responses required from players. The clinical training paradigm consisted of the Interactive Metronome rhythmicity-training (Interactive Metronome Inc., *www.interactivemetronome.com*). It was selected as it places very high demands on timing skills and motor coordination and has been associated with cognitive improvements while being otherwise very unlike typical video game play (see Section 2.2, for a detailed description of Interactive Metronome training and its posited effects). The training groups were as follows: *Experimental group* – played a standard first-person action game (Unreal Tournament 2004 – Epic Games Inc., 2004); *Control Group #1* – played a slower-paced, team-based, first-person shooting game (America's Army – Pragmatic Solutions Inc. and Army Game Project, 2002); *Control Group #2* – played a slower-pace, first-person, multi-ball sport game (Harry Potter: Quidditch World Cup – Electronic Arts Inc., 2003); *Control*

Group #3 – played a speeded visuo-motor puzzle game Tetris (Alexey Pajitnov, 1985), *Control Group #4* – received rhythmicity training (Interactive Metronome – Interactive Metronome Inc., 1993); and the *Baseline Group* – played a set of basic computer Card Games (Solitaire – Micosoft, 1990; Free Cell – Microsoft, 1992; Hearts – Microsoft, 1993) and Minesweeper (Microsoft, 1992).

Participants were tested before and after training on two different aspects of visual attention: the temporal dynamics of visual attention and the number of objects that can be attended. The temporal dynamics of visual attention was assessed by using the attentional blink paradigm (Raymond, Shapiro, & Arnell, 1992), which measures how attention, once allocated to an item, recovers over time. The number of objects that can be attended over time was investigated using the multiple object tracking task which tests the participants' ability to track multiple moving objects over the course of several seconds (Pylyshyn & Storm, 1988). Performance on the attentional blink and multiple object tracking before and after training was compared for each group. To control for test–retest improvements, change in performance was then compared between each game and the baseline control group (Card Games). This comparison also allows us to rule out any possible unforeseen effects of the training process, such as Hawthorne-like enhancements produced through attention and encouragement from the experimenters during training or changes in participants' expectations after completing training which they often deduce is meant to improve their post-test performance (Benson, 2001). Finally, to test the relative effectiveness of our selected experimental action game, improvements in participants trained on the experimental action game (Unreal Tournament) were compared to that of each of the other groups.

Although the training games shared many features, the standard action game proved more effective in improving performance on these two tests of the temporal dynamics of visual attention than the other training regimens.

2. Methods

2.1. Participants

All participants were undergraduate or graduate students at the University of Rochester with ages ranging from 18 to 29 years (mean age = 20.5). In a pre-screening interview, participants' vision was tested on a 10-ft Tumbling "E" eye chart. As all testing and training paradigms in the study were run binocularly, participants were accordingly allowed to use both eyes when viewing the eye chart. One participant was excluded because he was unable to correctly identify the E's orientation at a resolution of 20/20. Thus, all included participants were confirmed to have normal or corrected to normal vision. No participants reported any sensory, neurological, or attentional impairment. Written informed consent was obtained from each participant and all participants were paid $8 for each hour of participation. Participants were required to be non-action, non-sports video game players. The criterion for this classification was playing *less* than 1 h per week of action or sports games in the preceding 12 months. Each participant filled out a survey, describing any gaming experience they had over the previous year, and were asked to estimate the number of hours per week they had spent playing different types of video games. Participants were asked to categorize their game experience in four different categories: action (examples: Unreal Tournament, Medal of Honor), sports (examples: NBA Live, Madden NFL), strategy (examples: The Sims, SimCity), and others. Because video games are rather ubiquitous in today's 20-something age group, some experience with slower paced games was permitted (0–1 h per week). Participants who reported playing action or sports games for more than 1 hour per week were not included in the study. All qualifying participants underwent training as described in the next section.

This study reports results for all participants that successfully completed pre-testing, training, and post-testing. Thus, the final experimental group: Unreal Tournament ($N = 11$, mean age $= 21.6$) included 5 females and 6 males, control group #1: America's Army ($N = 14$, mean age $= 19.9$) included 6 females and 8 males, control group #2: Harry Potter ($N = 14$, mean age $= 20.1$) included 7 females and 7 males, control group #3: Tetris ($N = 14$, mean age $= 19.5$) included 6 females and 8 males, control group #4: Interactive Metronome ($N = 19$, mean age $= 20.8$) included 9 females and 10 males, and the baseline group: Card Games ($N = 12$, mean age $= 21.5$) included 6 females and 6 males.

2.2. Training

Training was randomly assigned and consisted of playing the pre-determined video game, or performing the prescribed exercises, for 1 h per day for a total of 12 h with a minimum of 3 h and a maximum of 5 h of training completed per week. These time requirements were set in accordance with the suggested training rate for Interactive Metronome. All groups, except for the Interactive Metronome group, that did rhythmicity training, played their respective games on 20 inch Dell Flat Panel displays.

The 11 members of the *experimental group* played the game Unreal Tournament 2004 (henceforth referred to as the action video game). This game was chosen to be similar to those played by the expert video game player group in our previous studies. It has a relatively simple interface, uses first-person point of view and requires effective monitoring of the entire visual field (extent from fixation about 13°-height $\times$16°-width). Unreal Tournament 2004 was chosen, in part, because there is no "script." Instead, the game is controlled by the action of 32 AI agents rather than linear story development. This keeps the game from becoming predictable after multiple hours of play and thus maximizes the attentional demands placed on the player. Each hour session of the action game was divided into three 20-min blocks. The difficulty of each block was adjusted based upon the kill/death ratio achieved by the participant. If, in a block, the player scored more than twice as many kills than they had deaths, the difficulty level was increased one level. Participants were not permitted to move back down after achieving a given difficulty rating; they were instead required to attempt to master the new level.

The 14 members of *control group #1* played the game America's Army, a free access game created by the US army. While also a first-person shooter, with an equivalent interface and point of view, this game differed from the experimental action video game in several significant respects. First of all, participants had to go through a series of "basic training" exercises – such as firing range practice and an obstacle course – before being able to participate in an actual mission, reducing the time actually spent in a battle context. Secondly, this game places a much stronger emphasis on strategy and teamwork. It encourages such tactics as laying ambushes and sniping enemies, which, though cognitively engaging, are much less demanding for visual attention than constant engagement in a shootout. Thus, the pace of the game was generally slower than that of the experimental action game. Finally, participants played online against other human players. Although this adds a new, interesting dimension in the training, it prevents fine adjustments of the difficulty of each session as the number and skill level of the opponents our trainees faced could not be kept under experimental control. The size of the visual field used by the game was the same as that presented to the experimental group.

The 14 members of *control group #2* played Harry Potter: Quidditch World Cup. The first-person ball game is relatively fast-paced and, like the first two groups, creates a rich three-dimensional environment for the participant to navigate. It places strong demands on visual attention; it presents multiple balls, teammates, and opponents who must be kept track of for successful play. The game also presents the advantage of being much less violent and much more child and parent-friendly than the above two games. However, the fact that it is a children's game does reduce the general challenge and speed of the game. Also, like America's Army, Harry Potter requires the completion of training exercises that take away from actual game-time. Furthermore, it requires many hours of play and the collection of a variety of achievements before it allows players to advance to a more difficult level. It is therefore difficult to continually fully engage an adult player in this game. Again, the visual field used in the game was identical to that presented in the first two groups.

The 14 members of *control group #3* played the game Tetris. This game was selected to control for the effect of improved visuo-motor coordination, while placing little demand on the simultaneous processing of multiple items. Accordingly, the version of Tetris on which participants were trained had the "preview block" option turned off. The game was displayed to cover the entire extent of the screen. As such, the field of view of the Tetris game was actually slightly larger than that of the action game. The effective control game area extended 18°-height ×13°-width from fixation.

The 19 members of *control group #4* underwent rhythmicity training as designed by Interactive Metronome Inc. (*www.interactivemetronome.com*). This training paradigm requires extremely accurate timing. Participants are required to perform a set of simple physical movements (e.g., clapping) in synch with a slow steady beat. Accuracy feedback is provided in the form of guide sounds as well as numbers presented on a screen. The numbers represent the discrepancy, in milliseconds, between the participant's trigger hit and the given beat. Training was administered using the standard 12-session set format recommended by IM Inc. for unimpaired adults. To avoid confounds introduced by extensive interactions between participant and experimenter, the "auto-train" feature was used. In this format, participants are automatically shown the desired movements by a video guide and are given text instructions as to the number of repetitions to be completed. The program provides positive reinforcement (Skinner, 1953), in the form of a video of fireworks, when a trainee breaks a previously established personal record. Interactive Metronome training has been tied to improvements in attentional and cognitive skills, including vigilance, temporal sequencing, and motor planning, at least in children with attentional problems (Schaffer et al., 2001). In this study, IM provides a way to look at the attentional enhancements produced by honing the precision and accuracy of timing skills while leaving out virtually all other components of typical video game play.

The 12 members of the *baseline group* played a set of games selected to place relatively low demands on visual attention (Card Games). To avoid extreme boredom, they were allowed to choose between four games during any given session: Solitaire, Free Cell, Hearts, and Minesweeper. While often cognitively demanding, these games do not require constant vigilance, distributed attention, tracking of multiple objects, or fast responses to suddenly presented stimuli. This group served to control for any unexpected effects produced by participation in the study or, in other words, any cognitive or perceptual changes produced by being required to play a game, on a computer, in a lab setting, for 12 h (spread over the course of several weeks) while logging the time played and game performance. Furthermore, this group served as a control for any possible effects due to familiarity with the attentional blink and multiple object tracking tasks. These attentional tasks are described in Sections 3 and 4, respectively.

2.3. Pre- and Post-Testing

Two experimental paradigms that measure different aspects of the dynamics of visual attention were tested within a week before the time participants began training and again within a week after the day they finished training (note: to avoid interpretations based on arousal, participants were always post-tested at least 24 h after their final video game training session). These paradigms are described in greater detail in Sections 3 and 4.

2.4. Analysis

All analyses proceeded by first looking at test–retest improvement within each training group separately on the totality of the data. Group comparisons were then performed on the task conditions that have been shown in our previous work, and that of others, to be most sensitive to training (intermediate range between ceiling and floor). The experimental and control groups were first compared to the baseline group to evaluate the source of any test–retest improvement, if present. Then, the control groups were compared to the experimental group to evaluate the relative efficiency of each of these new training regimens compared to that of the action video game we have used in our previous work. Given the high number of training groups (one experimental, four controls, and one baseline), multiple tests had to be performed. We did not correct for multiple comparisons as each of these comparisons was fully planned; however, although the majority of our a priori hypotheses called for one-tailed statistics, two-tailed statistics were used to be more conservative.

3. Experiment 1: The Attentional Blink

The Attentional Blink paradigm was used to measure the effect of our different training regimens on how quickly attentional resources recover after being directed toward a target. In the attentional blink paradigm (Raymond et al., 1992), observers are exposed to a rapid serial visual presentation of items. In this study, we administered a version of the attentional blink task where participants viewed a stream of black letters and were required to identify a first target – a letter presented in white (Target 1) – and detect a second target – a black X (Target 2) – when present. The number of intervening letters between Target 1 and Target 2 was manipulated, and performance accuracy at the different Target 1-Target 2 intervals was measured. The term "attentional blink" refers to the finding that Target 2 is frequently not detected if it occurs just after Target 1; however, as the amount of time elapsed between Target 1 and Target 2 increases, detection of Target 2 improves, indexing that attention can again be summoned toward a new target. To control for changes in participants' baseline ability to detect Target 2, two control blocks were interleaved between the experimental blocks such that the tasks were completed in the following order: (1) control, (2) experimental, (3) control, (4) experimental. In the control blocks, the visual input was the same as in the experimental condition but participants were instructed to ignore the white letter and were only required to report the presence or absence of the X. Data from the control blocks were then used to correct for possible differences in baseline Target 2–detection rates in the dual condition; in other words, corrections were made to control for differences in the simple ability to parse the rapid, serial stream of letters and detect the "X" that could otherwise confound interpretations about the depth of the attentional blink.

Our past work has shown that action video game play results in faster recovery from the attentional blink (Green & Bavelier, 2003). We therefore predicted that attentional recovery would be more rapid for experimental action game players than for the card game group (baseline group) in which fast and accurate processing is not of the essence. Based on our previous findings, we also expect improvement in the experimental group to exceed that of the Tetris group (control group #3). The effects of control games #1 (America's Army), #2 (Harry Potter), and #4 (Interactive Metronome) on this measure are more difficult to predict. Although America's Army and Harry Potter do place a premium on selective visual attention skills, it is not clear that the speed and difficulty of these two games will be sufficient to produce measurable improvements on the attentional blink after only 12 h of play. The effectiveness of Interactive Metronome training will depend on whether or not extremely precise timing in cognitive and motor planning skills will be sufficient to produce changes in the temporal resolution of attention.

3.1. Stimuli and Procedure

The same set-up as in Green and Bavelier (2003) was used. Participants viewed the stream of letters displayed on a gray background in the center of a computer monitor. Their heads were positioned in a chin-rest fixed 57 cm away from the screen. The letters subtended 1° of visual angle and were presented for 15 ms at a rate of 1 per 100ms. Participants pressed a key to begin each trial. This key press was followed by the presentation of 7–15 black letters (randomly chosen but excluding X), then the first target (Target 1) was presented in white followed by 15–18 black letters (including an X 50% of the time). Four blocks were administered: in the first and third blocks, participants were only required to respond to the presence of an X while in the second and fourth blocks they needed to identify the white letter prior to indicating whether or not they had seen an X. In these cases, once stimulus presentation was complete, participants typed in the identity of the white letter and pressed a "Y" or "N" key to indicate whether or not they had seen an X. The time elapsed between the white letter and the X is termed lag and measured in terms of number of items (the letter presented immediately after Target 1 was labeled as being displayed at "Lag 1" while the second letter was displayed at "Lag 2" and so forth).

3.2. Results

All groups exhibited a marked attentional blink with worse "X detection" performance when the time elapsed between the white letter and the X was short and a gradual recovery of performance as that time lag increased. A main test–retest improvement was noted in the experimental group as well as in the Interactive Metronome group and, marginally so, in the America's Army group. In isolation, a test–retest improvement does not establish that the training was successful in inducing better performance. It merely demonstrates that participants doing the task for a second time improve their performance when compared to the first time, an effect that can be due to practice at the task itself, independent of the intervening training regimen. To establish a causal effect of a training regimen, it needs to be shown that the test–retest improvement found in the training group is larger than that of the baseline group. Results are given in detail in Appendix 1 (see also Tables A1 and A2).

The efficacy of each of the training regimens was evaluated by contrasting test–retest improvement from each training regimen with that seen in the baseline group. As a faster recovery from the attentional blink following training is best seen at those lags where the attentional blink is typically most pronounced (lags 2, 3, 4), these analyses focused on these three lags. While the Interactive Metronome and America's Army groups improved significantly at these lags following training (comparing their pre-test score to their post-test score), only the experimental group showed a level of improvement that was significantly greater than what was observed in the baseline group (Figure 1). Therefore, only the experimental group can convincingly be said to exhibit a faster recovery from the attentional blink than what one would predict from simple test/retest benefits.

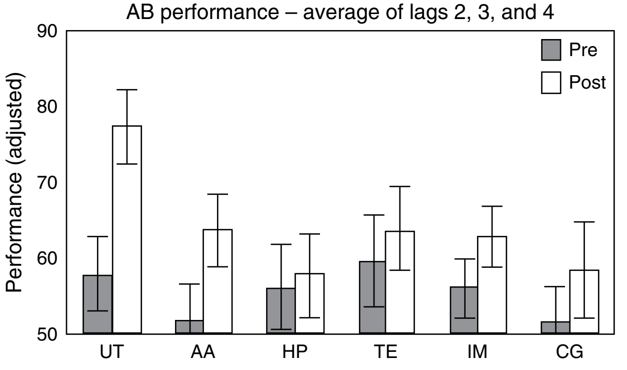

Figure 1. Only the experimental, action video game trained group demonstrated an improvement in performance greater than what was seen in the baseline group. The amount of improvement seen in the experimental group was greater than every other training game except America's Army.

In addition, when the test–retest improvement in the experimental group was compared to that seen in the other training regimens, the experimental group improved significantly more than all except America's Army. Taken together, the fact that the experimental group was the only group to improve significantly more than the baseline group and that this group improved significantly more than all but one of the other groups suggests that the experimental action game had the greatest effect on the speed of attentional recovery.

4. Experiment 2: Multiple Object Tracking

In the multiple object tracking paradigm, participants are required to track the movements of multiple moving objects. In this version of the task, participants viewed 16 randomly moving yellow circles. At the beginning of the trial, some subset of these circles was cued by being colored blue. After 2 s, the cued circles turned back to yellow, and participants were required to keep track of the circles that had been cued (now visually indistinguishable from uncued circles) as they continued to move randomly about the screen. After several seconds of tracking, one of the circles turned white and the participant was asked to decide whether or not it belonged to the cued set (yes/no decision). This method of response, rather than the more typical method of asking the participant to indicate each of the initially cued objects, was employed to reduce memory demands during the response process and thus allow a more accurate estimation of the tracking capabilities per se.

Based on our previous results, we expected significant improvements in the experimental – Unreal Tournament – group but little improvement in the Tetris group. Additionally, the Card Game and Interactive Metronome groups were not expected to improve much in multiple object tracking performance as these forms of training do not require participants to deploy attention across multiple objects. Possible improvements were expected in the America's Army and Harry Potter groups as these games do require the player to keep track of multiple objects. However, it was not known whether the speed and difficulty of these two control games would be sufficient to produce measurable changes after only 12 h of play.

4.1. Stimuli and Procedure

The same set-up as in Green and Bavelier (2006b) was used, except cued circles were colored blue instead of red and non-cued circles were colored yellow instead of green. This change was made to avoid any complications produced by possible undiagnosed red–green color-blindness in our participant group. Furthermore, as previous work has found eye movements to have few implications for performance on this task (Pylyshyn & Storm, 1988) and our previous work indicates that fixation performance on this task is roughly equivalent for gamers and non-gamers (Green & Bavelier, 2006b), participants were not eye-tracked during the task.

Each observer viewed the display binocularly with his or her head placed in a chin rest at a test distance of 57 cm. Participants were instructed to fixate within a center ring (radius = 0.25 deg). Participants pressed a key to begin each trial. Each trial began with 16 circles (radius 0.5 deg) moving randomly at a rate of 5 deg/sec on a circular gray background (radius of circular background = 10 deg). The circles repelled one another before contact (0.5 deg minimum separation), were repelled by the outer edges of the background and by the center fixation circle. At the beginning of the trial, 1–7 of the circles were cued by being colored blue while the remaining circles were drawn in yellow. Participants were instructed to attend to the blue circles. Participants were warned that the blue circles would shortly change back to yellow, after which time they had to continue tracking the same circles that were previously cued. The cued circles changed to yellow after 2 s, leaving all 16 circles visually indistinguishable as they continued to move randomly about the screen. After 5 s of tracking, motion ceased and one of the circles was highlighted in white (probe circle). At this point the participant was asked to press a Yes or No key to indicate whether or not the probe circle had been cued at the beginning of the trial. The probe circle was one of the originally cued circles 50% of the time. Each number of cued circles (1–7) was presented 20 times (10 yes, 10 no) for a total of 140 trials.

4.2. Results

In all groups, the ability to correctly identify whether or not a given circle had been cued at the beginning of the trial decreased as the number of circles to track increased, replicating the standard performance on the multiple object tracking task. Three training groups improved significantly between pre- and post-testing; these included the experimental (Unreal Tournament), the Interactive Metronome, and the Tetris group. As discussed earlier, such test–retest improvements do not allow one to distinguish between a genuine

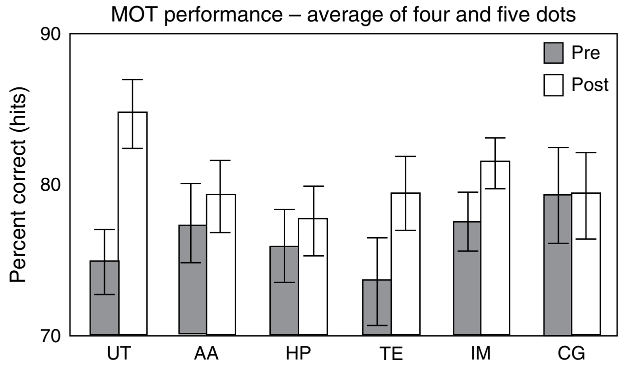

Figure 2. Only the experimental, action video game trained group demonstrated an improvement in performance greater than what was seen in the baseline group.

effect of training or a simple improvement as participants do the task for the second time. A causal effect of training regimen can only be shown by comparing test–retest improvements from the training group to that of the baseline group.

This was done by focusing the analysis on two critical attentional loads (4–5 circles) where performance is neither at ceiling nor at floor, and therefore most sensitive to experimental manipulation (Green & Bavelier, 2006b; Trick, Jaspers-Fayer, & Sethi, 2005). Only the experimental group showed a significantly greater test–retest improvement in accuracy than what was observed in the baseline group (Figure 2). When the improvement in the experimental group was compared to that in the other training groups, the experimental group was seen to improve significantly more than the Harry Potter group, with a discernable trend in the same direction also being seen for the America's Army and Interactive metronome groups. While the comparison between the experimental group and the Tetris group was not significant, previous research with approximately three times more training has demonstrated significant difference between these regimens (Green and Bavelier, 2006b). Taken together, these results suggest that the experimental action game is best suited to improving this aspect of attention. More detailed results are given in Appendix 2 (see Tables A3 and A4).

5. General Discussion

The goal of this study was to assess effects of a variety of training paradigms on dynamic aspects of visual attention, as represented by performance on the attentional blink and multiple object tracking tasks. On the attentional blink task, only the experimental action group showed significantly greater performance than the baseline group on the critical lags where recovery is occurring, suggesting that an enhancement in the temporal dynamics of visual attention occurred only as a result of playing the experimental action game. Other groups did show improvements in their post-test performance, but these effects were not significantly different than what was observed in the baseline group, and thus are likely to represent simple test–retest enhancements. Furthermore, when the experimental action game group was compared to each of the four control games, the improvement in the action game group (Unreal Tournament) was significantly greater than that achieved by all of the other training groups other than the America's Army group that was not seen to differ from the experimental group. The fact that the improvement in the America's Army group was neither significantly greater than the baseline group nor significantly less than the experimental action game group suggests that America's Army had a small effect on the recovery of attention.

Similar results were found for the multiple object tracking task. When focusing on the critical attentional load of 4–5 to-be-tracked circles, only the experimental action game group showed a significantly greater improvement on these object tracking measures when compared to the baseline group. Although performance did improve in some of the other groups, the size of the effect in those groups was not different from the expected improvement due to simple test–retest effects.

Overall this work stresses the importance of well-tailored control groups in studies of training-induced learning. Although several training groups showed significant improvements on the tests of visual attention, it cannot be concluded that the training itself produced these enhancements. Some improvement is to be expected based purely on experience with the testing paradigms (test–retest effects). This factor needs to be removed by comparison to a baseline group that was equally exposed to the test paradigms but not trained in a way that should enhance performance on these paradigms. Only if greater improvement is established in the experimental group as compared to the control group can one conclude that training with the experimental regimen has had a causal effect on performance. The standard action game, Unreal Tournament 2004, was the only training that, in 12 h, produced significant improvements in our measures of the temporal resolution of visual attention and of the ability to track multiple objects over time as compared to the baseline group.

The other two control training games that were most similar to Unreal Tournament in terms of action gaming had relatively less effect. America's Army did not produce improvements greater than the baseline group in either experiment. However, the degree of improvement in this group on the attentional blink task was also not significantly less than the experimental group, suggesting that America's Army might have an intermediate effect on the temporal dynamics of attention. Harry Potter, despite its seemingly high attentional demands, also did not produce enhanced performance on either task. The lack of an effect of training observed in the America's Army and Harry Potter groups may in part be explained by our inability to maintain an adequate difficulty level to continue to engage our adult participants in the game, which is likely an important factor for learning. While the role of challenge in learning remains to be firmly established, the present study cautions against the use of a control group that is just asked to perform test–retest without intermediate training or which is trained on an "easy" version of the same task as the experimental group as it is likely that any less challenging training will always lead to less learning, regardless of the nature of the training.

The other two control groups (Interactive Metronome and Tetris) bore little similarity to action gaming, except for the fact that they were quite engaging and challenging for the participants. Neither Tetris training, as in our previous studies (Green & Bavelier, 2003), nor Interactive Metronome training led to improvements in either experiment beyond what was observed in our baseline control group. The Card Games, as expected, produced no significant changes in performance on either task and functioned as a baseline for comparing the effects of the other groups.

This study highlights several key traits in computer games that may facilitate changes in visual attention. The most efficient game for training (Unreal Tournament) was the fastest and the least predictable one. It presented frequent, widely distributed, unexpected events that require a fast and accurate response from the player. Finally, it was formatted in such a way that players were not encouraged to use more passive strategies, such as sniping, which tend to reduce game pace and attentional load.

However, because the games used here are merely representative of several popular types of games and differ, from each other, in many respects, it is difficult at this stage to pinpoint precisely which criteria are essential for learning. This study instead serves to narrow our scope and highlight which games and which game traits are good candidates for training visual attention. Within this smaller range, it will now be feasible to select a specific game and vary only specific aspects of play, such as level of challenge or pace of the game, and observe the effect of these variations on the success of training.

Acknowledgments

We thank R. Abejuela and J. Wenck for help with participants and data collection. This research was supported by grants from the National Institute of Health and the Office of Naval Research to Daphne Bavelier.

Appendix 1: Results for Attentional Blink

Analysis of Ability to Detect Target 2 Given Target 1 Had Been Identified

In order to measure recovery from the attentional “bottleneck” produced by the identification of the white letter (Target 1), we looked at the percent of correct Target 2 detections as a function of the time elapsed between Target 1 and Target 2. The letter presented immediately after Target 1 was labeled as being displayed at “Lag 1” while the second letter was displayed at “Lag 2” and so forth. To ensure subjects were allocating attention to the white letter, only trials in which Target 1 was correctly identified were used in our analyses. Two-tailed *p*-values are reported throughout this section.

Correct Target 2 detection scores on the dual-task trials were adjusted using results from the X-detect-only task and the following formula was applied: performance = 100-([T2 alone] – [T2 given T1 correct]). Thus, this performance measure reflects subjects' ability to recover attention following successful identification of the first target and is not confounded by subjects' ability to detect the second target alone. It therefore provides a more direct measure of the recovery of attentional resources after they have been allocated to the first target.

Using this adjusted measure, separate ANOVAs for each training regimen with test (pre/post) and lag (1–8) as factors were performed for each of the training groups (Unreal Tournament: UT, America's Army: AA, Harry Potter: HP, Tetris: TE, Interactive Metronome: IM, Card Games: CG) which revealed significant effects of lag for all groups, with the expected gradual recovery at increasing lags (Table A1) (UT, $F(7,\ 70) = 19.75,\ \mathrm{MSE} = 0.032,\ \mathrm{p} < 0.001,\ \eta_p^2 = 0.66$; AA, $F(7, 91) = 44.68,\ \mathrm{MSE} = 0.031,\ \mathrm{p} < 0.001,\ \eta_p^2 = 0.77$; HP, $F(7, 91) = 25.09,\ \mathrm{MSE} = 0.025,\ \mathrm{p} < 0.001\ \eta_p^2 = 0.66$; TE, $F(7, 84) = 20.73,\ \mathrm{MSE} = 0.037,\ \mathrm{p} < 0.001,\ \eta_p^2 = 0.63$; IM, $F(7, 126) = 48.88,\ \mathrm{MSE} = 0.036,\ \mathrm{p} < 0.001\ \eta_p^2 = 0.73$; CG, $F(7, 77) = 31.58$, $\mathrm{MSE} = 0.041$, $\mathrm{p} < 0.001$, $\eta_p^2 = 0.74$).

Furthermore, the main effect of test was significant for the experimental group (UT, $F(1, 10) = 20.76,\ \mathrm{MSE} = 0.020,\ p = 0.001,\ \eta_p^2 = 00.67$), the Interactive Metronome group ($F(1, 18) = 4.70,\ \mathrm{MSE} = 0.035,\ p = 0.04,\ \eta_p^2 = 0.21$) and marginally for the America's Army group ($F(1, 13) = 3.89,\ \mathrm{MSE} = 0.046,\ p = 0.07,\ \eta_p^2 = 0.23$). Finally, an interaction between test and lag was observed for the experimental group (UT, $F(7,\ 70) = 2.25,\ \mathrm{MSE} = 0.032,\ p = 0.04\ \eta_p^2 = 0.18$). This interaction highlights the fact that the speed at which subjects recover from the blink is best indexed by performance at the early lags. Accordingly to look at the difference in recovery across training, we performed another set of ANOVAs to examine group interactions at lags 2, 3, and 4. Lag 1 was not included as the AB paradigm we chose shows the typical saving from the blink at lag 1 ("lag 1 sparing" – Chun & Potter, 1995).

Separate ANOVAs for each training regimen using data from lags 2, 3, and 4 with test (pre/post) as the only factor revealed highly significant effects of test for the experimental group (UT, $F(1, 10) = 44.74,\ \mathrm{MSE} = 0.005,\ \mathrm{p} < 0.001\ \eta_p^2 = 0.82$), and more modest effects for the America's Army group ($F(1, 13) = 6.46,\ \mathrm{MSE} = 0.014,\ p = 0.025,\ \eta_p^2 = 0.33$), and marginally for the Interactive Metronome group ($F(1, 18) = 4.25,\ \mathrm{MSE} = 0.010,\ p = 0.054,\ \eta_p^2 = 0.19$). The other training groups showed no significant improvement at these critical lags (see Figure 1).

A series of 2 (training group) $\times 2$ (pre vs. post training) $\times 3$ (lags 2, 3, 4) ANOVAs were run comparing each group to the baseline group (control group #5: CG) in order to estimate the effect of training above and beyond test–retest improvement. The experimental group was the only training group that showed significantly more improvement than the baseline group (pre vs. post-test by training group interaction) (UT, pre-test vs. post-test accuracy (lags 2,3,4): 57.72 ± 4.89 vs. 77.25 ± 4.94; CG, pre-test vs. post-test accuracy: 51.54 ± 4.33 vs. 58.23 ± 6.37) (CG vs. UT, $F(1, 21) = 6.44,\ \mathrm{MSE} = 0.021,\ p = 0.019,\ \eta_p^2 = 0.23$). When the action group was compared to the five control groups, the action group improved significantly more than all of these, except America's

Table A1
Mean accuracy (standard error) at each of the eight lags before and after training (pre vs. post) for each training group.

Lag	1	2	3	4	5	6	7	8
Unreal Tournament								
Pre	74.47(7.79)	53.87(5.71)	50.67(6.06)	68.63(6.82)	91.74(7.05)	97.41(3.84)	103.67(5.87)	100.23(1.70)
Post	83.69(5.55)	67.33(7.54)	76.10(8.23)	88.33(3.23)	98.58(1.81)	102.73(2.92)	101.31(2.83)	101.10(1.16)
Difference	9.22	13.46	25.43	19.7	6.84	5.32	−2.36	0.87
America's Army								
Pre	64.47(7.17)	40.44(7.17)	50.73(6.52)	65.69(4.89)	85.81(6.08)	101.73(3.74)	102.82(2.00)	98.97(2.80)
Post	68.51(6.99)	53.69(6.24)	55.97(4.67)	81.31(4.95)	92.41(3.93)	104.38(2.29)	102.28(2.77)	97.50(1.12)
Difference	4.04	13.25	5.24	15.62	6.6	2.65	−0.54	−1.47
Harry Potter								
Pre	63.28(9.53)	41.25(5.75)	52.29(7.31)	74.09(7.36)	85.38(5.19)	97.69(3.94)	103.92(4.64)	104.3(2.03)
Post	70.91(7.90)	46.95(6.72)	48.68(7.54)	77.24(5.52)	89.64(4.76)	94.95(3.57)	95.61(3.22)	96.46(2.76)
Difference	7.63	5.7	−3.61	3.15	4.26	−2.74	−8.31	−7.84
Tetris								
Pre	69.39(6.15)	55.69(7.55)	58.03(7.99)	64.41(6.01)	91.47(3.89)	87.97(4.89)	96.40(5.57)	96.84(4.36)
Post	78.13(6.66)	52.57(7.98)	66.05(6.65)	72.12(5.91)	88.48(4.00)	101.01(3.97)	95.17(3.73)	100.61(2.15)
Difference	8.74	−3.12	8.02	7.71	−2.99	13.04	−1.23	3.77
Interactive Metronome								
Pre	65.17(4.64)	41.96(5.24)	51.80(5.86)	73.87(4.04)	84.18(4.45)	95.37(3.82)	101.58(3.67)	101.54(1.42)
Post	70.52(5.91)	51.56(6.35)	56.79(4.61)	79.33(4.16)	93.12(2.79)	99.59(2.37)	101.7(2.79)	100.02(1.26)
Difference	5.35	9.6	4.99	5.46	8.94	4.22	0.12	−1.52
Card Games								
Pre	63.94(8.67)	34.03(5.91)	49.68(4.60)	70.63(5.78)	88.34(4.39)	96.32(5.06)	106.71(3.63)	99.07(3.05)
Post	67.93(6.54)	50.51(8.27)	54.15(7.63)	70.03(7.26)	91.26(3.13)	95.86(3.10)	101.57(2.53)	99.07(0.71)
Difference	3.99	16.48	4.47	−0.6	2.92	−0.46	−5.14	0

To calculate accuracy, correct T2 detection scores on the dual-task trials were adjusted using results from the X-detect-only task according to the following formula: performance = 100−([T2 alone]−[T2 given T1 correct]). Thus, the measure reflects participants' ability to recover attention following successful id.

Army (AA vs. UT, $F(1,23) = 2.06$, $\text{MSE} = 0.030$, $p = 0.164$, $\eta_p^2 = 0.08$; HP vs. UT, $F(1,23) = 20.81$, $\text{MSE} = 0.014$, $p < 0.001$, $\eta_p^2 = 0.48$; TE vs. UT, $F(1,22) = 5.18$, $\text{MSE} = 0.041$, $p = 0.033$, $\eta_p^2 = 0.19$; IM vs. UT, $F(1,28) = 7.11$, $\text{MSE} = 0.024$, $p = 0.013$, $\eta_p^2 = 0.20$). Taken together, the fact that the experimental group was the only group to improve significantly more than the baseline group and that this group improved significantly more than all but one of the other groups suggests that the experimental action game had the greatest effect on the speed of attentional recovery. Finally, it should be noted that none of the training groups demonstrated a pre/post difference in performance on the target absent trials, which were analyzed separately. In fact, only the experimental group approached a significant improvement on X-absent trials (UT, $F(1,10) = 4.25$, $p = 0.07$), thus eliminating the possibility that the experimental group's improvement on X-present trials was due to a shift in criteria rather than an increase in sensitivity.

Analysis of Ability to Detect Target 2-Only

Separate ANOVAs were carried out for each training group with test (pre/post) and lag (1–8) as factors on the % correct for Target 2 detection only. This corresponds to performance on the separate blocks of trials during both pre- and post-testing in which subjects were told to ignore the white letter and report only whether or not an X was present in the stream of letters. Because lag was included as a factor, only trials in which the target X was present could be used in this analysis, as "lag" is meaningless when no target is present. All groups improved significantly between pre- and post-testing in their ability to detect T2 when they were not required to identify the white letter (Table A2) (UT, $F(1, 10) = 9.02$, $\text{MSE} = 0.009$, $p = 0.013$, $\eta_p^2 = 0.47$; AA, $F(1,13) = 10.17$, $\text{MSE} = 0.012$, $p = 0.007$, $\eta_p^2 = 0.44$; HP, $F(1,13) = 18.59$, $\text{MSE} = 0.028$, $p = 0.001$ $\eta_p^2 = 0.59$; TE, $F(1,13) = 15.08$, $\text{MSE} = 0.016$, $p = 0.002$, $\eta_p^2 = 0.54$; IM, $F(1,18) = 13.64$, $\text{MSE} = 0.011$, $p = 0.002$, $\eta_p^2 = 0.43$; SO $F(1,11) = 43.72$, $\text{MSE} = 0.006$, $p < 0.001$, $\eta_p^2 = 0.80$).

All groups showed a significant effect of lag, which reflects the white letter's ability to exogenously capture attention, even when it is not processed deliberately (Moroni et al., 2000; Raymond et al., 1992) (UT, $F(7, 70) = 4.83$, $\text{MSE} = 0.008$, $p < 0.001$, $\eta_p^2 = 0.33$;

Table A2
Mean accuracy (standard error) for the X-detect-only task before and after training (pre vs. post) for each training group.

	UT	AA	HP	TE	IM	CG
Pre	85.94(1.32)	87.78(1.07)	83.48(1.26)	83.98(1.39)	84.17(0.99)	85.42(1.10)
Post	90.20(1.13)	92.41(0.80)	93.14(0.83)	90.46(1.00)	88.61(0.88)	93.16(0.80)
Difference	4.26	4.63	9.66	6.48	4.44	7.74

AA, $F(7, 91) = 3.95$, $\text{MSE} = 0.008$, $p = 0.001$, $\eta_p^2 = 0.23$; HP, $F(7, 91) = 3.73$, $\text{MSE} = 0.009$, $p = 0.001$, $\eta_p^2 = 0.22$; TE, $F(7, 84) = 3.74$, $\text{MSE} = 0.009$, $p = 0.001$, $\eta_p^2 = 0.22$; IM, $F(7, 126) = 6.17$, $\text{MSE} = 0.011$, $\text{p} < 0.001$, $\eta_p^2 = 0.25$; SO, $F(7, 77) = 4.78$, $\text{MSE} = 0.005$, $\text{p} < 0.001$, $\eta_p^2 = 0.30$). There were no significant group-by-test interactions when groups were compared to either the experimental group or the baseline group. This finding also held true in a separate analysis for X-absent trials.

This improvement in baseline Target 2 correct detection that occurred across all the training groups reflects a more general test–retest improvement in the ability to segment the rapid stream of letters rather than an improvement in attentional recovery following the processing of Target 1. It is likely, however, to be an added source of variance when studying between-group differences in improvement of recovery from the attentional blink. Accordingly, this difference was corrected for as described in the previous section when analyzing Target 2 detection given the correct identification of Target 1, after ensuring that it was not the source of a significant difference across training groups.

Appendix 2: Results for Multiple Object Tracking

Object tracking performance was measured as the subject's percentage of correct responses to the yes (was cued)/no (was not cued) decision made about the given probe circle. The number of correct responses was recorded separately for each possible number of cued circles (1–7). There were no interactions found with expected response (yes vs. no) so performance measures were collapsed across "yes" and "no" trials. Separate ANOVAs with test (pre/post) and cued circles (1–7) as factors were performed for each training regimen. As expected, a main effect of number of circles to track was found for each group with accuracy decreasing with increasing number of circles (Table A3) (UT, $F(6, 60) = 47.02$, $\text{MSE} = 0.007$, $\text{p} < 0.001$, $\eta_p^2 = 0.82$; AA, $F(6, 78) = 55.19$, $\text{MSE} = 0.009$, $\text{p} < 0.001$, $\eta_p^2 = 0.81$; HP, $F(6, 78) = 69.66$, $\text{MSE} = 0.008$, $\text{p} < 0.001$, $\eta_p^2 = 0.84$; TE, $F(6, 78) = 81.73$, $\text{MSE} = 0.006$, $\text{p} < 0.001$, $\eta_p^2 = .86$; IM, $F(6, 108) = 52.91$, $\text{MSE} = 0.009$, $\text{p} < 0.001$, $\eta_p^2 = .75$; SO, $F(6, 66) = 25.54$, $\text{MSE} = 0.012$, $\text{p} < 0.001$, $\eta_p^2 = 0.70$).

Three training groups improved significantly between pre- and post-testing (Table A4): our experimental group, UT ($F(1, 10) = 5.32$, $\text{MSE} = 0.015$, $p = 0.044$, $\eta_p^2 = 0.35$), IM ($F(1, 18) = 5.90$, $\text{MSE} = 0.006$, $p = 0.026$, $\eta_p^2 = 0.25$) and TE ($F(1, 13) = 9.02$, $\text{MSE} = 0.006$, $p = 0.01$, $\eta_p^2 = 0.41$). There were no significant interactions between test and the number of circles for any of the groups.

It was consistently seen that while groups showed comparable performance with relatively few items, differences in performance emerged as some critical attentional load (4–5 circles) was reached. This pattern is consistent with our previous findings (Green & Bavelier, 2006b) as well as recent findings in children who play action video games (Trick et al., 2005). Accordingly, we ran a second set of 2 (training group) $\times 2$ (pre vs. post training) ANOVAs, looking only at performance on trials where four or five circles were cued at the outset. A series of ANOVAs comparing each group to the baseline game

Table A3

Mean accuracy (standard error) at each number of circles to be tracked (1–7) before and after training (pre vs. post) for each training group.

No. of circles	1	2	3	4	5	6	7
Unreal Tournament							
Pre	97.73(1.04)	94.55(1.25)	90.00(1.65)	78.64(2.95)	71.36(3.64)	66.36(3.44)	65.00(1.91)
Post	98.18(1.02)	95.00(1.65)	93.64(2.44)	87.27(3.59)	82.27(2.27)	71.36(4.53)	68.18(4.68)
Difference	0.45	0.45	3.64	8.63	10.91	5	3.18
America's Army							
Pre	100(0)	94.29(1.56)	88.21(2.90)	85.36(2.94)	69.64(2.99)	71.43(3.21)	62.14(3.13)
Post	99.64(0.36)	97.50(1.01)	86.43(2.31)	83.21(2.75)	75.36(3.53)	67.86(4.22)	68.98(3.64)
Difference	−0.36	3.21	−1.78	−2.15	5.72	−3.57	6.84
Harry Potter							
Pre	96.79(1.45)	97.14(0.86)	83.93(2.35)	81.07(3.36)	70.71(2.91)	62.86(2.95)	63.57(1.92)
Post	98.57(0.82)	93.93(1.67)	89.29(2.02)	81.79(3.13)	73.57(2.94)	65.00(3.06)	64.29(3.35)
Difference	1.78	−3.21	5.36	0.72	2.86	2.14	0.72
Tetris							
Pre	96.79(1.13)	92.86(1.94)	87.86(1.72)	78.93(3.48)	68.57(2.64)	63.93(3.64)	62.86(3.62)
Post	96.79(1.35)	91.79(1.93)	90.71(2.21)	82.86(3.04)	76.07(2.88)	68.93(2.58)	67.50(3.22)
Difference	0	−1.07	2.85	3.93	7.5	5	4.64
Interactive Metronome							
Pre	97.37(0.89)	92.37(2.04)	87.37(3.00)	81.58(2.20)	73.42(3.27)	74.47(3.26)	64.47(2.56)
Post	97.89(0.79)	94.74(1.05)	88.95(1.69)	83.16(2.17)	80.00(2.06)	74.47(2.86)	67.89(2.92)
Difference	0.52	2.37	1.58	1.58	6.58	0	3.42
Card Games							
Pre	98.33(0.94)	95.00(1.74)	91.67(2.16)	79.17(4.39)	79.58(4.58)	74.17(2.88)	65.00(2.46)
Post	97.50(0.97)	93.75(2.05)	88.33(2.78)	82.08(3.77)	76.67(3.96)	75.00(3.43)	68.33(3.71)
Difference	−0.83	−1.25	−3.34	2.91	−2.91	0.83	3.33

Table A4
Mean accuracy (standard error) before and after training (pre vs. post) for each training group.

	UT	AA	HP	TE	IM	CG
Pre	80.52(1.70)	81.58(1.63)	79.44(1.61)	78.83(1.66)	81.58(1.34)	83.27(1.65)
Post	85.13(1.70)	82.70(1.60)	80.92(1.60)	82.09(1.43)	83.87(1.16)	83.10(1.57)
Difference	4.61	1.12	1.48	3.26	2.29	.17

on four and five cued-object trials indicated that only the action group improved significantly more that the baseline group (pre vs. post-test by training group interaction) (UT vs. CG, $F(1, 21) = 6.60$, $\text{MSE} = 0.008$, $p = 0.02$, $\eta_p^2 = 0.24$ with pre-test vs. post-test % accuracy of 75.00 ± 2.02 vs. 84.77 ± 2.30 for UT and 79.38 ± 3.10 vs. 79.38 ± 2.73 for CG). None of the other training groups showed a significantly greater improvement on object tracking performance when compared to our baseline group (see Figure 2). On these critical trials, the action game trainees (UT) also outperformed subjects trained on Harry Potter ($F(1, 23) = 4.46$, $\text{MSE} = 0.009$, $p = 0.046$, $\eta_p^2 = 0.16$). The greater improvement of the action group also approached significance when compared to that of the America's Army group ($F(1, 23) = 3.53$, $\text{MSE} = 0.011$, $p = 0.073$, $\eta_p^2 = 0.13$), but less so when compared to the Interactive Metronome group ($F(1, 28) = 2.12$, $\text{MSE} = 0.011$, $p = 0.157$, $\eta_p^2 = 0.07$).

Surprisingly, there was no pre–post by group interaction between the action group and the Tetris-playing controls ($F(1, 23) = 1.09$, $\text{MSE} = 0.009$, $p = 0.31$, $\eta_p^2 = 0.05$). However, there is a consistent trend for greater improvement in the action group, as would be expected based on our previous work which employed nearly three times as much training (Green & Bavelier, 2006b).

References

Ahissar, M. & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America, 90*, 12, 5718–5722.

Ahissar, M. & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*, 401–406.

Ahissar, M. & Hochstein, S. (2000). The spread of attention and learning in feature search: Effects of target distribution and task difficulty. *Vision Research, 40*, 1349–1364.

Ball, K. & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research, 27*, 953–965.

Benson, P. G. (2001). Hawthorne effect. In W.E. Craighead. & C. B. Nemeroff (Eds), *The Corsini encyclopedia of psychology and behavioral science* (3rd ed., Vol. 2, pp. 667–668). New York: John Wiley and Sons.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*, 68–79.

Chun, M. M. & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology – Human Perception and Performance*, *21*, 109–127.

Clark, J. E., Lanphear, A. K., & Riddick, C. (1987). The effects of videogame playing on the response selection processing of elderly adults. *Journal of Gerentology, 42*(1), 82–85.

Donchin, E., Fabiani, M., & Sanders, A. (1989). The learning strategies program: An examination of the strategies in skill acquisition. *Acta Psychologica, 71*, 1–311.

Dorval, M. & Pepin, M. (1986). Effect of playing a video game on a measure of spatial visualization. *Perceptual Motor Skills, 62*, 159–162.

Drew, B. & Waters, J. (1986) Video games: Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. *Cognitive Rehabilitation, 4*, 26–31.

Fiorentini, A. & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature, 287*, 43–44.

Fiser, J. & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*(6), 499–504.

Gagnon, D. (1985). Videogames and spatial skills: An exploratory study. *Educational Communication and Technology Journal, 33*, 263–275.

Goldstein, J., Cajko, L., Oosterbroek, M., Michielsen, M., van Houten, O., & Salverda, F. (1997) Video games and the elderly. *Social Behavior and Personality, 25*(4), 345–352.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*(3), 387–405.

Green, C. S. & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534–537.

Green, C. S. & Bavelier, D. (2006a). The cognitive neuroscience of video games. In P. Messaris & L. Humphreys (Eds), *Digital Media: Transformations in Human Communication*. New York, Peter Lang.

Green, C. S. & Bavelier, D. (2006b). Enumeration versus multiple-object tracking: The case of action video game players. *Cognition*, *101*, 217–245.

Green, C. S. & Bavelier, D. (2006c). Effect of action video games on the spatial distribution of visuo-spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1465–78.

Green, C. S. & Bavelier, D. (2007). Action video game experience alters the spatial resolution of vision. *Psychological Science*, *18*(1), 88–94.

Greenfield, P. M., DeWinstanley, P., Kilpatrick, H., & Kaye, D. (1994). Action video games and informal education: Effects on strategies for dividing visual attention. *Journal of Applied Developmental Psychology, 15*, 105–123.

Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in eye–hand motor coordination of video-game users and non-users. *Perceptual and Motor Skills, 57*, 155–158.

Kida, M., Oda, S., & Matsumura, M. (2005). Intensive baseball practice improves the Go/Nogo reaction time, but not the simple reaction time. *Cognitive Brain Research, 22*, 257–264.

Kosel, M. & Fish, M. (designers), Fish, M., Bank, L., & Grubbs, E. (programmers) (1983). *The factory*. New York: Sunburst Communications.

Lum, J., Enns, J. T., & Pratt, J. (2002). Visual orienting in college athletes: Explorations of athlete type and gender. *Research Quarterly for Exercise and Sport, 73*(2), 156–167.

McClurg, P. A. & Chaille, C. (1987). Computer games: Environments for developing spatial cognition. *Journal of Educational Computing Research, 3*(1), 95–111.

Moroni, C., Boucart, M., Humphreys, G. W., Henaff, M. A. & Belin, C. (2000). Failure to identify the target does not prevent the attentional blink effect. *Neuroreport, 11*(12), 2775–2780.

Orosy-Fildes, C. & Allan, R. W. (1989). Psychology of computer use: XII. Videogame play: Human RT to visual stimuli. *Perceptual and Motor Skills, 69*, 243–247.

Pylyshyn, Z. W. & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.

Ramachandran, V. S. & Braddick, O. (1973). Orientation-specific learning in stereopsis. *Perception, 2*, 371–376.

Raymond, J. F., Shapiro, K. L., & Arnell, K. M. (1992) Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849–860.

Rosenberg, B. H., Landsittel, D., & Averch, T. D. (2005). Can video games be used to predict or improve laparoscopic skills? *Journal of Endourology, 19*(3), 372–376.

Schaffer, R. J., Jacokes, L. E. Cassily, J. F. et al. (2001). Effect of Interactive Metronome ® training on children with ADHD. *The American Journal of Occupational Therapy, 55*, 155–162.

Skinner, B. (1953). *Science and human behavior*. New York: MacMillan.

Shui, L. P. & Pashler, H. (1992). Improvement in line orientation discrimination in retinally local but dependent on cognitive set. *Perceptual Psychophysiology, 52*, 582–588.

Subrahmanyam, K. & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology, 15*, 13–32.

Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The "Catch the Spies" task. *Cognitive Development*, *20*, 373–387.

Wanatabe, T., Náñez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848.

Yuji, H. (1996). Computer games and information processing skills. *Perceptual and Motor Skills, 83*, 643–647.

This page intentionally left blank

THE ROLE OF VISUAL MAPS TO STIMULATE LEARNING IN A COMPUTER GAME

Richard Wainess and Harold F. O'Neil
University of Southern California

Abstract

Research suggests it is the quality instructional methods embedded into a game, not the game itself, that will determine the game's effectiveness as a teaching or problem-solving tool. Navigation maps are a common instructional method to aid in navigation of complex game environments. This chapter examines the use of navigation maps in a variety of environments and conditions. The overall finding is that navigation maps do not aide in learning or problem-solving in all situations. Other conditions, such as task, performance objectives, other instructional methods, or issues related to cognitive load, may override the potential effectiveness of navigation maps.

Despite early expectations, research into the effectiveness of games and simulations as educational media has been met with mixed reviews (de Jong & van Joolingen, 1998; Garris, Ahlers, & Driskell, 2002; O'Neil, Wainess, & Baker, 2005). It has been suggested that the lack of consensus can be attributed to weaknesses in *instructional strategies* embedded in the media and to other issues related to *cognitive load* (Chalmers, 2003; Cutmore, Hine, Maberly, Langford, & Hawgood, 2000; Lee, 1999; Thiagarajan, 1998; Wolfe, 1997). Cognitive load refers to the amount of mental activity imposed on working memory at an instance in time (Chalmers, 2003; Sweller & Chandler, 1994, Yeung, 1999). Researchers have proposed that *working memory* limitations can have an adverse effect on learning (Sweller & Chandler, 1994; Yeung, 1999). Further, *cognitive load theory* suggests that learning involves the development of *schemas* (Atkinson, Derry, Renkl, & Wortham, 2000), a process constrained by limited working memory and separate channels for auditory and visual/spatial stimuli (Brunken, Plass, & Leutner, 2003). Cognitive load theory also describes an unlimited capacity, *long-term memory* that can store vast numbers of schemas (Mousavi, Low, & Sweller, 1995).

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

The inclusion of *scaffolding*, which provides support during schema development by reducing the load in working memory, is a form of *instructional design*; more specifically, it is an *instructional strategy* (Allen, 1997; Clark, 2001). For example, graphical scaffolding, which involves the use of imagery-based aids, has been shown to provide effective support for graphically based learning environments, including video games (Benbasat & Todd, 1993; Farrell & Moore, 2000; Mayer, Mautone, & Prothero, 2002). Navigation maps, a particular form of graphical scaffolding, have been shown to be an effective scaffold for navigation of a three-dimensional (3D) virtual environment (Cutmore et al., 2000). Navigation maps have also been shown to be an effective support for navigating and problem solving in a 2D hypermedia environment (Baylor, 2001; Chou, Lin, & Sun, 2000), which is comprised of *nodes* of information and *links* between the various nodes (Barab, Bowdish, & Lawless, 1997). In contrast, however, Wainess (2007) found no benefit from the use of navigation maps to assist in complex problem-solving tasks in a 3D video game.

A virtual environment creates a number of issues with regards to learning. Problem solving within a virtual environment involves not only the cognitive load associated with the to-be-learned material, referred to as *intrinsic cognitive load* (Paas, Tuovinen, Tabbers, & Van Gerven, 2003), it also includes cognitive load related to the visual nature of the environment, referred to as *extraneous cognitive load* (Brunken et al., 2003; Harp & Mayer, 1998), as well as navigation within the environment – either *germane cognitive load* or extraneous cognitive load, depending on the relationship of the navigation to the learning task (Renkl & Atkinson, 2003); It is germane cognitive load if navigation is a necessary component for learning; that is, it is an instructional strategy. It is extraneous cognitive load if navigation does not, of itself, support the learning process; that is, it is included as a feature extraneous to content understanding and learning. An important goal of instructional design within immersive environments involves determining methods for reducing the extraneous cognitive load and/or germane cognitive load, thereby providing more working memory capacity for intrinsic cognitive load (Brunken et al., 2003). This chapter examines the reduction of cognitive load through the use of graphical scaffolding in the form of visual maps and whether this instructional strategy results in better performance outcomes as reflected by *retention* and *transfer* (Paas et al., 2003). Retention refers to the storage and retrieval of knowledge and facts (Day, Arthur, & Gettman, 2001). Transfer refers to the application of acquired knowledge and skills to new situations (Brunken et al., 2003).

Research has examined the use of navigation maps, a particular form of graphical scaffolding, as navigational support for complex problem solving tasks within a hypermedia environment, where the navigation map provided an overview of the 2D, textually-based world which had been segmented into chunks of information, or nodes (Chou et al., 2000). Research has also examined the use of navigation maps as a navigational tool in 3D virtual environments. Studies involving 3D environments have examined either the effect of a navigation map on navigation within an occluded environment (where vision is blocked by walls, trees, buildings, etc.) with the singular goal of getting from point A to point B (Cutmore et al., 2000) or on navigation in a maze-like environment (hallways) that included a simple problem solving task; finding a key along the path in order to open a door at the end of the path (Galimberti, Ignazi, Vercesi, & Riva, 2001).

Wainess (2007) examined the effect of use of 2D topological maps (i.e., a floor plan) on navigation within a 3D video game environment in relationship to a complex problem-solving task. Other studies have examined the use of 2D maps (i.e., a 2D site map) to aid in navigation of the various nodes for complex problem-solving tasks (e.g., Chou & Lin, 1998). It is argued here that the role of the two navigation map types (2D site map and 2D topological floor plan) serve the same purpose in terms of cognitive load, which is, they reduce cognitive load by distributing some of load normally placed in working memory to an external aid, the navigation map. In other words, information (the structure of the environment) that would normally have been held in working memory is offloaded to an accessible, external map of the environment. However, it is also argued here that the spatial aspects of the two learning environments differ substantially. A larger cognitive load is placed on the visual/spatial channel of working memory with a 3D video game environment as compared to a 2D hypermedia environment, due to the more complex visual requirements of working within a 3D world as compared to a 2D world, thereby leaving less working memory capacity in the 3D video game for visual stimuli; the navigation map. Therefore, the cognitive load benefits of map usage in a 3D environment may not be as great as the cognitive load benefits of map usage in a 2D environment, particularly if the map is spatially separated from the main environment – the video game – a condition which adds cognitive load (Mayer & Moreno, 1998).

As immersive 3D video games have become more widespread as commercial entertainment, interest has also grown for the utilization of 3D video games as educational media, particularly because of the perceived motivational aspects of video games for engaging students (O'Neil & Fisher, 2004). According to Pintrich and Schunk (2002), *motivation* is "the process whereby goal-directed activity is instigated and sustained" (p. 405). Tennyson and Breuer (2002) contended that *motivation* influences both attention and maintenance processes, generating the *mental effort* that drives us to apply our knowledge and skills. Salomon (1983) described *mental effort* as the depth or thoughtfulness a learner invests in processing material. Therefore, the role of navigation maps to reduce the load induced by navigation and, thereby, reduce burdens on working memory is an important issue with regard to enhancing the effectiveness of video games as educational environments.

1. Cognitive Load Theory

In recent years, a number of theories have emerged which provide plausible yet competitive explanations of how humans acquire and use knowledge. These theories, which tend to overlap or augment one another, include fuzzy trace theory (Reyna & Brainerd, 1995, 1998), evolutionary psychology (Geary, 2002; Sweller, 2006), a hierarchical view of cognitive abilities (Stankov, 2000), short-term and long-term working memory (Ericsson & Kintsch, 1995), higher order thought (HOT; Rosenthal, 2000a, b), theory of mind

(Flavell, 2004; Leslie, 1987), adaptive character of thought theory (ACT-R, Anderson, 1996), and cognitive load theory (Sweller & Chandler, 1994). This chapter focuses on cognitive load theory.

Cognitive Load Theory is based on the assumptions of a limited working memory with separate channels for auditory and visual/spatial stimuli, and a virtually unlimited capacity long-term memory that stores schemas of varying complexity and levels of automation (Brunken et al., 2003). According to Paas et al. (2003), cognitive load refers to the amount of load placed on working memory. Working memory refers to the limited capacity for holding and processing chunks of information. Brunning, Schraw, and Ronning (1999) defined a chunk as any stimulus that is used, such as a letter, number, or word. According to Miller (1956) working memory capacity varies from five to nine chunks of information. More recently, Paas et al. (2003) argued that working memory can only handle two or three "novel" chunks of information.

Long-term memory has an unlimited permanent capacity (Tennyson & Breuer, 2002), and can contain vast amounts of schemas. Schemas are cognitive constructs that incorporate multiple elements of information into a single element with a specific function (Paas et al., 2003). Schemas have the functions of storing information in long-term memory and of reducing working memory load by permitting people to treat multiple elements of information as a single element or chunk (Kalyuga, Chandler, & Sweller, 1998; Mousavi et al., 1995).

There are three types of cognitive load that can be defined in relationship to a learning or problem-solving task: *intrinsic cognitive load, germane cognitive load*, and *extraneous cognitive load* (Ayres, 2006). Intrinsic cognitive load refers to the cognitive load placed on working memory by the to-be-learned material (Paas et al., 2003). Germane cognitive load refers to the cognitive load required to access and process the intrinsic cognitive load; For example, the problem-solving processes that are instantiated in the learning process so that learning can occur (Renkl & Atkinson, 2003). Extraneous cognitive load refers to the cognitive load imposed by stimuli that neither support the learning process (i.e., germane cognitive load) nor are part of the to-be-learned material (i.e., intrinsic cognitive load). An important goal of instructional design is to balance intrinsic, germane, and extraneous cognitive loads to support learning outcomes and to recognize that the specific balance is dependent on a number of factors (Brunken et al., 2003), including the need for motivation.

2. Problem Solving

Many tasks involve *problem solving*. Problem solving is "cognitive processing directed at transforming a given situation into a desired situation when no obvious method of solution is available to the problem solver" (Baker & Mayer, 1999, p. 272). According to Kirschner, Sweller, and Clark (2006), problem solving "places a huge burden on working memory" (p. 77). The O'Neil Problem-Solving model (Figure 1; see O'Neil, 1999) defines three core constructs of problem solving: *content understanding, problem-solving strategies*,

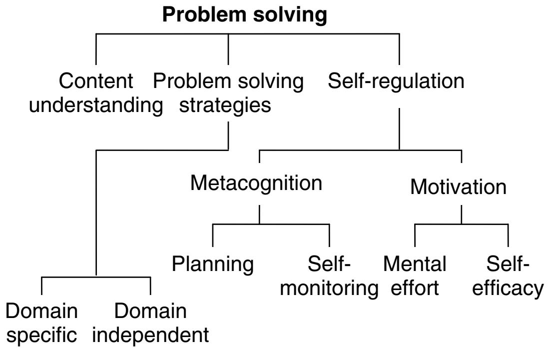

Figure 1. O'Neil Problem-solving model

and *self-regulation*. Content understanding refers to domain knowledge. Problem-solving strategies refer to both domain-specific and domain-independent strategies. Self-regulation is comprised of metacognition (planning and self-monitoring) and motivation (mental effort and self-efficacy; O'Neil, 1999, 2002).

3. Games, Simulations, and Simulation-Games

Computer-based educational games fall into three categories: games, simulations, and simulation games. While there is debate as to the specific characteristics of each of these three media (e.g., Betz, 1995–1996; Crookall & Arai, 1995; Crookall, Oxford, & Saunders, 1987; Dempsey, Haynes, Lucassen, & Casey, 2002; Duke, 1995; Garris et al., 2002; Randel, Morris, Wetzel, & Whitehill, 1992; Ricci, Salas, & Cannon-Bowers, 1996), Gredler (1996) provided definitions for the three media that offer some clear delineations among them. According to Gredler, games consist of rules, can contain imaginative contexts, are primarily linear, and include goals as well as competition, either against other players or against a computer (Gredler, 1996). Simulations display the dynamic relationship among variables which change over time and reflect authentic causal processes. Simulations are non-linear and have a goal of discovering causal relationships through manipulation of independent variables. Simulation games are a blend of games and simulations (Gredler, 1996). Wainess (2007) added an important modification to Gredler's definitions.

When Gredler (1996) described games a linear and simulations as non-linear, she was referring to their goal structures – games have linear goal structures and simulations have non-linear goal structures. According to Wainess (2007), there is another aspect of games or simulation interaction that can be described as either linear or non-linear – the intervention structure of the media. Intervention refers to the actions players or users are allowed to take at any given moment of the game or simulation. In almost all instances of intervention, both games and simulations give at least two choices (e.g., quit or continue, turn left or turn right, fight or run, increase something or decrease it). Therefore, for both games and simulations, the intervention structure is non-linear.

There has also been disagreement as to the definition of video games and, more importantly, whether a computer-based game is a video game. However, the definitions of computer-based game and video game are beginning to coincide and the two terms are beginning to be used interchangeably (e.g., Greenfield et al., 1996; Greenfield, deWinstanley, Kilpatrick, & Kaye, 1996; Kirriemuir, 2002; Okagaki & Frensch, 1994). In this chapter, we will also use the terms computer-based game and video game interchangeably.

4. Educational Benefits of Games and Simulations

While numerous studies have cited the learning benefits of games and simulations (e.g., Adams, 1998; Baker, Prince, Shrestha, Oser, & Salas, 1993; Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Betz, 1995–1996; Khoo & Koh, 1998; Lainema & Nurmi, 2006), others have found mixed, negative, or null outcomes from games and simulations, specifically in the relationship of enjoyment of a game to learning from the game (e.g., Brougere, 1999; Dekkers & Donatti, 1981; Druckman, 1995; O'Neil et al., 2005). One of the problems appears to be that many non-empirical studies claim learning outcomes that cannot be substantiated by their own data (see O'Neil et al., 2005 for a discussion of this issue), as well as empirical studies making claims not supported by the data (see O'Neil et al., 2005). Another problem seems to be the paucity of empirical studies.

Of several thousand articles on game and simulation studies with adult subjects published in peer reviewed journals in the last 15 years, only 19 were empirical (see O'Neil et al., 2005). Of those 19 studies, two compared game-based training to other methods of instruction; one study found in that game-based training resulted in higher performance on retention tests (Ricci et al., 1996) while the other found it resulted in lower performance (Parchman, Ellis, Christinaz, & Vogel, 2000). Another of the studies found that playing a game that fostered certain generalizable skills lead to greater performance on a far transfer task (jet fighter training) as compared to those who did not play the game (Gopher, Weil, & Bareket, 1994). The remaining 16 studies compared game performance based on differing instructional strategies or due to individual differences. In general, those studies found that higher visual abilities resulted in better game performance and prior game playing experience resulted in better game playing performance. A few found that instructional method affected game performance.

Another issue in claims attributed to games or simulations is the inaccurate use of media definition. For examples, the medium used by Ricci et al. (1996) was defined by the researchers as a game, but the description of the medium met the criteria for a simulation game. Therefore, any outcomes attributed to the use of games would have been inaccurate.

Another claim related to the proposed educational benefit of games and simulations is their motivational characteristics. The assumption is that motivation always leads to learning. However, a number of researchers suggest otherwise (e.g., Brougere, 1999; Druckman, 1995; Salas, Bowers, & Rhodenizer, 1998). Salomon (1983) even contended that a positive attitude can actually indicate less learning. Dekkers and Donatti (1981) found that motivation wanes over time, as the novelty of the game or simulation subsides. While these various arguments potentially explain the mixed findings with regards to the learning outcomes in games and simulation research, there is another argument which may provide a better explanation of the mixed findings.

There appears to be consensus among a large number of researchers that the negative, mixed, or null findings might be related to a lack of sound instructional design embedded in the games (de Jong & van Joolingen, 1998; Garris et al., 2002; Gredler, 1996; Lee, 1999; Leemkuil, de Jong, de Hoog, & Christoph, 2003; Thiagarajan, 1998; Wolfe, 1997). These researchers suggest that it is the instructional design embedded in a medium and not the medium itself that leads to learning. Instructional design involves the implementation of various instructional strategies and methods.

5. Scaffolding

As discussed earlier, cognitive load theory (Paas et al., 2003) is concerned with methods for reducing the amount of cognitive load placed on working memory during learning and problem-solving activities. Clark (2003) commented that instructional messages (an instructional method) must also keep the cognitive load from instructional presentations to a minimum. Scaffolding is considered a viable instructional method that assists in cognitive load reduction. There are a number of definitions for *scaffolding* in the literature. Chalmers (2003) defined scaffolding as the process of forming and building upon a schema. In a related definition, Van Merriënboer, Kirschner, and Kester (2003) defined the original meaning of scaffolding as all devices or strategies that support students' learning. More recently, Van Merriënboer, Clark, and de Croock (2002) defined scaffolding as the process of diminishing (fading) support as learners acquire more expertise. Allen (1997) defined scaffolding as the process of training a student on core concepts and then gradually expanding the training. To summarize, the four definitions of scaffolding involve the development of simple to complex schema, all devices that support learning, the process of diminishing (fading) support during learning, and the process of building learning from basic concepts to complex knowledge, respectively. Ultimately, the core principle embodied in each of these definitions is that scaffolding is concerned with controlling the amount of cognitive load imposed by learning, and each reflects a philosophy or approach to controlling or reducing that load. In this chapter, all four definitions of scaffolding will be considered.

As defined by Clark (2001), instructional methods are external representations of internal cognitive processes that are necessary for learning but which learners cannot, may not, or will not provide for themselves. Instructional methods provide learning goals (e.g., demonstrations, simulations, and analogies: Alessi, 2000; Clark 2001), monitoring (e.g., practice exercises: Clark, 2001), feedback (Alessi, 2000; Clark, 2001; Leemkuil et al., 2003; Sitzmann, Kraiger, Stewart, & Wisher, 2006), and selection (e.g., highlighting information: Alessi, 2000; Clark, 2001). Alessi (2000) added that instructional methods include: giving hints and prompts before student actions; providing coaching, advice, or help systems; and providing dictionaries and glossaries. Jones, Farquhar, and Surry (1995) added advance organizers, graphical representations of problems, and hierarchical knowledge structures to the list of instructional methods. Each of these examples is a form of scaffolding.

In learning by doing in a virtual environment, students can actively work in realistic situations that simulate authentic tasks for a particular domain (Mayer et al., 2002). A major instructional issue in learning by doing within simulated environments concerns the proper type of guidance (i.e., scaffolding), that is, how best to create *cognitive apprenticeship* (Mayer et al., 2002). Mayer and colleagues (2002) also commented that their research shows that discovery-based learning environments can be converted into productive venues for learning when appropriate cognitive scaffolding is provided; specifically, when the nature of the scaffolding is aligned with the nature of the task, such as pictorial scaffolding for pictorially based tasks and textually based scaffolding for textually-based tasks.

5.1. Graphical Scaffolding

According to Allen (1997), selection of appropriate text and graphics can aid the development of mental models, and Jones et al. (1995) commented that visual cues such as *maps* and *menus* as *advance organizers* help learners conceptualize the organization of the information in a program. A number of researchers support the use of maps as visual aids and organizers (Benbasat & Todd, 1993; Chou & Lin, 1998; Chou et al., 2000; Farrell & Moore, 2000; Ruddle, Howes, Payne, & Jones, 2000). Chalmers (2003) defined *graphic organizers* as organizers of information in a graphic format, which act as spatial displays of information that can also act as study aids. Jones et al. (1995) argued that interactive designers should provide users with visual or verbal cues to help them navigate through unfamiliar territory.

5.2. Navigation Maps

Cutmore et al. (2000) defined *navigation* as "... a process of tracking one's position in a physical environment to arrive at a desired destination" (p. 224). A route through the environment consists of either a series of locations or continuous movement along a path. Cutmore et al. further commented that "Navigation becomes problematic when the whole path cannot be viewed at once but is largely occluded by objects in the environment" (p. 224). The occluding objects may include internal walls or large environmental features such as trees, hills, or buildings. Under these conditions, one cannot simply plot a direct visual course from the start to finish locations. Rather, knowledge of the layout of the space is required. Navigation maps or other descriptive information may provide that knowledge (Cutmore et al., 2000).

Effective navigation of a familiar environment depends upon a number of cognitive factors. These include working memory for recent information, attention to important cues for location, bearing and motion, and finally, a cognitive representation of the environment which becomes part of a long-term memory, a cognitive map (Cutmore et al., 2000). A 2D navigation map can help users to navigate and orient themselves, and may facilitate an easier learning experience.

A number of experiments have examined the use of navigation maps in virtual environments. Using a web-based hypermedia environment, Chou and Lin (1998) and Chou et al. (2000) examined various navigation map types, with some navigation maps offering global views of the environment (global navigation map) and others offering more localized views (local navigation map), based on the learner's current location. In the Chou and Lin (1998) study, 121 college students were divided into five groups, based on four navigation map variations; (1) a global map of the entire 94 node hierarchical knowledge structure (the entire hypermedia environment), (2) a series of local maps for each knowledge area of the environment, (3) a tracking map that updated according the participant's location with the participant's location always in the center and showing one level of nodes above and two below the current position, (4) a no-map situation, and (5) an all-map situation (global, local, and tracking). After being given instruction on their respective navigation tools and time to practice, subjects were given 10 search tasks and an additional 30 min to browse the hypermedia environment, after which they answered posttest questions and an attitude questionnaire and created a knowledge map.

Results of the Chou and Lin (1998) study indicated that the search efficiency (search speed) for the all map and global map groups were significantly faster than for the other three groups (local map, tracking map, and no map), indicating benefits from using the global map or all maps (which included the global map). There was not significance between the all map and global map groups. Knowledge map creation for the all map and global map groups were also significantly higher than for the tracking map group, but not significantly higher than the local map or no map groups. Overall, the results of the Chou and Lin (1998) study suggest that use of a global map or use of a combination of maps, including the global map, results in greater search efficiency but not greater content understanding (as indicated by knowledge map development) then either local maps or no map. Additionally, there were no differences found between the use of a local map versus no map, suggesting no cognitive value to a local map. With regard to attitude, there were no differences by map type for any of the attitudinal scales, including attitude toward the learning experience, usability of the system, and disorientation.

As with Chou and Lin (1998) study, the Chou et al. (2000) study, which involved over 100 college students, showed that the type of navigation map used affected performance. However, some findings were in contrast to the earlier study. With regards to development of a knowledge map, the no map group's performance was significantly higher than the local map's performance. The global map group's score almost reached significance over the local map group ($p = 0.057$). There was no difference between the global map scores and the no map scores. This differed from the earlier study which found a significant difference between the global map over no map, suggesting that, for the tasks involved in one or both of these studies, map use might not have been a primary factor in developing content understanding.

Results of the Chou et al (2000) study indicated that map type can affect performance in a search task. A global map resulted in better performance than a local map or no map with regards to navigation (search speed and revisiting sites), while performance on knowledge map creation by the no map group was significantly better than for those who used a local map and only slightly better than for those who used a global map. In other words, accomplishment of a problem-solving task was best with a global map while understanding of a problem-solving task or environment was best with either no map or with a global map. Results of the two Chou and colleague studies (Chou & Lin 1998; Chou et al., 2000) suggest that global map use can improve search speed and reduce revisiting locations. The mixed results of the two studies suggest that map use may not influence content understanding.

According to Tkacz (1998), soldiers use navigation maps as tools, which involve spatial reasoning, complex decision making, symbol interpretation, and spatial problem solving. In her study involving 105 marines, Tkacz examined the procedural components of cognitive maps required for using and understanding topographic navigation maps, stating that navigation map interpretation involves both top-down (retrieved from long-term memory) and bottom-up (retrieved from the environment and the navigation map) procedures. Therefore, Tkacz examined the cognitive components underlying navigation map interpretation to assess the influence of individual differences on course success and on real-world position location. In addition, Tkacz related position location ability to video game performance in a simulated environment. Performance measures consisted of real-world position location (in the field), a map reading readiness exercise (using a map), and simulated travel in a videogame environment (a 3D maze, with movement in six directions; North, South, East, West, Up, and Down). The goal of the maze game was for the player to move as quickly as possible through the 125 room structure to a goal room and open the exit door.

All participants completed a map reading pretest, after which the treatment group received 15 hours of geographical training. After that training, all groups completed spatial tests and a geography test. Additional participant data was obtained from armed services vocational aptitude tests that had been administered to each participant at enlistment.

According to Tkacz (1998), the geographical instruction significantly improved the ability to perform terrain association and relate the real-world scenes to topographical map representations. Results of the study also indicated that orientation and, to a lesser extent, reasoning ability are important for map interpretation. Video game performance was affected by all spatial skills, but particularly by orientation and mental rotation (visualization), with high ability subjects escaping the maze faster than lower ability subjects. Video game performance was also affected by map reading ability, with better performance by those demonstrating better map reading performance.

Mayer et al. (2002) commented that a major instructional issue in learning by doing within simulated environments concerns the proper type of guidance, which they refer to as cognitive apprenticeship. The investigators used a geological gaming simulation, the Profile Game, to test various types of guidance structures (i.e., strategy modeling), ranging from no guidance to illustrations (i.e., pictorial aids) to verbal descriptions to pictorial and verbal aids combined. The Profile Game is based on the premise, "Suppose you were visiting a planet and you wanted to determine which geological feature is present on a certain portion of the planet's surface" (Mayer et al., 2002, p. 171). While exploring, you cannot directly see the features, so you must interpret data indirectly, through probing procedures. The experimenters focused on the amount and type of guidance needed within the highly spatial simulation.

Though a series of experiments, Mayer et al. (2002) found that pictorial scaffolding, as opposed to verbal scaffolding, is needed to enhance performance in a visual-spatial task. In the final experiments of the series, participants were divided into verbal scaffolding, pictorial scaffolding, both, and no scaffolding groups. Participants who received pictorial scaffolding solved significantly more problems than did those who did not receive pictorial scaffolding. Students who received strategic scaffolding did not solve significantly more problems than students who did not receive strategic scaffolding. While high-spatial participants performed significantly better than low-spatial students, adding pictorial scaffolding to the learning materials helped both low- and high-spatial students learn to use the Profile Game. Students in the pictorial-scaffolding group correctly solved more transfer problems than students in the control group. However, pictorial scaffolding did not significantly affect the solution time (speed) of either low- or high-spatial participants. Overall, adding pictorial scaffolding to the learning materials lead to improved performance on a transfer task for both high- and low-spatial students in the Profile Game (Mayer et al., 2002).

Wainess (2007) examined the use of navigation maps to assist in problem solving in a video game, SafeCracker® (Daydream Interactive, Inc., 1995/2001). SafeCracker is a non-violent, PC-based game where players must find and open safes, through the use of clues, objects, and in some cases, prior knowledge or trial-and-error. Data was collected from 65 college students (43 male and 22 female). The purpose of the study was to examine the effect of a navigation map on a complex problem-solving task in a 3D, occluded, computer-based video game. With one group playing the video game while using a navigation map (the treatment group) and the other group playing the game without aid of a navigation map (the control group), the study examined differences in problem-solving outcomes as informed by the O'Neil (1999) Problem-Solving model (see Figure 1).

The Wainess (2007) study examined the effect of navigation maps on content understanding and problem-solving strategy retention and transfer, as well as the correlation between trait self-regulation and performance, and between navigation map usage and continuing motivation, as exhibited by continued optional play of the game.

All participants in the Wainess (2007) study were given instruction in use of the knowledge mapping software and the game SafeCracker, including strategy instruction.

The treatment group was then given instruction on using the navigation maps, as well as strategies for effectively navigating the environment. The control group was also given navigation strategies. Participants then played SafeCracker twice. Participants were also given a task completion form at the start of each game, which listed the safes to be opened and the rooms the safes were in, which acted as an advance organizer. After each game, participants created a knowledge map and responded to problem-solving strategy retention and transfer questions. At the conclusion of the study, participants were offered an opportunity to play the game for an extra half hour (to examine continuing motivation).

Results of the Wainess (2007) study indicated that the use of navigation maps did not affect content understanding, problem-solving strategy retention, or problem-solving strategy transfer. In addition, higher levels of self-regulation were unrelated to higher levels of performance regardless of whether or not a map was used, except for a positive correlation between amount of mental effort and the amount of improvement in the problem-solving strategy retention scores for the navigation map group, $r = 0.38$, $p < 0.05$. Lastly, those who used the navigation map (the treatment group) did not exhibit higher continuing motivation than those who did not use the map (the control group).

Two directions were considered to explain the lack of results in the Wainess (2007) study; one that would reduce or suppress the effects of the treatment and one that would inflate the outcomes measures for the control group. Suppression of outcomes by the treatment group might well be explained by the contiguity effect and by extraneous cognitive load. Inflation of the control group might well be explained by priming.

The *contiguity effect*, which refers to the cognitive load imposed when multiple sources of information are separated (Mayer, Moreno, Boire, & Vagge, 1999; Mayer & Moreno, 2003; Mayer & Sims, 1994; Moreno & Mayer, 1999). Spatial contiguity occurs when information is physically separated (Mayer & Moreno, 2003). The contiguity effect results in the *split attention effect*, where the process of integrating spatially (or temporally) separated information may place an unnecessary strain on limited working memory resources, resulting in impairment in learning (Atkinson et al., 2000; Mayer & Moreno, 1998; Tarmizi & Sweller, 1988). The study by Wainess (2007) imposed spatial contiguity, since the navigation map was presented on a piece of paper while SafeCracker was viewed on a computer monitor.

Extraneous cognitive load refers to the cognitive load imposed by unnecessary (extraneous) materials (Harp & Mayer, 1998; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2000; Renkl & Atkinson, 2003; Schraw, 1998) and can affect both retention and transfer (Moreno & Mayer, 2000). The navigation map in the Wainess (2007) study may have been an extraneous detail. The game environment may not have been complex enough to benefit from use of a navigation map. Therefore, adding the map would have added extraneous cognitive load for the treatment group, negating any expected benefits from the map.

Priming is a cognitive phenomenon where a stimulus (e.g., a word or sound) readies the mind to allow or engage particular relevant schema. Timely exposure to stimuli results in enhanced access to stored stimuli or information (retrieved October 7, 2005 from http:// filebox.vt.edu/8080/users/dereese2/module8/module08bkup/IDProjectWebpage/lesson4.htm). Priming asserts that providing cues can help focus attention on important tasks or details, which ultimately helps with metacognitive process involved in learning and problem solving. Both groups in the Wainess (2007) study were primed a number of times with search and problem-solving strategies, which might have aided both groups in understanding procedures necessary for doing well in the SafeCracker. All participants were primed during knowledge map training, SafeCracker training, navigation map training for the treatment group, navigation training for the control group, and at the start of each game (see Wainess 2007 for detailed descriptions of the priming events). Those strategies may have influenced both groups enough to offset any differences that might have been fostered by navigation map usage, ultimately resulting in similar outcomes for the two groups.

6. Conclusions and Implications

Graphical scaffolding has been shown to provide effective support for graphically based learning environments, including video games (Benbasat & Todd, 1993; Farrell & Moore, 2000; Mayer et al., 2002). Navigation maps, a particular form of graphical scaffolding, have been shown to be an effective scaffold for navigation of a 3D virtual environment (Cutmore et al., 2000). Navigation maps have also been shown to be an effective support for navigating in a problem-solving task in a 2D hypermedia environment (Baylor, 2001; Chou et al., 2000). In contrast, navigation maps were shown to be ineffective for a complex problem-solving task in a 3D, occluded computer-based video game (Wainess, 2007). Results of the Wainess (2007) study indicate that use of a navigation map does not guarantee improvements over not using a navigation map.

More and more, educational institutions seem to be embracing the use of video game and simulation environments as a way of modernizing teaching. The primary impetus for this change in learning strategy is the belief that the motivational aspects of games and simulations will lead to improvements in learning. Yet, as research has shown, it is the quality and appropriateness of the instructional strategies embedded in learning environment that determine whether or not learning will occur. Little is known about the use of immersive 3D games for learning, and this chapter highlights the fact that what works in one learning environment may not work in another. To ensure that 3D games provide the necessary features to foster learning, studies examining instructional strategies that have been previously shown to be effective in other learning environments must be carefully examined for effectiveness in this new learning environment. One such strategy is the use of navigation maps. As a potentially useful instructional strategy, it is important to discover the circumstances under which navigation maps are beneficial to learning. Only then can we begin to prescribe their use.

References

Adams, P. C. (1998, March/April). Teaching and learning with SimCity 2000 [Electronic Version]. *Journal of Geography, 97*(2), 47–55.

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

Allen, R. B. (1997). Mental models and user models. In M. Helander, T. K. Landauer, & P. Prabhu (Eds), *Handbook of human computer interaction: second, completely revised edition* (pp. 49–63). Amsterdam: Elsevier.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, *51*(4), 355–365.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214.

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287–298.

Baker, D., Prince, C., Shrestha, L., Oser, R., & Salas, E. (1993). Aviation computer games for crew resource management training. *The International Journal of Aviation Psychology, 3*(2), 143–156.

Baker, E. L. & Mayer, R. E. (1999). Computer-based assessment of problem solving. *Computers in Human Behavior, 15*, 269–282.

Barab, S. A., Bowdish, B. E., & Lawless, K. A. (1997). Hypermedia navigation: Profiles of hypermedia users. *Educational Technology Research & Development*, *45*(3), 23–41.

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

Baylor, A. L. (2001). Perceived disorientation and incidental learning in a web-based environment: Internal and external factors. *Journal of Educational Multimedia and Hypermedia, 10*(3), 227–251.

Benbasat, I. & Todd, P. (1993). An experimental investigation of interface design alternatives: Icon vs. text and direct manipulation vs. menus. *International Journal of Man-Machine Studies, 38*, 369–402.

Betz, J. A. (1995–1996). Computer games: Increase learning in an interactive multidisciplinary environment. *Journal of Educational Technology Systems, 24*(2), 195–205.

Brougere, G. (1999, June). Some elements relating to children's play and adult simulation/gaming. *Simulation & Gaming, 30*(2), 134–146.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist 38*(1), 53–61.

Brunning, R. H., Schraw, G. J., & Ronning, R. R. (1999). *Cognitive psychology and instruction* (3rd ed.). Upper Saddle River, NJ: Merrill.

Chalmers, P. A. (2003). The role of cognitive theory in human-computer interface. *Computers in Human Behavior, 19*, 593–607.

Chou, C. & Lin, H. (1998). The effect of navigation map types and cognitive styles on learners' performance in a computer-networked hypertext learning system [Electronic Version]. *Journal of Educational Multimedia and Hypermedia, 7*(2/3), 151–176.

Chou, C., Lin, H, & Sun, C.-T. (2000). Navigation maps in hierarchical-structured hypertext courseware [Electronic Version]. *International Journal of Instructional Media, 27*(2), 165–182.

Clark, R. E. (Ed.) (2001). *Learning from media: Arguments, analysis, and evidence*. Greenwich, CT: Information Age Publishing.

Clark, R. E. (2003, February). Strategies based on increasing student motivation: Helping learners invest maximum mental effort. In H. F. O'Neil (Ed.), *What works in distance education. Report to the Office of Naval Research by the National Center for Research on Evaluation, Standards, and Student Testing* (pp. 22–23).

Crookall, D. & Arai, K. (Eds). (1995). *Simulation and gaming across disciplines and cultures: ISAGA at a watershed*. Thousand Oaks, CA: Sage.

Crookall, D., Oxford, R. L., & Saunders, D. (1987). Towards a reconceptualization of simulation. From representation to reality. *Simulation/Games for Learning, 17*, 147–171.

Cutmore, T. R. H., Hine, T. J., Maberly, K. J., Langford, N. M., & Hawgood, G. (2000). Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human-Computer Studies, 53*, 223–249.

Day, E. A., Arthur, W., Jr, & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill. *Journal of Applied Psychology, 86*(5), 1022–1033.

de Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research, 68*, 179–202.

Dekkers, J. & Donatti, S. (1981). The interpretation of research studies on the use of simulation as an instructional strategy. *Journal of Educational Research, 74*(6), 64–79.

Dempsey, J. V., Haynes, L. L., Lucassen, B. A., & Casey, M. S. (2002). Forty simple computer games and what they could mean to educators. *Simulation & Gaming, 43*(2), 157–168.

Druckman, D. (1995). The educational effectiveness of interactive games. In D. Crookall & K. Aria (Eds), *Simulation and gaming across disciplines and cultures: ISAGA at a watershed* (pp. 178–187). Thousand Oaks, CA: Sage

Duke, R. D. (1995, December). Gaming: An emergent discipline. *Simulation & Gaming, Silver Anniversary Issue* (Part 4), 426–438.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245.

Farrell, I. H. & Moore, D. M. (2000). The effect of navigation tools on learners' achievement and attitude in a hypermedia environment. *Journal of Educational Technology Systems, 29*(2), 169–181.

Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly, 50*(3), 274–290.

Galimberti, C., Ignazi, S., Vercesi, P., & Riva, G. (2001). Communication and cooperation in networked environment: An experimental analysis. *CyberPsychology & Behavior, 4*(1), 131–146.

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441–467.

Geary, D. C. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences, 12*, 317–345.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*, 387–405.

Gredler, M. E. (1996). Educational games and simulations: a technology in search of a research paradigm. In D. H. Jonassen (Ed.). *Handbook of Research for Educational Communications and Technology*. (pp. 521–540). New York: Simon & Schuster Macmillan.

Greenfield, P. M., Camaioni, L., Ercolani, P., Weiss, L., Lauber, B. A., & Perucchini, P. (1996). Cognitive socialization by computer games in two cultures: Inductive discovery or mastery of an iconic code? In P. M. Greenfield & R. R. Cocking (Eds), *Interacting with Video* (pp. 141–167). Norwood, NJ: Ablex Publishing Corporation.

Greenfield, P. M., deWinstanley, P., Kilpatrick, H., & Kaye, D. (1996). Action video games and informal education: Effects on strategies for dividing visual attention. In P. M. Greenfield & R. R. Cocking (Eds), *Interacting with Video* (pp. 187–205). Norwood, NJ: Ablex Publishing Corporation.

Harp, S. F. & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90(3)*, 414–434.

Jones, M. G., Farquhar, J. D., & Surry, D. W. (1995, July/August). Using metacognitive theories to design user interfaces for computer-based learning. *Educational Technology, 35*(4), 12–22.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40*(1), 1–17.

Khoo, G.-S. & Koh, T.-S. (1998). Using visualization and simulation tools in tertiary science education [Electronic Version]. *The Journal of Computers in Mathematics and Science Teaching, 17*(1), 5–20.

Kirriemuir, J. (2002, February). Video gaming, education, and digital learning technologies: Relevance and opportunities [Electronic Version]. *D-Lib Magazine*, *8*(2), 1–8.

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), 74–86.

Lainema, T. & Nurmi, S. (2006). Applying an authentic, dynamic learning environment in real world business. *Computers in Education, 47*, 94–115.

Lee, J. (1999). Effectiveness of computer-based instructional simulation: A meta analysis. *International Journal of Instructional Media*, *26*(1), 71–85.

Leemkuil, H., de Jong, T., de Hoog, R., & Christoph, N. (2003). KM Quest: A collaborative Internet-based simulation game. *Simulation & Gaming, 34*(1), 89–111.

Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind." *Psychological Review, 94*(4), 412–426.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187–198.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171–185.

Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence of dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*(4), 638–643.

Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389–401.

Miller, G. A. (1956). The magical number, seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Moreno, R. & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Moreno, R. & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case of minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117–125.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334.

O'Neil, H. F., Jr. (1999). Perspectives on computer-based performance assessment of problem solving: Editor's introduction. *Computers in Human Behavior, 15*, 255–268.

O'Neil, H. F., Jr. (2002). Perspective on computer-based assessment of problem solving [Special Issue]. *Computers in Human Behavior, 18*(6), 605–607.

O'Neil, H. F. & Fisher, Y.-C. (2004). A technology to support leader development: Computer games. In D. V. Day, S. J. Zaccaro, & S. M. Halpin (Eds), *Leader development for transforming organizations: Growing leaders for tomorrow* (pp. 99–121). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005, December). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474.

Okagaki, L. & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*, 33–58.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71.

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*, 73–87.

Pintrich, P. R. & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Pearson Education.

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 & Games, 23*, 261–276.

Renkl, A. & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 13–22.

Reyna, V. F. & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*(1), 1–75.

Reyna, V. F. & Brainerd, C. J. (1998). Fuzzy-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology, 71*, 194–209.

Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology, 8*(4), 295–307.

Rosenthal, D. M. (2000a). Consciousness, content, and metacognitive judgments. *Consciousness and Cognition, 9*, 203–214.

Rosenthal, D. M. (2000b). Metacognition and higher-order thoughts. *Consciousness and Cognition, 9*, 231–242.

Ruddle, R. A., Howes, A., Payne, S. J., & Jones, D. M. (2000). The effects of hyperlinks on navigation in virtual environments. *International Journal of Human-Computer Studies, 53*, 551–581.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8*(3), 197–208.

Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychology, 18*(1), 42–50.

Schraw, G. (1998). Processing and recall differences among seductive details. *Journal of Educational Psychology, 90*(1), 3–12.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology, 59*, 623–664.

Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*, 35–51.

Sweller, J. (2006). Discussion of 'emerging topics in cognitive load research: using learner and information characteristics in the design of powerful learning environments.' *Applied Cognitive Psychology, 20*, 353–357.

Sweller, J. & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.

Tarmizi, R. A. & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424–436.

Tennyson, R. D. & Breuer, K. (2002). Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior, 18*(6), 650–668.

Thiagarajan, S. (1998, September/October). The myths and realities of simulations in performance technology. *Educational Technology, 38*(4), 35–41.

Tkacz, S. (1998, September). Learning map interpretation: Skill acquisition and underlying abilities. *Journal of Environmental Psychology, 18*(3), 237–249.

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research & Development, 50*(2), 39–64.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking a load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5–13.

Wainess, R. (2007, April). The Effect of Navigation Maps on Problem Solving Tasks Instantiated in a Computer-Based Video Game. *Dissertation Abstracts International*, *67*(10), 3714A (UMI No. ATT 3237181).

Wolfe, J. (1997, December). The effectiveness of business games in strategic management course work [Electronic Version]. *Simulation & Gaming Special Issue: Teaching Strategic Management, 28(4)*, 360–376.

Yeung, A. S. (1999). Cognitive load and learner expertise: Split attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *Journal of Educational Media, 24*(2), 87–102.

This page intentionally left blank

INQUIRY ON THE ROLE OF CONTEMPORARY CHESS SOFTWARE TO ENRICH HUMAN LEARNING AND COGNITION

William M. Bart and Saahoon Hong
University of Minnesota

Tae Seob Shin
Michigan State University

Abstract

The game of chess provides opportunities to utilize higher reasoning skills and to enhance human learning and education. It involves comprehension of chess game position, identification of candidate moves, and analysis and evaluation of possible sequences of moves. This chapter explores various issues associated with the game of chess and the utility of chess software in educational interventions. Special attention is given to three recent studies that made use of contemporary chess software. Those three studies provided results that lent support to the assertion that chess instruction can lead to salutary cognitive effects among children and adults.

1. Introduction

Video games have become an exciting focus of academic inquiry (e.g., Gedler, 1996; O'Neil & Fisher, 2004; O'Neil, Wainess, & Baker, 2005). Despite their technical sophistication and popularity, examples of contemporary chess software such as Chesssmaster 10th Edition and Fritz 9 that provide video game experiences have not received scholarly attention that they deserve, In support of that assertion, some definitions will be examined.

Games may be defined as consisting of "rules that describe allowable player moves, game constraints and privileges (such as ways of earning extra turns), and penalties for illegal (non-permissible) actions. Further, the rules may be imaginative in that they need not relate to real-world events" (Geder, 1996: 523). The complex game of chess certainly satisfies that definition with its array of well-established rules, game constraints, and

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

severe penalties for illegal actions. Chess is a popular game played worldwide. In fact, chess is viewed by many as "the immortal game" (Shenk, 2006).

Video games, also termed computer games, may be defined as games that "typically require a screen (television, monitor, or liquid crystal display) on which the game is viewed and any of a number of input devices (such as a controller, joystick, keyboard, or keypad) through which one interacts with the game" (O'Neil & Fisher, 2004: 104). Contemporary chess software such as Chessmaster 10th Edition and Fritz 9 satisfies that definition, as such software allows individuals to play artificial opponents on practically any computer. Chess is a visual game and, as a result, contemporary chess software may be viewed as providing a video game experience.

With contemporary chess software being very widely used, one may inquire as to the psychological and educational correlates and predictors of skill at the game of chess. This chapter will explore, in part, the utility and the role of contemporary chess software in the context of human learning and education.

To many, chess software may be viewed as "edutainment that is defined in the Merriam-Webster Online Dictionary (2007) as "entertainment (as by games, films, or shows) that is designed to be educational." To members of the chess community including tournament chess players, contemporary chess software provides instruction and guidance in the game of chess. Contemporary chess software may also be viewed as providing instruction on reasoning skills such as planning, decision-making, problem solving, and planning in the visual-spatial context of chess (Bart, 2004; Horgan, 1987). The educational value added by the use of chess software would be its potential to facilitate the development of higher reasoning skills.

The authors of this chapter are interested in the empirical investigation of the role and utility of contemporary chess software in human learning and education. This chapter will explore various issues associated with the game of chess, chess software, and the utility of chess software in educational interventions.

2. The Game of Chess

Chess is a game that has been played for many centuries. Some contend that chess originally comes from China; whereas, others contend that chess originally comes from India. Chess may also have begun in China and then passed through India on the way to Central Asia, the Middle East, and then Europe (Hooper & Whyld, 1996).

Millions of people play chess throughout the world. In addition, there are fine players and chess teachers on all the continents. The universal appeal of the game emanates from features of the game itself and the interaction of the game with computer technology. Among the features of chess that likely give the game universal appeal are the following: (a) fair in that all players begin with the same resources, (b) honest in that all players must follow the same set of rules including a prohibition on cheating; (c) competitive in that players play to win; (d) cognitively stimulating in that chess requires higher reasoning skills including skills at critical thinking, decision-making, problem solving, and planning;

and (e) cognitive enhancing in that chess can permit players to practice and improve their higher reasoning skills.

Chess is a game played on a square grid with 8 rows and 8 columns. The chessboard has 64 component squares with 32 being white (or light-colored) and 32 being black (or dark-squared). The squares are arranged in an alternating manner so that bordering any white square will be 2–4 black squares and bordering any black square will be 2–4 white squares. The chessboard is identical to a checkerboard. Sensitivity to square color is a feature of advanced chess skill (Weeramantry & Eusebi, 1993).

Chess is a game played typically between two individual players, although chess can be played between two groups. One player plays a set of white (or light-colored) pieces and the other player plays a set of black (or dark-squared) pieces. The player who plays the white (or light-colored) pieces is termed "White" in a chess game and the player who plays the black (or dark-colored) pieces is termed "Black." The pieces available to the two players are equivalent. In that sense, chess is a fair game.

Figure 1 displays the starting position for a chess game. The White pieces are arrayed on one side of the board and the Black pieces are arrayed on the opposite side of the board. From the perspective of the White side, the rows of the chessboard are termed ranks with each rank identified by a number from 1 to 8. The rank closest to White is termed rank 1 and the rank farthest from White is termed rank 8. Thus from the White side to the Black side, the ranks are arranged in the following order: rank 1, rank 2,. . . , rank 8.

From the perspective of the White side, the columns of the chessboard are termed files with each file identified by a letter from a to h. The file on the left side of White is termed file a and the file on the right side of White is termed file h. Thus from the left

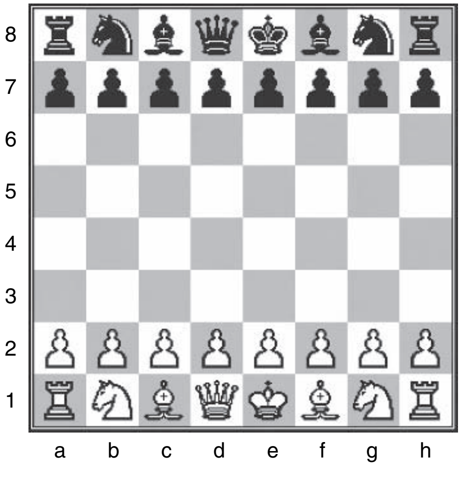

Figure 1. The starting position for the game of chess.

side of White to the right side of White, the files are arranged in the following order: file a, file b,. . . , file h.

As a result, a rank number and a file letter identifies each component square. This system of marking squares permits a system of notation, termed the algebraic notation, for moves. The algebraic notation permits chess games to be recorded and then studied in some other location or time. There are Internet websites such as www.chessgames.com and DVDs such as Mega Database 2005 that contain over 200 000 chess games with most of these games being games between very competent chess players. In fact, Mega Database 2005 contains over 2 500 000 chess games.

The pieces available to each player are the following: eight pawns, two knights, two bishops, two rooks, one queen, and one king. The White pawns are arrayed on the eight squares of rank 2 and the Black pawns are arrayed on the eight squares of rank 7 with one pawn on each square in those two ranks. Each chess piece of a chess player can move between squares, but not to squares that are occupied by other pieces of the same player.

The pawn is the least-valued piece. It may move one or two squares forward on its file on its first move. After its first move, it can only move one square at a time. It can capture an opponent's piece that is located on a forward diagonal square either to its immediate left or to its immediate right. For example, a White pawn on e4 may capture a Black piece on d5 or on f5. The pawn is a short-range piece.

If a pawn reaches the rank farthest from its original location, the pawn may be transformed into a higher-valued piece such as a queen or rook. For example, a Black pawn that starts on c7 can become an additional Black queen if it reaches c1. Each pawn has thus a "Cinderella" quality. A popular first move for White is moving the pawn on e2 to e4 and is simply written as "e4." A popular first move for Black is moving the pawn on c7 to c5 and is simply written as "c5."

The knight is worth approximately three pawns. The two White knights are initially placed on b1 and g1 and the two Black knights are initially placed on b8 and g8. Knights move in a L-pattern, two squares in one direction (either along a rank or along a file) and then one square in a direction perpendicular to the first direction. For example, the White knight on b1 may be moved to c3 or a3. The knight is the only piece that may jump over pieces.

The knight may capture an opponent's piece that is located on any square to which it could move. For example, a Black knight on c6 may capture a White piece on b4. When a piece captures an opposing piece, it occupies the square that was occupied by the opposing piece. The knight is a medium-range piece.

A popular first move for a White knight is moving the knight on g1 to f3 and is simply written as "Nf3." A popular first move for a Black knight is moving the knight on b8 to c6 and is simply written as "Nc6."

The bishop is also worth approximately three pawns. The two White bishops are initially located on c1 and f1 and the two Black bishops are initially located on c8 and f8. Both White and Black have a bishop on a white (or light-colored) square termed a white bishop and a bishop on a black (dark-colored) square termed a black bishop.

Bishops move only along diagonals. The white bishops can only move on white (or light-colored) squares and the black bishops can only move on black (or dark-covered) squares. The bishop is a long-range piece that can move from one corner of the chessboard to the opposite corner of the chessboard. However, it cannot jump over pieces.

The bishop may capture an opponent's piece that is located on any square to which it could move. For example, a White bishop located at b2 may capture a Black rook located at h8 provided that there are no pieces, White or Black, in between those two squares, that is, on the diagonal squares c3, d4, e5, f6, and g7.

A popular move for a White bishop is moving the bishop on f1 to g2 and is simply written as "Bg2." A popular move for a Black bishop is moving the bishop on c8 to d7 and is simply written as "Bd7."

The rook is worth approximately 5 pawns. The two White rooks are located on a1 and h1 and the two Black rooks are located on a8 and h8. Squares a1 and h8 must be black (or dark-colored) squares and a8 and h1 must be white (or light-colored) squares.

Rooks move only along ranks or files. For example, a White rook on a3 can move to d3 if there are no pieces (White or Black) on b3 and c3. The rook is a long-range piece that can move along ranks and files; however, it cannot jump over pieces.

The rook may capture any opponent's piece that is located on any square to which it could move. For example, a White rook located at e1 may capture a Black rook located at e8 provided that there are no pieces, White or Black, in between, that is, on the file e squares of e2, e3, e4, e5, e6, or e7. A rook that captures an opponent's piece replaces the piece on the square that was occupied by the opponent's piece.

A popular move for a White rook is moving the rook on a1 to d1 and may be simply written as "Rd1." A popular move for a Black rook is moving the rook on h8 to e8 and may be simply written as "Re8."

The queen is worth approximately 9 pawns. The White queen is located on d1 and the Black queen is located on d8. Square d1 must be a white (or light-colored) square and d8 must be a black (or dark-colored) square.

The queen moves only along ranks, files, or diagonals; however, it cannot jump over pieces. For example, the White queen on d3 can move to f5, d5, or b5 if there are no pieces (White or Black) on e4, d4, and c5. A queen may be viewed as a combination of a rook and a bishop. The queen is the most powerful chess piece.

A queen may capture any opponent's piece that is located on any square to which it could move. For example, the White queen located at f3 may capture a Black rook located at c6 provided that there are no pieces, White or Black, in between, that is, on the diagonal squares of e4 and d5. A queen that captures an opponent's piece replaces the piece on the square that was occupied by the opponent's piece.

A popular move for the White queen is moving the queen on d1 to c2 and is simply written as "Qc2." A popular move for a Black queen is moving the queen on d8 to c7 and is simply written as "Qc7."

The king is worth the entire game. The White king is located on e1 and the Black king is located on e8. If a king is threatened by pieces of the opponent, if the squares to which the king could move are either occupied by other pieces or attacked by pieces of the opponent, and if the player cannot obstruct the piece(s) of the opponent that threatens the king, then the opponent may say "checkmate" and win the game.

The king moves along ranks, files, or diagonals. For example, the White king on d3 can move to e4, d4, or f4. The king is a short-range piece that can move only one move in any straight direction (horizontally, vertically, or diagonally). It cannot jump over pieces.

A king may capture any opponent's piece that is located on any square to which it could move. For example, the White king located at g3 may capture a Black rook located at f4 provided that the Black rook is not protected by another Black piece. A king that captures an opponent's piece replaces the piece on the square that was occupied by the opponent's piece.

A possible move for the White king is moving the king on e1 to f1 and is simply written as "Kf1." A possible move for the Black king is moving the king on e8 to f7 and is simply written as "Kf7."

If a player places a piece on a square from which it can directly attack the opponent's king, then the player says "check" indicating that the other player's king is in danger of being captured. The king can never be placed on a square from which it is attacked by a piece of the opponent. In other words, neither king can ever be placed in harm's way. As a result, one king can never be placed on a square that is adjacent to a square on which the opponent's king is placed. The White king and the Black king can never be placed next to each other on the chessboard.

Chess players are expected to manage their pieces to "checkmate" their opponent. The task of the chess player is similar to the task of the coach of a soccer team managing players in order to win games or the task of a business leader managing business staff in order to promote financial growth of the business. With each side initially having 16 pieces from six types of chess pieces, chess may be viewed as a complex, multivariable game.

To appreciate the complexity of chess, please consider that, after only three moves by both White and Black, there are 1 000 000 possible resulting chess positions if we assume a very conservative 10 possible move available to each player for each of the three moves. 10 move choices for White for move 1×10 move choices for Black for move 1×10 move choices for White for move 2×10 move choices for Black for move 2×10 move choices for White for move 3×10 move choices for Black for move $3 = 1\ 000\ 000$ possible resulting chess positions.

Chess games tend to have three phases: the opening, the middle game, and the endgame. There are hundreds of openings and variations of those openings. Mastery of this phase of the game requires memorization of many opening move sequences. The minimum viable number is three: one opening for White, one defense for Black in response to move e4 for White, and one defense for Black in response to move d4 for White. The opening tends to be the initial 10–15 complete moves (White and Black) of a game.

The middle game tends to be the most demanding phase of the game. Middle games often require reasoning skills on the part of the chess players, including the comprehension of chess positions, the identification of candidate moves, the analysis of move sequences resulting from those candidate moves, and the evaluation of chess positions resulting from selected move sequences. The numbers of chess positions possible in chess middle games are staggering. Middle games tend to be the next 11–30 complete moves (White and Black) of a game.

The endgame in some chess games may be rather mechanical. For certain endgames, a chess player may know exactly the sequence of moves that will lead to a win. For other endgames, a chess player may need to be sensitive to subtle features of the endgame positions and may need to do much calculation, that is, determine viable sequences of candidate moves, in order to win the games.

Chess is a complex game that involves tactics and strategies. Tactics relates to the usage of several pieces in a region of the chessboard to win a piece or to gain a spatial or material advantage in a chess game. A tactical player looks at specific regions on the chessboard. Strategies relate to the positioning of all of the pieces in order to gain a spatial or material advantage in a chess game. A strategic (or positional) player looks at the entire board and attempts to array the pieces into positions that are advantageous to the player.

In addition to chess having a notation and involving tactics and strategies, chess has a metric to index chess skill. The metric is termed the Elo rating originally developed by Arpad Elo. The Elo rating is the measure of chess skill used throughout the United States and elsewhere.

Elo ratings range from 0 indicating a total absence of chess skill to approximately 2800 indicating super grandmaster or even world championship level of chess skill. Elo rating intervals determine classes of chess skill (Goichberg, Jarecki, & Riddle, 1993):

Rating range	Name
2400 and above	Senior Master
2200–2399	Master
2000–2199	Expert
1800–1999	Class A
1600–1799	Class B
1400–1599	Class C
1200–1399	Class D
1000–1199	Class E

In addition to those classes, there are the very advanced classes of International Master and Grandmaster. To attain the title of international master, one needs to perform competitively at tournaments in which the opponents are titled players such as International Masters or Grandmasters. In order to attain the title of Grandmaster, one needs to perform competitively at tournaments in which the opponents are Grandmasters.

Consider an individual to be named Leonard, who is provisionally rated at 1200 after two rated games. Leonard plays three more rated games. With the first game, Leonard wins against an opponent with 1500 rating. With the second game, Leonard loses to an opponent with a 1300 rating. With the third game, Leonard draws against an opponent with a 1300 rating. Leonard's new rating is calculated in the following manner:

Step One: The Computation of Total Elo Rating Points.

1.	2×1200	=	2400 Elo rating points
2.	$1 \times 1500 + 400$	=	1900 Elo rating points
3.	$1 \times 1300 - 400$	=	900 Elo rating points
4.	1×1300	=	1300 Elo rating points
	Total		6500 Elo rating points

Step Two: The Computation of the Elo Rating

6500 total Elo rating points/5 chess games = 1300 Elo rating.

Leonard received additional points (e.g., 400 points) when defeating an opponent and lost additional points (e.g., 400 points) when losing to an opponent. Elo ratings based on less than 30 rated chess games are termed provisional. Elo ratings based on 30 or more rated games are no longer provisional, but rather generally accepted as reliable and relatively accurate by tournament officials.

3. Chess Reasoning

Gobet and Simon (1996) have contended that there are two prominent psychological mechanisms underlying skilled chess-playing performance. They contended that one such mechanism is "the recognition of cues in chess positions that evoke information from the expert's memory about possible moves and other implications of recognized patterns of pieces." This first mechanism may be referred to as "chess game comprehension."

They also contended that the second mechanism is "planning by looking ahead at possible moves, possible responses by the opponent, possible responses to those responses, and so on." This second mechanism may be termed "chess planning and calculation."

A more delineated interpretation of how competent chess players think when faced with the task of moving a piece would involve the following steps:

1. Comprehension of the chess game position.
2. Identification of candidate moves.
3. Analysis of possible sequences of moves following selected candidate moves.
4. Evaluation of chess positions that result from the possible move sequences.

As a result, one could infer that chess involves critical thinking, because chess players should comprehend chess positions, analyze move sequences, and evaluate resulting chess positions. According to the American Philosophical Association (1990), critical thinking involves comprehension, analysis, and evaluation.

In addition to critical thinking, chess requires problem-solving skills, as chess players attempt to determine how to move pieces to certain squares or how to trade their weak pieces for the strong pieces of their opponents. Chess also requires decision-making skills, as chess players need to make decisions as to which of several candidate moves they will actually make. Chess thus provides a context in which individuals can engage in critical thinking, problem solving, and decision-making. As a result, chess permits the development of higher-order reasoning skills such as planning, problem solving, decision-making, and critical thinking.

Charness (1989) explicated what constitutes chess expertise. Charness tested a chess players named DH with the use of an array of cognitive tasks using chess positions at two times; when DH was an average tournament chess player and 9 years later when DH became a chess master. During that 9 year period, DH did not improve in the depth of search in terms of the length of chess move sequences to be analyzed. However, DH perceived board configurations during the second testing in terms of larger chunks (or arrays) of pieces than the chunks perceived during the first testing. In addition, DH assessed endgame positions faster during the second testing. Also, DH was more selective in the selection of candidate moves during the second testing than during the first testing. Drawing from his inquiry with DH and other investigations of chess expertise, Charness concluded that chess experts perceive chess configurations with the use of larger chunks than do chess novices and that chess experts are more selective in the consideration of candidate moves than chess novices.

4. Research on Chess and Education

Certain studies on chess have relevance for education and learning. Frydman and Lynn (1992) assessed 33 young tournament chess players from Belgium. These children with a mean age of 11.0 years completed the French Wechsler Intelligence Scale for Children. The children were placed into three groups: (a) Group 1 consisted of players with Elo ratings greater than 1550; (b) Group 2 consisted of players with Elo ratings in the 1350–1550 range; and (c) Group 3 consisted of players with Elo ratings in the 1000–1350 range.

One prominent result was that the participants had a mean full-scale IQ of 121. The second prominent result was that the mean performance IQ of 122 was significantly greater than the mean verbal IQ of 109. These results suggest that competent tournament-level chess skill requires superior general intelligence and superior visual-spatial ability.

A study of school-aged American chess players by Horgan and Morgan (1990) had a similar finding. Horgan and Morgan found that above-average chess players scored higher than average on the Raven's Progressive Matrices Test, an excellent measure of general intelligence.

But neither of those studies was an experiment. As a result, it is not clear whether chess training increases general intelligence and visual-spatial ability or whether general intelligence and visual-spatial ability are prerequisites for tournament-level chess skill. In addition, neither the Frydman and Lynn study nor the Horgan and Morgan study provided information as to whether their young participating chess players were superior students. One would predict that their participants were superior students, because IQ including visual-spatial intelligence is positively correlated with scholastic achievement (Mackintosh, 1998).

Doll and Mayr (1987) reported that 33 chess Master in the 18–34 years of age range scored significantly higher on the processing of complex information, numerical thinking, reasoning speed, and general intelligence on subtests of the Berlin Intelligence Structure test than did non-chess players, but not differently on a visual-spatial reasoning task. This study of the interrelationship between adult reasoning skills and chess skill was also not an experimental study.

Christiaen and Verholfstadt-Daneve (1978) randomly assigned 40 young 5th-grade boys in the 10–11-year age range to an experimental treatment and to a non-treatment. The experimental treatment participants received chess instruction for 1 h, Friday afternoons, for two school years. The dependent variables were school grades and scores on two Piagetian tasks, a balance task and fluid task. The two Piagetian tasks served as measures of formal operational reasoning that marks advanced cognitive development that can emerge in adolescence (Inhelder & Piaget, 1958).

One prominent result was that the experimental participants registered school grades that were significantly higher than the school grades of the control group. However, their Piagetian task performances were higher but not significantly greater than the Piagetian task performances of the non-treatment group participants. Christiaen and Verholfstadt-Daneve contended that chess instruction can contribute to improved school performance among students in middle childhood.

Smith and Cage (2000) investigated the effect of chess instruction on mathematics achievement among rural, southern, African-American high school students. The participants were from rural regions of northern Louisiana near the city of Monroe. The treatment group consisted of 11 females and 9 males, 20 upper-class high school students in total. The control group consisted of 20 upper-class high school students randomly selected from the general student population.

The students in the treatment group received 120 h of chess instruction over a 5-month period. The instruments used in the study were all cognitive measures. One measure was the mathematics section of the California Achievement Test (Level 20). A second measure was the Guilford–Zimmerman Test of Spatial Visualization (SV) that was used to assess spatial ability. A third measure was the Group Embedded Figures Test (GEFT) that was used to assess level of field dependence/independence. The fourth measure was the Naglieri Nonverbal Ability Test (NNAT) that was used to assess nonverbal reasoning ability. All four measures were used to provide pre-test and post-test data.

One prominent result was that the treatment group had significantly greater means for the measures of mathematics achievement, spatial ability, and nonverbal reasoning than the control group. In addition, the treatment group was, on the average, more field independent than the control group. A third result was that the male participants registered, on the average, significantly greater gains in mathematics achievement than the female participants. The gains in CAT scores for the males were 13.9% for the treatment males and only 8.5% for the control males. The gains in CAT scores for the females were 8.2% for the treatment group females and only 1.7% for the control group females.

Smith and Cage interpreted these results as indicating that the chess instruction engendered improvements in concentration, creativity, patience, and perseverance among the treatment participants and that those positive changes led to improvements in mathematics achievement and related abilities.

In summary, the few studies on the relationship between chess and human learning that were published in refereed journals support the view that chess instruction engenders improvements in scholastic achievement – especially mathematics achievement – and basic cognitive abilities such as spatial ability. However, much more research is needed to investigate the cognitive and affective effects of chess instruction on various groups of participants. Correlational research is certainly helpful to examine relationships between chess instruction and various cognitive and affective variables. However, only experimental research will reveal patterns of causation between chess instruction and changes in cognitive and affective variables.

It is likely that future studies on the cognitive and affective effects of chess instruction will involve contemporary chess software in the chess interventions, because of the capacity of contemporary chess software to assess chess skill, to provide the chess instruction, and to provide corrective feedback to participants.

5. Chess Software

The impact of chess on computer software and the Internet is extraordinary Websites devoted to chess include Chessgames.com at http://www.chessgames.com/and the Internet Chess Club at http://www.chessclub.com/. In addition, software devoted to chess is often very sophisticated.

For Mac users who use Mac OS 9, MacChess 5.0.1 is exceptional chess software that is freeware. It provides artificial opponents of different levels up to the Master-level. It provides a hint option during games to provide move suggestions for a chess player. In addition, the board and piece settings are pleasant and conducive to game concentration. Many Mac users undoubtedly hope that a new version of MacChess is formulated that can operate with Mac G4 and G5 machines that use Mac OS X.

Sigma Chess 6.1 Pro is another fine chess software for Mac users. This software designed by Ole Christensen is a commercial product, modestly priced, that can work with G4s and G5s that use the Mac OS X. It is a Master-level program that permits the chess player to play against artificial opponents at different skill levels. It allows one to analyze and annotate personal chess games and to print them for collections of personal chess games. It allows for an array of designs for pieces and boards and even changes in perspective.

Sigma Chess 6.1 Lite is related chess software that is freeware. It has many of the features of Sigma Chess 6.1 Pro including a Master-level artificial opponent. It could satisfy many recreational chess players who do not need many of the special features such as 3D board settings of Sigma Chess 6.1 Pro.

Another chess software for Mac users is Chessmaster 9000 by Feral Interactive. It is a commercial product that has an impressive set of features. It permits games against 150 different artificial opponents, from novice to grandmaster levels. It provides instruction on openings, endgames, and chess theory and principles. It even has a new feature termed "Blunder Alert" that notifies the chess player of moves that are errors and then suggests better moves.

For PC users, there is excellent chess software available. First of all, there is Chessmaster 9000 that provides chess drills, lessons, and tutorials. It permits chess games against a wide variety of artificial opponents at different skill levels. It also provides a "blunder alert" in case that the chess player is about to make a problematic move in an unrated chess game. With features such as "blunder alert," Chessmaster 9000 provides guided instruction in chess to players (Kirschner, Sweller, & Clark, 2006). Chessmaster 9000 for the PC is a commercial product of the UbiSoft company. UbiSoft is now distributing a new version of that chess software, termed Chessmaster 10th edition.

Another chess software for PC users is Fritz 9 that is arguably the most popular commercial chess software available. Fritz 9 is often used by professional chess players including Experts, Masters, and Grandmasters. For novice players, it provides games against artificial opponents at different skill levels, indicates threatened squares, and suggests and explains moves. For intermediate players, it provides instruction on openings and endgames and analyzes game positions. For highly proficient players, Fritz 9 provides games against artificial opponents of Expert and above levels, permits a full analysis of chess games, and provides review of hundreds of tournament games. The Chessbase company makes Fritz 9 and is now distributing a new version of that chess software, termed Fritz 10.

Basic to all of these forms of chess software is the chess engine, that is, the program that can actually play chess. The most famous chess engine is arguably Deep Blue 2 that defeated the former world chess champion, Gary Kasparov, in a chess match in 1997. Feng-Hsiung Hsu led the IBM-sponsored team that developed Deep Blue 2. He recounted that entire effort including the drama of the match in a book entitled *Behind Deep Blue* (Hsu, 2002).

In that volume, Hsu clearly indicated the three main software components of a chess engine: the move generator, the search control, and the evaluation function. The move generator is software that finds the chess moves. The search control is software that analyzes the move sequences identified by the chess engine. The evaluation function assesses the worth of the position reached if the chess engine makes certain moves. An evaluation function may produce a number from $+\infty$ to $-\infty$. For example, in MacChess, if a chess player plays an artificial opponent and if the evaluation function value after a move is positive, then the artificial opponent has a better position; if the value is negative, then the chess player has the better position. The evaluation function of such chess software provides a quantitative index to help psychological researchers determine exactly at which move a chess player makes an error and how serious the error occurs.

Chess software provides a means to determine the level of chess skill of a chess player and a means to identify errors in moves. Such software could be used in any series of research studies on the cognitive and affective effects and correlates of chess instruction. Such research could involve a wide spectrum from young children to adults and the elderly. Following will be three reports of how chess software can be used in education.

The first report will describe a study of the relationship of chess instruction to learning among at risk students. The first study provides an example of how one can examine the effects of chess instruction using chess software on the cognitive abilities of learners.

6. A Study of Chess Instruction Among at Risk Students

Regardless of chess skill level, chess playing provides numerous opportunities in which players can practice higher order thinking skills. We face an emerging question; can chess instruction have salutary cognitive effects with other types of students such as at risk students or disabled students? Storey (2000) suggested that chess instruction could benefit children with disabilities including behavior disorders (e.g., hyperactivity) and cognitive disabilities (e.g., mental retardation, slow learning). This study examines this issue with students at risk for academic failure.

6.1. Method

6.1.1. Participants

Thirty-eight students, ages 8–12, from three elementary schools participated. The schools are located in Seoul, Korea. Of these students, all students enrolled in an after school program for students at risk for academic failure. Students at risk were identified according to criteria by using the Basic Skill Test (BST) developed by the Korean Ministry of Education. They possessed poor math, reading, and writing skills. Approximately 3–5% of students per school fell into this category. They showed significant deficits in more than one area among the domains of reading, writing, and math. Playing chess was a novel experience to most of the students.

The control group consisted of 15 males and 5 females with an average age of 9.74 years and the experimental group consisted of 12 males and 6 females with an average age of 9.71 years. The control group had 17 students at risk and 3 students with learning disabilities and the experimental group had 15 students at risk and 3 students with learning disabilities.

The students were assigned randomly to each group. Also, the two groups were compared on demographic variables and intelligent test scores. Using statistical procedures such as one-way analysis of variance (ANOVA), the groups appeared to be equivalent on gender, age, grade, school and disabilities.

6.1.2. *Procedure*

After school personnel identified students at risk and parents returned consent forms, the study began with the administration of the two pre-tests. A researcher and a research assistant administered tests in the first week of this study. The Test of Nonverbal Intelligence-3 (TONI-3) was administered to the students. The TONI-3 was administered individually. Then the participants were randomly assigned to an experimental group or a control group. The experimental group received a 90-min chess lesson once per week over 3 months, and the control group students attended regular school activities after class.

At the end of the chess intervention, the TONI-3 was again given to the students. Chess instruction consisted of 12 separate lessons over a 3-month period. Each lesson included three segments: reviewing, lecturing, and chess playing. The chess instructor developed and provided a set of quizzes. Each quiz was used to identify student difficulty in understanding chess moves and rules. Each following lesson started with a review of the previous lesson and a quiz. The last six lessons were implemented in a computer lab with chess software and were intended to allow students to practice higher order cognitive skills.

Overall, the student was asked to follow four steps to develop his/her chess skills: (a) understand chess rules; (b) think ahead for a plan; (c) implement the plan; and (d) feedback and rehearsal. The researcher and the chess instructor developed twelve sessions that were taken from the Comprehensive Chess Course (Pelts & Alburt, 1992). The benefit of chess software was embedded in the session by using it as a tool to practice and generalize the contents of each lesson.

6.1.3. *Instruments*

TONI-3. The Test of Nonverbal Intelligence (TONI-3) (Brown, Sherbenou, & Johnsoen, 1997) is one test used in the study and is a norm-referenced, language-free measure of cognitive ability. In particular, the TONI-3 was designed to measure problem solving, aptitude, and reasoning skills. Two equivalent forms are available. Each form of the TONI-3 has 50 items. Converted scores from obtained raw scores are provided with a mean 100 and a standard deviation of 15. It is often administered for individuals who are believed to have difficulties in taking tests, disabilities, or lack of exposure to the English and the United States cultures. In this study, two forms (A and B) were administered. Alternate forms reliability has ranged from 0.79 to 0.92. The converted TONI scores were age-based standard scores.

6.1.3.1. Chess Rating In this study, chess software, Chessmaster 9000, estimated the chess ratings of the participants, as a result of the performances of the participants in chess games against artificial opponents at different Elo skill levels such as 300. The software not only rated student chess skill after each game, but also provided chess lessons and drills designed for novice and intermediate chess players.

6.2. Results

The TONI-3 pre-test means were 96.50 for the experimental group and 85.60 for the control group with standard deviations of 17.12 for the experimental group and 20.49 for the control group. The TONI-3 post-test means were 100.83 for the experimental group and 97.25 for the control group with 11.78 for the experimental group and 13.18 for the control group.

The results of repeated measures ANOVA indicated that the main effect for chess instruction was not significant for the TONI-3 with $F(1,\ 36) = 2.395,\ \ p > 0.05$. The treatment X-time interaction effect, reflecting differences among the groups in amount of change, was also not statistically significant for the TONI-3 with $F(1,\ 36) = 2.481,\ \ p > 0.05$.

Partial correlation analysis was employed to explore relations among variables in the experimental group. The partial correlation of the TONI-3 post-test score and the chess rating with the TONI-3 pre-test score being held constant was .520, $p < 0.05$. The mean chess rating was 131.39 with a standard deviation of 84.94. Thus, there was significant relationship between TONI-3 post-test scores and chess ratings.

In addition, a stepwise regression was conducted to determine whether either of the variables, TONI-3 pre-test score and chess rating, were necessary to predict TONI-3 post-test score. The stepwise regression analysis confirmed a model that included two significant predictors, $F(2,\ 15) = 12.254,\ \ p < 0.001$. The two predictors were TONI-3 pre-test scores and chess rating. Those two variables accounted for the 62% of the variance in the post-test TONI-3 score. Although the sample size was small, this result suggests that chess rating was related to the TONI-3 post-test score.

6.3. Discussion

One result was that chess instruction did not significantly affect nonverbal reasoning ability as measured by the TONI-3. However, student chess rating was significantly correlated with nonverbal reasoning as measured by the TONI-3, after controlling for TONI-3 pre-test scores. This result lends credence to the view that the attainment of higher chess ratings, rather than mere exposure to chess, was related to the improvement of nonverbal reasoning. For the chess group only, the result supported the view that transfer of cognitive skill occurred, even though more research on the effects of chess instruction on cognitive skills of students at risk for academic failure is warranted.

In summary, we suggest that neither merely playing chess nor merely receiving chess instruction will engender salutary cognitive effects among participants. What is crucial is that the chess instruction engenders gains in chess skill (e.g., Elo rating), which, in turn, could lead to changes in the cognitive capacities of participants. Thus, as an example, any chess intervention with military personnel intended to improve nonverbal and verbal reasoning, subject-matter learning, and certain attributes such as perseverance, patience, creativity, concentration, and problem-solving ability should involve contemporary chess software that engenders gains in chess skill (e.g., Elo rating), as well provides chess instruction, measures level of chess skill, and provides corrective feedback to the military participants.

One limitation of the study is that there were no explicit assessment of the following chess skills: (1) understand chess rules; (2) engage in planning activities; and (3) implement plans in playing chess. As a result, there was no assessment as to the extent to which the intervention engendered improvements in those three capabilities. Such assessment is suggested for future research inquiry on the cognitive effects of educational interventions making use of contemporary chess software such as Chessmaster 9000.

A second limitation of the study is that there were no dependent variables of academic skills as were used by Smith and Cage (2000). As a result, there was no assessment of the extent to which the intervention engendered improvements in academic skills and scholastic achievement. Such assessment is suggested for future research inquiry on the scholastic effects of educational interventions making use of contemporary chess software.

We also recommend that the cognitive effects of chess performance on students at risk for academic failure and other groups of individuals continue to be studied. The heretofore-mentioned methods for improving chess instruction could be incorporated in such inquiry. Chess instruction specially configured may be proved to be very efficacious in producing salutary cognitive effects among students at risk for academic failure as well as other groups such as adults and the elderly in the United States, and elsewhere in the world. Future research may convincingly demonstrate the utility of contemporary chess software in enhancing human learning and in improving education.

The second report will describe a study of transfer of visual-spatial problem solving using chess software. The second study provides an example of how one can study transfer of higher order reasoning capabilities such as problem-solving skills using chess software.

7. A Study of Transfer of Problem-solving Skills Using Chess Software

The second study was designed to investigate the likelihood that novice chess players can solve simple chess problems and the extent to which skill at solving a simple chess problem transfers to a similar simple chess problem among novice chess players.

7.1. Method

7.1.1. Participants

The participants in the study were 98 undergraduate and graduate students at the University of Minnesota, who had little or no experience at chess. The average age of the participants were 27 years and 4 months. There were 31 males (31.6% of the sample) and 67 females (68.4% of the sample).

7.1.2. Procedure

The participants were asked to solve two chess problems. Prior to the two chess problems, the investigator instructed each participant as to the basic rules of chess, how the King and Rooks move, and what "check" and "checkmate" mean. The investigator used the chess software, Chessmaster 8000, in this instruction. In addition, the investigator used the basic rules of the chess section from the website of the United States Chess Federation to teach these rudimentary features of the game of chess. The duration of this instruction was approximately 40 min.

After this initial instruction, the investigator allowed the participant to play with Kings and Rooks on a chessboard provided by Chessmaster 8000. The chessboard displayed the endgame of a King and Rook against a King as an exercise prior to the first problem. After the participant demonstrated correct movements of those pieces, the investigator provided the two simple chess problems in a specific order.

In the first task, the participants were provided an endgame position involving the White King on e1, White Rooks on a1 and h1, and the Black King on e8. Each participant was asked to play White and attempt to win the game by checkmating Black. The Figure 2 depicts the chessboard used in the first task.

In the second task, the participants were provided an endgame position involving the White King on e1, Black Rooks on a8 and h8, and the Black King on e8. Each participant was asked to play Black and attempt to win the game by checkmating White. The Figure 3 depicts the chessboard used in the first task.

In both tasks, they played against the computer, which was an artificial player at the level of Chessmaster (Elo score: 2957). The participants had 12 moves to finish the game in each problem. Only the first 10 moves were used for our analysis.

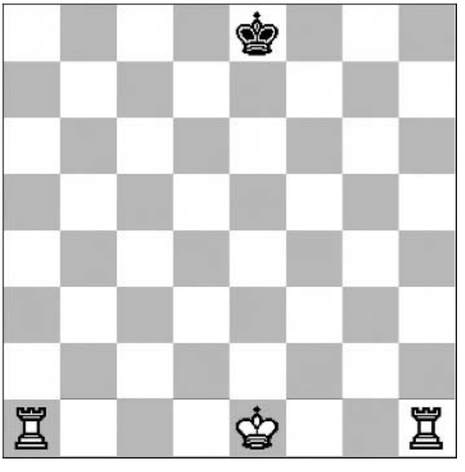

Figure 2. The chess position for the first simple chess problem.

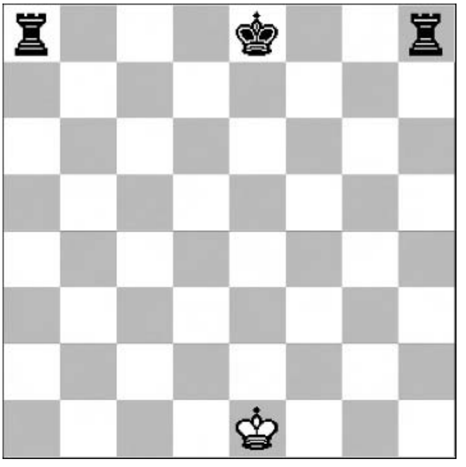

Figure 3. The chess position for the second simple chess problem.

Solving the chess problem in 12 moves served as evidence that a participant had the strategy involving rook coordination to solve the problem in task 1, because that strategy is required to solve the endgame chess problem in task 1 in less than 12 moves. As a result, there was no need for an independent measure of whether the participants actually learned the strategy in task 1.

7.2. Results

The research question was "Does performance on chess problem 1 transfer to performance on chess problem 2?" It was hypothesized that the participant who solved the first problem is likely to solve the second one, because the same strategy is applied to both problems. Among the 98 participants, 22 (22.4% of the participants) were able to solve the first problem. Since the participants were all novice chess players, the low success rate was not that surprising. However, only eight of the 22 (36.4%) of the participants, who solved the first chess problem correctly, also correctly solved the second problem. This result indicated that problem-solving skill with one simple chess problem does not readily transfer to a similar simple chess problem.

7.3. Discussion

One conclusion of this study is that most novice chess players lack problem-solving skills at solving simple chess problems. An inference of this conclusion is that problem-solving skills with visual-spatial problems are not readily available to adult learners. This condition may thus necessitate the construction and implementation of training programs in visual-spatial problem-solving skills in order for adult learners to develop such higher order reasoning skills.

A second conclusion is that problem-solving skill with one simple chess problem does not readily transfer to similar simple chess problems. An inference of this conclusion is that problem-solving skills with visual-spatial problems do not readily transfer to similar visual-spatial problems. This condition may thus necessitate the construction and implementation of training programs in visual-spatial problem-solving skills to increase the likelihood of positive transfer of such higher order reasoning skills across visual spatial problems.

Charness (1989) found that expert chess players are better able to recall chess positions and the strategies to be employed by opponents than novice players. As a result, chess training may lead to improvements in visual-spatial memory and in perspective taking.

The third report will provide an informal commentary on the experience of the first author in teaching a course on chess and reasoning to first year undergraduates. The course made judicious usage of chess software and chess-related websites.

8. Observations on Teaching Chess to College Freshmen

Contemporary chess software lends itself to research and instruction opportunities that eclipse inquiry on students at risk for academic failure. The first author (W. Bart) has taught a popular freshman seminar on chess and higher-order reasoning skills to first-year undergraduates for 4 years. In the course, the instructor begins the course by teaching the rudimentary features of chess such as how specific pieces move and how one checkmates an opponent. In addition to the basic components of playing chess, the instructor introduces the students to algebraic chess notation so that they can record their games and to instructive classical games to illustrate some basic principles of chess such as control of the center. The instructor provides the students chess exercises in the form of simple chess endings with only a few pieces on the board. The students are expected to win the games by making appropriate moves.

After the initial instruction in the basic components of chess, the instructor has the students play each other and well as play artificial opponents presented by the chess software, In addition, the instructor presents and leads discussions of highly instructive tournament chess games and evaluations of chess games produced by the students. In fact, a major mechanism of chess improvement and assessment in the course is the evaluation of personal chess games. Students in the course are expected to prepare written evaluations of three of their personal chess games.

In addition to the contemporary chess software of Fritz 8, this freshman seminar makes judicious usage of the website termed Chessgames.com. That website has a repository of over 200 000 important tournament chess games, some of which are used by the instructor to illustrate various key ideas in chess. The freshman seminar makes substantial usage of contemporary chess software and free chess-related websites.

Initially, many of the first-year students are impatient and impulsive in playing chess. They are often oblivious to viable candidate moves and do not formulate any plans for how they will proceed in their games. They tend not evaluate their moves before they actually make their moves. Also, they took full advantage of the restroom breaks that would occur near the middle of the class sessions.

By the end of the seminar, the students became patient and careful when playing chess. They became more aware of candidate moves and tended to formulate plans for their chess games. They also tended to evaluate their moves before they actually made their moves. Also, they periodically disregarded restroom breaks in favor of studying chess positions! The students in general expressed their enthusiasm for the game of chess and recommended that more learners study chess to develop the higher reasoning skills required in chess. In end-of-course reflection papers, students indicated that the freshman seminar instruction improved their higher reasoning skills in other university courses.

There are certain implications of these observations. First, training in chess that explicitly highlights components of higher reasoning skills including problem solving, decision-making, critical thinking, and planning will likely engender the development and utilization of higher reasoning skills in academic courses. Second, training in chess that explicitly highlights the merits of affective traits extolled in chess such as patience and determination will likely engender the development and employment of such affective traits in other settings.

During the Fall of 2005, elementary school students in cities such as Philadelphia, Tampa, Seattle, San Diego, and New York received chess instruction. Paul Vallas, chief executive of the Philadelphia school system and former head of the Chicago public schools, contends that "chess is a great educational tool. 'Chess seems to improve problem-solving skills,' he said. 'It improves discipline. It improves memory. It certainly seems to improve mathematical skills' " (McClain, 2005).

Chess instruction and mathematics instruction are quite compatible, For example, the Chess and Math curriculum in the Canadian Province of Quebec incorporates chess instruction in the mathematics instruction in schools. The Chess and Math curriculum is also used in other Canadian Provinces. Chess instruction is also required in the schools of countries such as Turkey, Bulgaria, and Iceland. In addition, there are extensive scholastic chess activities in countries such as China, India, Spain, Israel, and Venezuela. Chess is also part of the regular school curriculum in New Jersey.

With all of this worldwide chess instructional activity, there should be extensive research on the cognitive and affective effects of chess instruction among adult learners as well as child and adolescent learners. For example, worthy of consideration would be research on the effect on academic performance of teaching "key ideas" in chess. Such chess instruction may likely have salutary cognitive and affective effects on adult learners as well as pre-adult learners.

Acknowledgments

The authors wish to thank the educators in Korea for their cooperation and the students in Korea for their participation in the study that occurred in Korea. Also, the authors want to thank the Center for Cognitive Sciences at the University of Minnesota for room

space to do a study on chess reasoning and the university students at the University of Minnesota, who participated in the studies that occurred in Minnesota.

If you are interested in the study of contemporary chess software to enhance human cognition and learning, please contact William M. Bart, Ph.D. at bartx001@umn.edu.

References

American Philosophical Association. (1990) Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. The Delphi Report: Research findings and recommendations prepared for the committee on pre-college philosophy. (ERIC Document Reproduction Service No. ED 315–423)

Bart, W. (2004, May). Cognitive enhancement: An approach to the development of intelligence. Poster session presented at the annual meeting of the American Psychological Society, Chicago, Illinois.

Brown, L., Sherbenou, R. J., & Johnsoen, S. K. (1997). *Examiner's manual: Test of Nonverbal Intelligence* (3rd ed.). Austin, TX: PRO-ED.

Charness, N. (1989). Expertise in chess and bridge. In D. Klahr & K. Kotovsky (Eds), *Complex information processing: The impact of Herbert Simon* (pp. 183–208). Hillsdale, NJ: Erlbaum.

Christiaen, J. & Verholfstadt-Daneve, L. Chess and cognitive development. *Nederlands Tydschrift voor de Psychologie, 36*, 561–582.

Doll, J. & Mayr, U. (1987). Intelligence and success in chess playing: An examination of chess experts. [German]. *Psychologische Beitrage, 29*, 270–289.

Frydman, M. & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology, 83*, 233–235.

Gedler, M. (1996). Educational games and simulations: A technology in search of a research paradigm. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 521–540). New York, NY: Simon & Schuster Macmillan.

Gobet, F. & Simon, H. A. (1996). The roles of recognition processes and look-ahead search in time-constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science, 7*, 52–55.

Goichberg, B., Jarecki, C., & Riddle, I. (Eds). (1993). *U. S. Federation's official rules of chess*. (4th ed.). New York, NY: David McKay.

Hooper, D. & Whyld, K. (1996). *The Oxford companion to chess*. Oxford, England: Oxford University Press.

Horgan, D. D. (1987). Chess as a way to teach thinking. *Teaching, Thinking and Problem Solving, 9*, 4–11.

Horgan, D. D. & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology, 4*, 109–128.

Hsu, F. H. (2002). *Behind Deep Blue*. Princeton, NJ: Princeton University Press.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Kirschner, P., Sweller, J., & Clark, R. (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, 4*, 75–86.

Mackintosh, N. (1998). *IQ and human intelligence*. Oxford, England: Oxford University Press.

McClain, D. (2005, October 13). More schools learn power of checkmate. *Chicago Tribune*, p. 1–30.

Merriam-Webster Online Dictionary. (2007). Retrieved January 2, 2007, from http://www.m-w.com/dictionary/.

O'Neil, H. & Fisher, Y. (2004). A technology to support leader development: Computer games. In D. Day, S. Zaccaro, & S. Halpin (Eds). *Leader development for transforming organizations* (pp. 99–102). Mahwah, NJ: Erlbaum.

O'Neil, H., Wainess, R., & Baker, E. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*, 455–474.

Pelts, R. & Alburt, L. (1992). *Comprehensive chess course: Volume 1* (4th ed.). New York, NY: Chess Information and Research Center.

Shenk, D. (2006). *The immortal game*. New York, NY: Doubleday.

Smith, J. & Cage, B. (2000). The effects of chess instruction on the mathematics achievement of southern, rural, black secondary students. *Research in the Schools, 7*, 19–26.

Storey, K. (2000). Teaching beginning chess skills to students with disabilities. *Preventing School Failure, 44*, 45–49.

Weeramantry, S. & Eusebi, E. (1993). *Best lessons of a chess coach*. New York, NY: David McKay.

WOULD YOU LIKE TO PLAY A GAME? EXPERIENCE AND EXPECTATION IN GAME-BASED LEARNING ENVIRONMENTS

Jacquelyn Ford Morie, Rebecca Tortell and Josh Williams
University of Southern California

Abstract

We present results from a series of experiments that looked at how previous experience and immediate priming affect a user's arousal state, performance and memory in a virtual environment used for training. We found that people's game play experience had effects on these measures, and that if participants expected the environment to be a game, they approached it with expectations that were not always conducive to optimal training. We suggest that the type of game being used for training will have the best outcome if users are familiar with that mode and have the appropriate schema to approach the training.

1. Introduction

Games have been used to train their participants for millennia. Games such as "Go" (c. 200BCE) and Chess (c. 600CE) were designed as high-level abstractions to instruct battlefield tacticians and to teach leaders skills in strategy, decision-making and courage (Encyclopedia Britannica, 2006; Fairbairn, 1995). In more recent centuries we have "war gaming," defined by the US Department of Defense as "a simulation, by whatever means, of a military operation involving two or more opposing forces using rules, data, and procedures designed to depict an actual or assumed real life situation." War games are now the preferred mode of training military tacticians (US Department of Defense, 2003).

Unlike Chess and Go, games delivered using modern day "game media," such as computer platforms and videogame consoles, have consistently decreased the games' levels of abstraction in favor of the realism offered by contemporary computer graphics. The excitement surrounding videogames and the degree of realism of which they are capable have attracted expert instructors in multiple fields to adapt them for education and

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

training. But one important question still needs to be addressed: What is the relationship between games to train and games to entertain?

Game media in the service of education are still, fundamentally, games. To successfully use digital games to teach we must take into account a number of factors surrounding how and why people play games. These factors may be crucial to using game media successfully for educational purposes. They include how well the game mode (the type of game being played) matches the content of the lesson to be presented, the quality of the content and what added value the game media itself can bring to the educational table. In our work we have found that it is also important to consider how the previous familiarity of the student with the game medium (or lack thereof) affects a learning experience that is delivered by a game.

Games require students to be completely involved in the experience. Game-based learning is not about lecturers or a film or video pouring information into passive students, but rather a stream that requires the learner to come and drink. Choices must be made in real time. There exists an engagement that is tangible, utilizing the learner's body and mind in ways traditional learning rarely does. Learners using games are not sitting at a desk taking down notes or listening attentively (as in the best traditional situations) but are actively engaged in a dialogue. The learner's actions are key to the lessons being taught. Though it may be mediated by a machine, if the content is well designed, the game learning experience can provide exceptional learning opportunities.

2. What We Did at ICT?

The work described in this chapter comes from a research program within the University of Southern California's Institute for Creative Technologies (ICT). The ICT was formed in 1999 from the United States Army's interest in harnessing the talent and methods of the Hollywood entertainment community, toward new research efforts to improve training for soldiers. The Army was focused on two issues. First, soldiers needed training beyond basic boot camp and skills instruction (shooting guns, driving tanks, working in teams, etc.) to deal with complex situations on the battlefield. Cognitive or "soft" skills were not being fully addressed in traditional training programs, leaving soldiers lacking in such areas as cultural awareness, negotiation skills and improvisational problem-solving when presented with situations impossible to address with standard protocols.

The second issue of concern to the Army was that young soldiers were, like others of their generation, enamored of entertainment media, especially films and videogames. How could the magic of Hollywood – its ability to engage an audience member – be applied to training to make it more effective?

In its first 5 years, the ICT initiated several projects that addressed the Army's concerns and desires. One of these was a program to both create and evaluate more effective immersive virtual reality environments that could be used for training. That program, called the Sensory Environments Evaluation Program (SEE) was set up to explore methods that could result in multi-sensory and cognitively real virtual environments that were also emotionally compelling (Morie et al., 2002). SEE's core approach was that emotional connection was the key to unlocking the magic of Hollywood engagement, and that this aspect was often missing in the way virtual training environments, or simulations, were designed (Rizzo, Morie, Williams, Pair, & Buckwalter, 2005).

Digital training simulations have been important to the US Military for decades, primarily because they can replace costly live action exercises, in which ammunition alone can cost up to tens of thousands of dollars per exercise. Much research has gone into increasing the fidelity of simulations, as measured by both photo- and physics-realism. The latter is frequently at the forefront of military concern, to ensure no "negative training" ensues from faulty ballistics trajectories or inaccurate effects of a specific bomb (Page & Smith, 1998).

As cognitive training goals became more desirable, absolute photo- and physics-realism became less important than a mental picture of believability. Everything in the system needed instead to produce a net effect of the user feeling as though he actually was in the place being simulated. Such an effect is called "achieving Presence". Presence is generally considered to be the subjective state where a participant who is exposed to a technologically delivered experience is unaware of the technology, accepting the experience as he or she would a similar real experience. Presence is a desired goal of most virtual environments, but achieving Presence can involve a complex interweaving of many factors, some of which are still not fully understood. These can include degree of realism in the visuals being presented, number of different senses or modalities (sight, sound, smell, touch), richness of the environment, degree of interaction permitted and techniques to more fully involve the participant, such as including content with emotional valance. There are also various types of Presence, such as *spatial*, where a person feels physically within a virtual space, and *social*, where interaction with others (virtual or real) in the environment provides a strong sense of connection and therefore immersion in the environment. Because of these factors, there is still intense debate about an exact working definition of Presence in the research community that studies it (Lee, 2004). Despite this, there have been several attempts to develop tests that can help determine if someone has achieved such a sense. Most notable are the two associated tests designed by Army Research scientists, Witmer and Singer (1998). The first is called the *Immersive Tendencies Questionnaire*, which is administered before a person enters into an experiment in a virtual environment. It gives an indication of how susceptible that person is to becoming immersed in various forms of media. After the experiment the *Presence Questionnaire* is administered and is designed to give a measure of how much Presence a person experienced. Witmer and Singer not only developed the tests but also established a high degree of reliability based on analysis of test data from several experiments forming a large sample base (Witmer & Singer, 1998).

We believed we could increase the degree of perceived Presence in measurable ways by adding a focus on the user's emotional engagement with the simulation. This concept formed the basis for a series of experiments of the SEE Project performed between 2003 and 2005. What we found was surprising and has very much to do with the digital game zeitgeist prevalent in today's world.

3. The Nature of SEE's Experiments

The SEE Program focused on the individual experience of a soldier trainee within a (computer-mediated) hostile environment. Since we knew that the ultimate purpose of this mediated environment would be to substitute for live training experiences, we were especially interested in the concept of choice and the trainee's ability to exercise free will within the space. While most experiments are designed with rigidly bound conditions, we felt that training for cognitive decision-making skills must require freedom of choice within the situation. This decision made our experiments a challenge to design. For our initial investigations we created a simple "mission" that resembled a nighttime reconnaissance task and emphasized observational skills and stealth (not being discovered) over a series of more traditional pedagogical goals to be met.

3.1. Approach

We had several questions we wanted to investigate with SEE. The underlying goal was to create a virtual experience that evoked emotional responses in the participants and felt 'cognitively real', which we defined to mean that, for all *cognitive* effects, the virtual world was accepted as a real experience. We approached the design process using an iterative series of small pilot studies. When we were satisfied we were at least in the right "emotionally rich" ballpark, we had to decide which of the many potential variables would make the best focus for our first tests.

Self-reports of emotional reactions tend to be less direct than in situ measurements, and so we decided to monitor users' physiological arousal as an indication of emotional state (Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2004). As described above, the task was initially designed to engage participants in an observation task, but when our pilot study was run (six male, four females) it seemed to be too open-ended for them to get really engaged. A concrete task was needed to give users a stronger sense of purpose. We therefore included an action for the user to perform as a result of his/her observations. If he/she reasoned that there were hostile forces in the area, the user was instructed to place a marker by the main building.

We also realized that the participant needed adequate instruction to understand the kind of situation they were going into and a backstory that would permit them to "buy into" the situation. We produced an instructional video to deliver a "mission briefing", and constructed a hypothetical but plausible story with the help of a professional interactive narrative writer, about suspected rebels in a remote area of the Balkans. This backstory, along with the mission details, were delivered to participants after they had been fitted with the virtual reality (VR) equipment, in order to keep them "in story" from video briefing to finish. This instruction video, also termed the instructional set or "priming" had a twofold purpose. First and foremost was the description of the participant's required tasks. Second was the thematic focus of the manner in which those instructions were delivered. For reasons described below, we presented one instructional video that advised participants they were going into a serious military mission and another version that invited them to play a game. This themed priming was extensive. Participants, randomly assigned to each priming group, were brought into a room decorated with either military posters or game posters. These posters were never explicitly referred to, but they were in plain sight while the participants filled out their consent forms and questionnaires.

3.2. Methods

Our virtual reality scenario was run on a mobile three-node PC cluster in full stereo mode with 5.1 spatialized 3D sound. Participants viewed the virtual environment through a Kaiser Electro-Opticals ProView XL-50 head-mounted display (HMD) with a functional 50-degree field of view. Worn on the head, this display contains a pair of optical lenses that deliver a right and a left eye view of the virtual world, updated 30 times per second, permitting a full stereo view of the virtual objects, buildings and land features. Participants moved through and interacted with the environment using a Polhemus magnetic 6 degree of freedom tracking system. This device records how a participant moves his/her head and body within the virtual reality environment so that the visuals are calculated according to where the participant is looking and moving in the system. A tracking unit on the head sent information about the side-to-side rotation of the person's head as well as the up and down rotation. This enabled us to know where they were looking at any time throughout the experiment. Navigation was accomplished with a small, game-like controller, which gave a person the ability to move forward and back as well as side-to-side via a small joystick-located on the device. With this information we had a record of all movements for each participant for later analysis and playback.

The results of our early pilot studies were surprising: participants were not taking advantage of our fully immersive VR system. They did not turn their heads or their bodies to look around, nor did they look up or squat down to get a better look at objects. The sort of experience to which we were exposing people was usually presented straight ahead, on a screen. Users needed to be taught all the capabilities of the VR system, and so we created a virtual "tutorial room". With the tutorial room, we were able to familiarize users with navigation and interaction within the 3D environment, providing them a fuller range of behavior.

Once we were sure the system was fully functional, we began evaluating our first participants: healthy adults between 18 and 40 with good vision, drawn from the local university population.

3.3. The Environment and the Mission

Detailed design of an environment that was both realistic and provided a cognitively engaging task was challenging. However, a byproduct of our process was a design approach for creating engaging computer-mediated environments that involved a gestalt of many factors working in concert to provide a situation that *felt real* to the participant. That design approach is detailed in Morie and Williams (2003).

The virtual environment resembled a small rural area of a typical Eastern European country. The inhabitants might be either civilian refugees or militant rebels. The reconnaissance or "recon" mission was set at night, which inspired the name "DarkCon" for our scenario.

The participant-trainee plays the role of a military scout who, after receiving the video briefing mentioned above, begins the mission at the entrance to a culvert or drainage tunnel. The scout must make his/her way through it to reach the mission target: a cluster of abandoned buildings on the far side of a river. Various clues in the environment can suggest whether the inhabitants of the area are displaced villagers or dangerous paramilitary forces. The mission itself is simple: to go in, observe and if evidence suggests the rebels are in control, mark their headquarters with a small GPS locator device so an air strike force can target the hideout. If no evidence of paramilitary activity is found, the scout keeps the transmitter and returns back through the culvert to the starting point undetected, to report on his/her findings.

The difficulties of essentially allowing the user open-ended choices within the simulation were mitigated by separating the environment into distinct regions. For example, in the first segment, the user traveled through a culvert, which only afforded forward movement. Beyond the end of the tunnel, the outside area allowed for more choice in where to go, hide or explore. The final segment of the experience involved the scout being discovered by inhabitants of the area, leading to what we hoped would be a high arousal state.

To facilitate maximum immersion in the simulation itself, we chose not to talk to or aid participants during their time in the virtual environment. This decision to not talk directly to the user caused us to rely heavily on cues that were part of the environment itself to provide essential information to the person playing the role of the scout. For example, objects placed in the culvert served as evidence of refugees living there and were also chosen to elicit emotional reactions. (See Figure 1.) A broken baby doll strategically placed near the culvert entrance squeaks "mama" if stepped on. Nearby, a family photo album has been left open in the mud, with personal photos in view. Near the culvert's exit, an observant scout would notice blood spattered on the wall and bullet casings on the ground. Taken together, these and other scenario elements describe a tragic story of fleeing refugees, unable to save their most treasured belongings.

3.4. The Experiments

Our ultimate goal was to test soldiers in the scenario using their recollection of the environment and stealth (not being caught) as the measures of successful performance, but our preliminary round of testing took place with civilians, as mentioned above.

We wanted to gather general and specific user characteristics, especially those that would be informative about the degree of Presence the participant experienced in the virtual environment, using the two instruments designed by Witmer and Singer described earlier. However, the Immersive Tendencies Questionnaire, administered before the experience, had only one very general question about the participant's game play habits. Since

Figure 1. Objects in DarkCon culvert that show evidence of civilian usage.

half of all participants would be experiencing the task as a fun videogame, we decided to get more specific. We expanded the single videogame question into four questions, asking about frequency of play for first person shooter, strategy/simulation, role-playing and puzzle games. Specific questions asked, "How often do you play [one of these types of] games?" Responses were structured on a Likert scale of 1–7, with 1 being "never" and 7 "often," resulting in values from 4 to 28. We also added another question: "Do you ever become so involved in a videogame that it is as if you are inside the game rather than moving a joystick and watching the screen (respond Never if you do not play videogames)?"

We recorded participants' heart rate and skin conductance as measurements of physiological arousal state with a Cleveland Medical BioRadio 110 wireless data acquisition system. Heart rate and skin conductance sensors have been shown to be two of the least invasive and most reliable measures for arousal states in a complete study of physiological sensors done by Matthews, MacDonald, and Trejo (2005). In addition, research has shown that emotional arousal and the physiological changes that result from it affect the consolidation of memory as reported in the classic studies of Cahill and McGaugh (1998) and McGaugh (2000). High emotional arousal tends to increase activity in the amygdala, which in turn mediates hippocampal processes resulting in consolidation (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000). Newer studies have shown a distinction between the valence of an emotional stimulus and the arousal such a stimulus might provoke. A recent study by Kensinger and Corkin (2004) showed that arousal was most important in formulating this amygdala–hippocampal network, while valance alone affected other brain areas and did not have the same lasting memory benefits. We reasoned that if we could generate a certain level of arousal, we could facilitate users' long-term memories of the experience.

We had selected the wireless BioRadio device to lessen the amount of cables needed to connect the participant to the system, but a number of issues with it ultimately caused the loss of half of the data; only heart rate could be analyzed with any degree of reliability. However, there is a high correlation between the two measures (Wilson, 2001), so we felt that the heart rate measures alone could provide sufficient data for our purposes.

After participants were outfitted with the physiological sensors, they were ready to don the virtual reality equipment described earlier. Once this was done users were presented with the tutorial room, in stereo in the head-mounted display, to bring participants up to speed on the use of the equipment. An added benefit of this guided training was the final exercise, in which users found their way back to the beginning of the tutorial environment on their own. This time gave us the opportunity to record a free-walking physiological baseline for each user.

The proper beginning of the DarkCon experience was the presentation of video instructions for the DarkCon task. We felt the instructions for such a military-oriented task ought to be delivered as a "mission briefing". In part because of the sponsor's interest in videogames for training, we decided to utilize the two versions of the briefing. One version encouraged participants to treat the scenario as a military or serious mission, and the other introduced it as a cool role-playing game. The actual instructions given to all participants were identical, but presented by speakers who were dressed and acted according to military or game-oriented intent. In one instruction video, the speaker introduced himself as "Major O'Neill", addressed the participant as "soldier", projecting a stern look and grave tone. In the other, the speaker was a smiling and friendly civilian, who told participants about the fun role-playing game they were about to experience. (See Figure 2.) We hypothesized that the difference in authority of the speaker would affect the way the participant perceived the consequences of the scenario, and that the

Figure 2. Game and serious instructional videos.

participant would be 'primed' to treat the experience as trivial or serious appropriate to the type of instructional video received. Participants were told the nature of their task (previously described), told that they would have 30 min to complete it and instructed to be both cautious and observant. From that point on, there was no further interaction between the experimenter and the participant.

Participants' physiology and behavior were recorded throughout the entire experience. The DarkCon task was designed to be nearly impossible for even fully trained soldiers to complete without being caught. This was done to ensure that all participants would be exposed to a situation of high arousal at the end of the scenario, which could consist of shouting voices, the appearance of snarling attack dogs or stepping on a landmine.

After their DarkCon experience, participants were given an "after-action review", an open-ended spoken interview of 11 questions, intended to elicit the participant's observations from the experience, such as "What did you understand your goal to be as outlined in the video instruction?" and "How many weapons did you observe in the environment? Please describe." In this review we wanted to test the participants on what they recalled and remembered of their observations in the experience. Two variables were employed for this purpose. "Immediate Recall" measured what a participant could report immediately after the experience. "Delayed Recall" measured what they retained 1 week after the experiment, collected via a telephone interview.

3.5. Insights from the Civilian Study

Preliminary analysis of data gathered from civilians, in preparation for our study of a military population, showed an important trend; participants who had higher overall levels of arousal throughout the scenario also had higher scores on the Delayed Recall variable, in agreement with the literature regarding emotions' effects on memory consolidation.

Our initial hypothesis was that priming would be the most significant variable of interest, but we saw no significant differences between priming groups. In searching for an explanation, we realized that the civilians we tested were university students: they had no military experience to apply to the situation. Lacking a schema or set of knowledge, for behavior in a military context, priming could not have produced large differences between civilians.

We theorized that maybe, in the absence of such experience, if the scenario "looked like game, smelled like a game and played like a game" it was perceived as a game no matter how it was presented. Gaming experience therefore became a variable of much more interest.

Yet another interesting trend surfaced: all high-frequency game play groups, with one exception, fell within the higher-arousal higher-recall correlation. The one exception was first person shooter gamers. Interestingly, the interface to our virtual training environment was most similar to this game mode. It therefore surprised us that the high-frequency first person shooter participants had lower-arousal levels than other high-frequency gamers and a subsequent lower Delayed Recall score.

At first we thought that a plausible reason for this was that these people were more "game fit" than the others. After all, first person shooter games are always intense. However, an intern working with us, when he heard this explanation, suggested we did not know how this type of gamer approached playing the game. He said the way to succeed at first person shooter games was to detach and not let yourself get aroused by all the events (e.g., gun blasts, blood and violence) or you would not be able to concentrate on shooting your targets. This was an awakening moment for us. Game play characteristics might have much more important indications than we had imagined. Because of these findings we decided to focus on it more directly as an important variable of interest for the military study.

4. Soldier Testing

As the preliminary civilian testing proceeded, we were establishing contacts with a military base so we could test soldiers. We shipped our portable VR system to Fort Benning, Georgia in May of 2005 to test a select group of Army Rangers who had almost completed their training. Army Rangers are the elite force of the Army and so could be expected to be exceptionally well trained for the type of mission our experiment presented.

Anecdotally we found that overall the soldiers were much more attuned to the experiment than were the civilians we tested. A number of factors could have accounted for this condition, but the most likely is that the task was perceived with the import of any other volunteer duty in which soldiers may participate. Though they were on their own time, they maintained their conduct representing the Army. Their demeanor was as cooperative as if our team had been composed of their immediate superiors.

Serious priming, as might be expected, had a better effect on the soldiers. It was tied to increased arousal states, which will be described in more detail. Unfortunate circumstance precluded our being able to gather Delayed Recall score from the soldier participants, but the rest of the data from the 34 participants was analyzed.

We learned a great deal more about the effectiveness of our methods from the soldier tests, but we learned the most from a comparison of the soldier and civilian groups. These discoveries are discussed below.

5. Findings from Combining the Data Sets

The first and most dramatic indication of the divide between participant groups was with the Immersive Tendencies Questionnaire, which measured pre-existing individual characteristics for the simulation participants. Significant differences appeared between civilians and soldiers on this test. Soldiers had a total Immersive Tendencies score on average 5.13 points lower than civilians ($t[80] = -2.094,\ p = 0.04$). In addition, soldiers scored on average 2.46 points lower than civilians on the "involvement"' subscale of the test. This subscale measures a person's susceptibility to becoming deeply involved in pursuits like watching movies or reading books. Because of these significant results, the Immersive Tendencies Questionnaire score was therefore included in all subsequent analyses. This finding indicated a priori that soldiers might become less immersed in the DarkCon virtual environment than their civilian counterparts.

Civilians and soldiers also recalled the environment differently. Soldiers remembered significantly more of their goals than civilians ($t[77] = 3.444,\ p = 0.001$). They also correctly remembered more of the people in the virtual environment ($t[77] = 3.141,\ p = 0.002$). Though not significant, soldiers tended toward correct recall of more items in total ($t[77] = 1.829,\ p = 0.07$). Soldiers also spent significantly more time inside the culvert than civilians ($t[74] = 2.023,\ p = 0.03$).

During the in-culvert epoch, soldiers in the serious-primed group had the highest maximum heart rates ($t[41] = 2.610,\ p = 0.013$). During the outside-culvert epoch, serious priming had a significant positive effect on mean heart rate of all subjects ($t[41] = 2.333,\ p = 0.025$).

The most striking results appeared in the end-of-scenario period, an intentionally stressful period during which the subject knows he has been caught. Serious-primed subjects had significantly higher mean heart rates during this epoch ($t[41] = 2.210,\ p = 0.033$). They also had higher minimum heart rates ($t[41] = 3.056,\ p = 0.004$). Soldiers had significantly higher maximum heart rates than civilians ($t[41] = 2.019,\ p = 0.05$). The total Immersive Tendencies Questionnaire score had a negative effect on mean heart rate during end of scenario ($t[41] = 2.952,\ p = 0.005$).

Over the entire scenario, serious priming had a significantly positive effect on mean heart rate ($t[41] = 2.144,\ p = 0.038$). Soldiers in the serious-priming group had the highest maximum heart rates over the entire scenario ($t[41] = 2.058,\ p = 0.046$). The total Immersive Tendencies Questionnaire score showed a negative trend that was nearly significant on overall mean heart rate ($t[41] = 1.974,\ p = 0.055$).

First of all, these results demonstrate that the priming manipulation was successful. The hypothesis that the context from which an individual approaches a training environment will affect his/her performance in that environment was supported, in terms of both recall and physiology.

Second, these results highlight the differences between the civilian and soldier populations. It is unclear whether these differences arise as a result of self-selection for military service or as an actual effect of military training. However, these differences in population should be taken into consideration when evaluating training tools for military use. Results from civilian evaluations should not be assumed to be generalizable to soldiers. This requires that tools being prepared for use by soldiers should be tested on soldiers – research must have access to a military subject pool.

It was expected that some differences might appear between soldiers and civilians, but the number and magnitude of these differences that appeared was surprising. We believed the soldiers receiving the game-primed instruction set would show higher arousal than civilians who had received the same instruction because their military training, we reasoned, would override the non-serious priming. We also believed their previous training would support higher recall across both priming groups. While differences in recall may be attributable to military training, the differences between civilian and soldier physiology do not seem explicable simply by conscious factors. Where differences were significant, soldiers invariably had higher arousal levels than civilians, regardless of whether they were primed to take DarkCon seriously or treat it as a game. Interestingly, this is the opposite of what would be predicted by the significant differences in immersive tendencies: soldiers having lower Immersive Tendencies Questionnaire scores, it might be expected, would be less engaged in the scenario, and yet they seemed to be the most engaged. We concluded that civilians and soldiers responded differently to the exercise that they were presented because of the experience that they brought to it.

Differences in priming appeared when soldiers were included because they were equipped with distinct strategies to employ for each potential presentation of the exercise. The civilians were unlikely to have developed a distinct schema for military situations that never arose in their daily life. It seems that serious priming successfully allowed participants to take the exercise more seriously – on the subconscious level at least (which is what the physiological signals measure). The differences in performance between civilians and soldiers also suggest that prior training is helpful. Let us now focus, for clarity, on the somewhat more homogenous group of soldiers.

Taking the same measures as before, priming had little effect on recall generally – not surprising, since soldiers are trained in observation and reconnaissance. In recalling the goals of the mission, however, videogame play and priming condition, while not significant, trended toward an interaction ($p = 0.092$). The effect of videogame play on recall depended upon the priming condition of the participant: when participants were game-primed, videogame play had a negative effect on their report of goals. When participants were serious-primed, videogame play had a positive effect on their report of goals. In soldiers, it seems that videogame play can be a detriment to performance, when the importance of that performance is downplayed.

Despite having similar or identical training for this sort of exercise, the priming led soldiers to react to the experience differently. When the participant was primed to treat DarkCon as simply a videogame, there may have been no reason for him to pay special attention to what was, essentially, just another game. The effects and trends observed on heart rate indicate the possibility that greater videogame play frequency led to lesser momentary arousal, both inside the culvert and during the end-of-scenario epoch. This may have occurred exclusively in these two epochs due to startling or generally stressful stimuli, which the outside-culvert epoch was generally lacking. In that case, it could be speculated that frequent game play tempered the magnitude of response to momentarily arousing events.

Especially during the end-of-scenario epoch, where we designed in several stressful events, participants with less videogame experience showed greater maximum heart rates than those with more experience. This indicates activity in the sympathetic nervous system, colloquially referred to as the “fight-or-flight” response, which is responsible for increases in arousal (Pumprla, Howorka, Groves, Chester, & Nolan, 2002). Videogame-playing participants may have been able to disengage themselves from the virtual danger of the end-of-scenario period, such that their fight-or-flight response was less dramatic. It appears that non-playing participants, on the other hand, were not able to respond in this way – the threat was perceived to be more genuine (i.e., not like a game, with which they had little experience) and therefore resulted in a typical response to danger. It is likely that these effects reflect users' subconscious involvement in the scenario. This might be indicative either of game-playing participants' reactions solely to games or their arousal characteristics in general. This demonstrates that on a subconscious level, at least, users with more videogame familiarity approach the environment differently than those with less.

6. Discussion

Our research efforts, were at best, beginning experiments that served to uncover areas needing further exploration. What they did was to clearly point out several ideas that seem to be important for those wanting to use videogames as learning tools. These ideas coalesce to the relationship of the *mode* of the videogame to the content or *domain* being presented, along with the unique *characteristics* and *background* of the learner.

6.1. Game Mode: What Do Games Buy You?

Much Army training is done with exercises that allow soldiers to rehearse what they are to do in their actual operational situations. This "Mission Rehearsal", or learning by doing, has a long-standing tradition in the military. In such training, there is no interference by any intermediate delivery mechanism. The best place for a soldier to learn lessons of combat is in a situation that either is, or closely resembles, real combat, such as boot camp, mock urban settings at controlled training centers or short tours of duty. Rehearsals in a virtual environment have been shown to be almost as effective as live training for activities that have a wayfinding or search-type task (Witmer, Bailey, & Knerr, 1995) and therefore present an area of great training interest for the military.

Games are commonly believed to be rehearsal learning mechanisms, but they are not "live training". Game modes must be examined both for what they both bring to the learning situation and what they leave out. In our work, we observed the possibility that an expert in a specific game mode may also have increased performance in a similar mediated training mode. This seems obvious and is certainly borne out by more mundane modalities, such as text. If one can read, the act of reading affords some familiarity with the mode of any lesson presented as text to be read.

Our experience has underscored for us the importance of the *game mode*, or the way in which the material is delivered (e.g., the material presented as a specific type of game) as opposed to the content of the lesson (or the *domain* being presented via the mode) and how these affect one another. Having game mode experience, a learner is likely to have certain expectations, a significant portion of which must be addressed in order for them not to feel frustrated. For example, in addition to being able to move about a game space, choose items, and explore, gamers may also expect to be able to repeat the game if they fail or die, to have cheat codes and to be able to save their experience before performing a risky move. They might want the interaction controls for the learning game to be mapped to standard devices or paradigms. In our experiments, there were several instances where we did not meet gamers' expectations. For example, a common strategy in first person shooter games is the "turkey look", using the controller to look around a corner from a safe place. Several participants commented on how this move, which we did not include, would have been useful in the scenario.

It was clear to the researchers throughout the civilian experiments that participants tended to always approach the virtual scenario as a game, despite our best efforts, via priming and design, to make it seem otherwise. Priming established how these participants would react to a game, as opposed to a training exercise, *to some degree*. It is the case, however, that the participants applied their previous understanding of videogames – to whatever extent they were able – regardless of their instructions.

There were several reasons we postulated for this. The first is familiarity. Virtual environments are so esoteric that few people have had first hand experience with them. Games are a different story, with large numbers of both children and adults playing videogames regularly. Due to this, we can assume that a large percentage of our potential participant pool had had experience with games. Of course they have also had a great deal of experience with more traditional media such as film and television. Film and television are passive media, requiring no action of the part of the audience. Games do require such interaction, which is the second reason they saw the scenario as a game: it had agency.

Agency, as defined by Janet Murray in her seminal book "Hamlet on the Holodeck", is the "satisfying power to take meaningful action and see the results of our decisions and choices" (Murray, 1997, p. 126). A chief characteristic of games is the agency they demand of people. If we had locked our participants in place they might have approached the scenario as a film, but because we gave them controls, they recognized it as a game. Reinforcing this recognition was the fact that the navigation controller was essentially the same as standard game controllers. Not only was there agency, but the method of obtaining that agency utilized a very familiar device.

During the experiments conducted with Rangers at a military base, researchers noted that the soldiers remarked far less on the game-like qualities of the virtual training environment both before and after the DarkCon experience that the civilians had. Despite this, their results supported the same postulates made for the civilian participants – soldiers who were experienced gamers were likely to be familiar with the interface, game controls and agency of the DarkCon scenario and thus have seen the environment being more a game than a training exercise.

6.2. Domains and Schema Considerations

Having looked at issues surrounding the game mode, we will now discuss some of the concerns around domain that our work has raised. The main purpose of a game for training is, of course, to present lessons to be learned in the context of the *domain* being addressed. The context for our scenario was tactical reconnaissance, a typical military task.

The background information for the task our participants were to undertake in DarkCon was given to them with the video briefing, which had either a serious military theme or a playful game attitude. The creation of the briefing was an iterative process that taught us many things about readying people for an unfamiliar task. The first serious briefing we produced was militarily accurate. We used subject matter experts to get the jargon right and took pains to ensure that it would pass muster from our Army sponsors. However, when we started using it in our pilot studies, no one understood a word of what was said. The civilian participant population simply had no related domain knowledge that would allow them to put the words into a context they understood. For our revised serious briefing video we still used an actor dressed appropriately as a Major in the Army, but toned down the language so it was suitable for public consumption. When we did the second experiment with soldiers we had to use the version of the instructional video that had militarily correct language. We used the exact same game-themed briefing for both civilian and soldier groups. This briefing contained precisely the same instructions as each version of the serious video, though presented by a different actor, with a more casual attitude and a "have fun!" closing.

Our intent in providing the two introductions was also to see if they had specific effect on what people would recall. Did making the exercise seem "frivolous" make them less likely to recall elements of the experience? Our results indicate that this was true only for domain experts (soldiers) who reported a high frequency of previous game play. Telling them it was a game and providing an interface that resembled a game cause soldiers in this (game-priming and high-game play) to report fewer of the targets indicated in the instruction set, as if they were not taking the instructions as seriously. We also wondered if we would see differences in arousal states (physiological measure) based on this variable. Could you introduce a lesson as a game in which people are supposed to learn or remember something and still achieve your desired effect? Even though the DarkCon exercise was one soldiers were familiar with from their military training, introducing it as a game produced undesired results, such as low arousal and low recall. And finally, would a more serious introduction to the experience provide a student greater motivation to learn? Our work suggests that in the case where the experience is game-like and the participants have both game experience and domain training (as our high game frequency soldiers), a serious introduction will produce more desired results, such as high recall.

To recap, we wanted to see if the serious introduction to a serious learning experience would produce serious results, or if the playful introduction of this serious experience would result in it being approached as a game and, given a familiarity with the game mode, have improved results. We discovered that the best performance by participants in our DarkCon environment was where the domain knowledge best matched the learners' schema *and* their familiarity with the game mode was also high. The "performance" of the military group was better because DarkCon at its base was military in nature and they were well versed in that schema. Participants coming in with military training and a high frequency in game play were well suited for the tasks to be carried out in the environment and exhibited the best performance *except* in the case where they were made to believe it was "just a game" (game-primed group).

To generalize, the best situation is when the learner knows enough about the domain to benefit from the lesson *and* enough about the game mode that it contributes to, rather than obfuscates, the lesson. The worst result should be seen when the learner is in an unknown domain using a game mode with which they have no experience. Between these two extremes there are possibilities we suggest might happen, but further work should be done to confirm these intuitive assumptions. When the learner knows something about the domain but is Not a gamer, then he/she may become frustrated by the game mode. If, however, the learner is a gamer playing a game to learn, he/she may not get the unfamiliar domain right away, but perhaps repeated tries will bring him/her up to speed quickly.

7. Conclusions

7.1. Are Games a Significant Learning Tool?

Our work indicates that games are a potentially important learning tool *if* the player is familiar with the game mode and the domain is matched to their level of expertise. Games used for learning are still games, however (even though in recent years they have become designated as "serious games" for the purpose of public debate and funding opportunities). The trick is to take advantage of what they offer as games and to look at how the game mode itself fits into the full learning picture. Games used to teach may be either frustrating *or* advantageous for the learner. If the user is familiar with the game mode and the learning tool has been designed to match the gamer's expectations, then the learning game has the best chance for success.

However, our investigations have raised a great many more questions than they answered and suggested new avenues for further study. Such ideas may be well known in more familiar learning domains but are not well understood when it comes to videogames used for learning. We will list a few of the questions we would like to see explored in more detail in the future.

7.2. Future Work

Our first suggestion for continued research asks the question: Would a more closely coupled game medium and mode result in improved performance by participants with experience in both?

We were not able to run participants through the DarkCon environment using a more traditional videogame presentation device, which would enable us to determine the effects of mediating the experience with VR equipment. The tutorial environment designed to familiarize our participants with the VR medium was an attempt to compensate for this inability, but to establish a robust baseline with which to continue investigations in this vein, a more exacting definition for the relation of medium/mode experience and performance in an unfamiliar domain must be determined.

It would also be interesting to investigate if repeated exposures to the learning environment would reveal an advantage of either domain experience or game mode experience in rate of improved performance. Presenting the same subject matter over and over in games is a convention of most game modes, and therefore may appear to give participants that are more familiar with games, but less familiar with the domain, a predictable upper-hand over those with inverse characteristics. The same, though, could be said of instruction across many domains using more traditional media and modes; there are few learning environments where a student is expected to "get it" the first time. The "gamers learning more by way of games" intuition is also threatened by our own delayed recall data. Civilian participants that experienced a heightened degree arousal exhibited greater delayed recall, but the same group's familiarity with games was revealed to have a negative impact on arousal state. While a single exposure to the environment showed no significant relationship between game experience and delayed recall, repeated exposures might create a familiarity with the domain that overcomes the relationship between game experience and lower degrees of arousal, yielding improved delayed recall.

This leads to yet another question: at what point are participants in a game-based learning environment only learning to "game the game"? This remains the fundamental question for research on games for education and training. Participants can only be known to have acquired worthwhile knowledge in a domain if they can apply this knowledge within the actual medium in which the domain is practiced. Our studies indicate that a pre-existing familiarity with the domain trumps a pre-existing familiarity with the game mode in terms of single-exposure performance. Regardless of how multiple exposures affect this result, only when participants are once again (or for the first time) immersed within the native medium and mode of the domain will their performance measures within the game-based learning environment be truly understood.

Acknowledgment

The project described herein has been sponsored by the US Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the US Government; no official endorsement should be inferred.

References

Cacioppo, J. T., Berntson, G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2004). The psychophysiology of emotion. In M. Lewis & J. M. Haviland-Jones (Eds), *Handbook of emotions*, 2nd ed. (pp. 173–191). New York: Guilford Press.

Cahill, L. & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neuroscience (TINS), 21*(7), 294–299. Irvine, CA: Elsevier Science Ltd.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience*, 20 (RC99), 1–5.

Chess (2006). In *Encyclopedia Britannica*. Retrieved December 5, 2006, from Encyclopedia Britannica Online: http://www.britannica.com/eb/article-9105867.

Fairbairn, J. (1995). Go and war. Retrieved November 22, 2006, from Go in Ancient China, http://gobase.org/reading/history/china/?sec=part-3.

Kensinger, E. A. & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Science*, USA 101: 3310–3315.

Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*, 27–50.

Matthews, R., MacDonald, N. J., & Trejo, L. J. (2005). Psycho-physiological sensor technologies: An overview. In D. Schmorrow (Ed.), *Foundations of augmented cognition*. New Jersey: Lawrence Earlbaum, Associates Publications.

McGaugh, J. L. (2000). Memory – A century of consolidation. *Science*, *287*(14), 248–251.

Morie, J., Iyer, K., Valanejad, K., Sadek, R., Miraglia, D., & Milam, D. (2002). Emotionally evocative environments for training. *Proceedings of the 2002 Army Science Conference*.

Morie, J. & Williams, J. (2003). The gestalt of virtual environments. *Proceedings of the International Workshop on Presence 2003* (Abstract only published. p. 51).

Murray, J. (1997). *Hamlet on the Holodeck*. New York: The Free Press.

Page, E. H. & Smith, R. Introduction to military training simulation: A guide for discrete event simulationists. *Proceedings of the 30th Conference on Winter Simulation* (pp. 53–60), December 13–16, 1998, Washington, DC.

Pumprla, J., Howorka, K., Groves, D., Chester, M., & Nolan, J. (2002). Functional assessment of heart rate variability: Physiological basis and practical applications. *International Journal of Cardiology*, *84*, 1–14.

Rizzo, A., Morie, J., Williams, J., Pair, J., & Buckwalter, J. (2005). human emotional state and its relevance for military VR training. *11th International Conference on Human–Computer Interaction*, Las Vegas, NV.

US Department of Defense (2003). *Military and associated words*. US Department of Defense.

Wilson, G. F. (2001). An analysis of mental workload in pilots during flight using multiple psychophysiologic measures. *International Journal of Aviation Psychology*, *12*(1), 3–18.

Witmer, B. G., Bailey, J. H., & Knerr, B. W. (1995). *Training dismounted soldiers in virtual environments: Route learning and transfer*. U.S. Army Research Institute for the Behavioral and Social Sciences Technical Report 1022.

Witmer, B. G. & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.

SUMMARY AND DISCUSSION

Ray S. Perez

Abstract

This chapter summarizes each of the prior chapters in the order of the table of contents, followed by a discussion and conclusions. The chapter melds both results of the empirical game literature in general and the chapters in this book specifically. In general, authors conclude that there are few empirical game studies and of the empirical studies, the results are mixed. In addition to theoretical frameworks and reviews of game literature, the contribution of the authors has been to document the results of nine empirical game studies. These empirical studies were reviewed and lessons learned developed.

This chapter begins with a summary of each of the prior chapters in the order of the table of contents, followed by a discussion and conclusions. The book was divided into four sections: framework, team learning, individual learning, and summary/discussion.

1. Framework

There were four chapters in the framework section. Each chapter was meant to be mainly a think piece to provide context for the remaining chapters. Topics covered were an overall conceptual framework, assessment and formative evaluation, and a view of games from a software perspective.

1.1. A Conceptual Framework for the Empirical Study of Games

The first chapter in the framework section, by Drs Robert Tennyson and Robert Jorczak, lays out a conceptual framework to study the impact of games. According to the authors, this chapter was motivated in part by three factors. The first is the observation that players of complex and highly interactive entertainment games are willing to spend large amounts of their personal time in game playing. Second, there is a lack of empirical

Computer Games and Team and Individual Learning
Edited by Harold F. O'Neil and Ray S. Perez

studies on the impact of games on learning and performance. Thus, current research has little empirical guidance for the developers of educational games. Third, there is a lack of a conceptual framework to focus research in the development of effective educational games.

The conceptual framework described in this chapter is based on an information-processing learning theory and the author's earlier work (Tennyson & Breuer, 1997, 2002). Their Interactive Cognitive Complexity (ICC) theory includes learner affect, cognitive strategies, a knowledge base, and executive control of internal processing components. The authors use this model to identify key variables or features of games that are felt to promote learning. Their conceptual framework of the Interactive Cognitive Complexity theory was used to identify key variables and features of games that relate to their model's key components and thus to learning. Identification of key features and variables that are felt to promote learning could serve both as a guide for game developers to include in their game design and help trainers and educators in selecting games that would promote learning.

In addition, they highlight the importance of individual differences in game player's traits, e.g., gamists, narrativists, and simulationists (Edwards, 2004). According to this research, gamers have quite distinct goals in playing a game. For example, gamists seek competition and challenge, narrativists seek story and characters, and simulationists like exploration and experience. These categories of players can be seen as having a primary motivation tied to the major components of the Interactive Cognitive Complexity framework: gamists are primarily motivated by aspects of game goals (problems), narrativists by game context, and simulationists by game interactions and representations. Although no empirical proof is offered for these three categories of game players, the authors argue that these game player traits may play an important role in determining the effectiveness of games. Therefore, games that offer explicit goals may be more motivating and thus more effective to gamists and less so for narrativists and simulationists. Lacking are measures of these traits, thus limiting their utility. However, the chapter is interesting in that it lays out a potential R&D agenda based on a creative conceptual framework.

The strength of their conceptual framework lies in the characteristics of games and less specifically at how one would measure the processes and outcomes of game playing. The next chapter in this book provides a conceptual framework for these assessment issues.

1.2. A Framework for the Assessment of Learning Games

This chapter, by Dr Eva Baker and Ms Girlie Delacruz, provides some of the most interesting and creative seminal ideas and issues for the assessment of learning in games. Issues discussed include design requirements, outcome and process measures, new concepts of validity, use of wrap-around approaches to evaluating existing games, and the notion of using a game for assessment when the initial training intervention was not a game.

The chapter focuses on specific use of games – the assessment of the learner. These authors propose that an assessment system should be concerned with more than the measurement of the player's knowledge of the game's mechanics and the knowledge of the game's topics. Assessment should also be concerned with the extent to which the game develops and sustains the transfer of skills, strategies, and other social behaviors (e.g., perseverance) to other games and other settings (e.g., on the job).

With respect to the required technical features for game assessment they outline the process to ensure that tests are aligned with what the game is purported to teach. For example, they discuss in some detail the methods for determining the cognitive demands of the game. They describe two different approaches for model-based assessment. First, the author's Model-Based Assessment approach (MBA, Baker, 1997) identifies five families of learning that include key learning outcomes for education and training. These are content understanding, problem solving, teamwork/collaboration, communication, and self-regulation. The second approach, Evidence-Centered Design (Mislevy, Steinberg, Breyer, Almond, & Johnson, 1999), begins by performing a cognitive task analysis which results in a student model (i.e., what set of knowledge skills or attributes are to be assessed). The student model specifies the variables the student is expected to demonstrate proficiency on. The model is framed in a mathematical structure (Bayesian nets) which provides probability estimates of the relationships among the variables of interest. Both of these models require that the process of assessment include a clear definition of the nature of assessment processes as well as outcomes. Such process data could also be useful in the diagnosis and remediation of errors in game play.

For assessment outcomes, they suggest the value of wrap-around assessments. Many of the chapters in this volume using commercial off-the-shelf (COTS) games as their experimental treatment have designed wrap-around assessments based on the authors' Model-Based Assessment approach (see chapters by Chen & O'Neil, Shen & O'Neil, and Wainess & O'Neil). Examples of wrap-around assessment instruments include use of computerized concept maps where measurement of a game player's knowledge is assessed before and after game playing. Such players' maps are scored in real time against an expert's map.

Another very intriguing idea of the authors is the use of games to serve as a proxy for typical tests, such as on-the-job performance tests. The value of such a game/test would be enormous in that it has the potential to provide trainers in the operational community (e.g., on the job) with the ability to create game scenarios that would require the trainee to exercise the knowledge and skills that s/he has acquired in other settings (e.g., schoolhouse). Another practical use of such a game would be to measure the speed and agility with which a learner acquires a new set of skills in an unfamiliar game environment. Such an assessment of a learner's speed and agility in learning new tasks could be key variables in predicting whether learners will acquire new skills and knowledge on the job.

1.3. A Formative Evaluation of the Training Effectiveness of a Computer Game

This chapter, by Drs Hsin-Hui Claire Chen and Harold F. O'Neil, describes a formative evaluation study of the COTS game *SafeCracker* (Daydream Interactive, Inc., 1995/2001). They argued that the game was assumed to teach problem solving and such problem solving could be measured by a wrap-around problem solving assessment. This evaluation study consisted of a pilot study where they refined their design, training protocols, and measures of problem solving and a main study which consisted of implementing the findings of the pilot study and conducting the evaluation. This chapter is very interesting and timely; as mentioned previously there has not been very much research on the effectiveness of games and the research that does exist does not address the issue of how to improve the effectiveness of games and what assessment framework should be used to evaluate their utility.

They applied the formative evaluation model developed by one of the authors (O'Neil, Baker, & Fisher, 2002). The purpose of this evaluation model is to determine the feasibility and utility of a game and improve it by providing information on its impact on problem solving. Based on Mayer and Wittrock's work (1996), the authors defined problem solving as a cognitive process directed at achieving a goal when no solution method is obvious to the problem solver. In their view problem solving consists of domain knowledge, problem solving strategies, and self-regulation. In their pilot study they paid careful attention to the development and refinement of reliable and valid instruments to measure each of the components of their problem solving model. They also included as dependent variables retention and transfer adapted from the work of Mayer and Moreno (Mayer & Moreno, 1998; Moreno & Mayer, 2004) where these researchers assess transfer by counting the number of predefined major idea units stated by subjects regardless of wording. The development of valid and reliable measures of problem solving is perhaps the most important aspect of this study. Such approaches could be used in the assessment of other problem solving COTS games.

They note that the *SafeCracker* game, as most COTS games, as it presently exists does not have embedded in its design any effective instructional strategy. This is an important point because the authors predict that games that lack an effective instructional strategy will not produce significant learning. This prediction is based on the assertion by media researchers (e.g., Clark, 1983, 1994, 2001; Kirschner, Sweller, & Clark, 2006; O'Neil, Wainess, & Baker, 2005; Wolfe, 1997) that media themselves do not facilitate learning but the embedded instructional strategy does. The results of their main study met their expectation that *SafeCracker* would produce very little learning (i.e., 2–4% compared to experts) because the game lacked an effective instructional strategy in its design. The main finding of this research is that playing a commercial game had small but significant effects on college students' problem solving. The generalizability of these results is of course limited given the lack of an experimental design (e.g., random assignment and a control group).

According to these researchers, if effective instructional strategies had been embedded in the game, learning gains would have been far more dramatic. Perhaps the most important aspect of this evaluative study is the development of valid and reliable wrap-around assessment instruments to measure components of problem solving; content understanding (knowledge maps), problem solving strategies (the retention and transfer questions), and self-regulation (the trait self-regulation questionnaire).

1.4. Foundations for Software Support of Instruction in Game Contexts

This chapter, by Dr Allen Munro, describes the architecture of *iRides* and the design of an Application Programming Interface (API) of a simulation engine that could serve as a model API for use in instructional games. He views games as simulations that are enjoyable. He states that important game characteristics are voluntary play, pursuit of a goal state, constraints and/or rules, and competition – which may include competition with oneself. His claim is that it is all of these characteristics that make games enjoyable. The author brings up a very important point about computer games – that they are built for entertainment, not for education or training per se. He then asks the question, "Would a game provide good value if it could be quickly mastered?" How many game buyers would buy a game where they could find that they saw everything and attained maximum performance in short order? (p. 57) The entertainment value of a game according to this author and others in the game industry requires that the goals of the game not be achieved too early, so that the consumers' expenditure on the game provides the value of many hours of diversion.

The author argues that the goal of game developers is not to teach the players the rules, characters, roles, strategies, and behavior of a game in the most direct and efficient manner. Rather, the goal is to teach players just enough to make the game interesting and challenging to maintain their motivation and keep them playing the game. Here we have a clear case of competing goals, education vs. entertainment. Training designers presumably would design the game to teach knowledge more directly and efficiently, whereas game designers are more interested in enabling game players to discover through unguided experiences the essential elements of game playing (e.g., rules, characters, roles, strategies), preserving the entertainment value of the game. However, such a discovery instructional strategy is not an effective one (Kirschner et al., 2006).

The challenge for designers of games for training or educational purposes is to increase the training and educational value of a game without decreasing the game's entertainment value. The author provides some suggestions as to how effective instructional techniques can be added to the game context. One suggestion is the development of a universal set of software services that might be provided for games through the use of a universal API. These services would be in support of his eight illustrative types of learning objectives for game contexts. For example, a learning objective would be learning causal relationships qualitatively and quantitatively or learning to solve novel problems in the game context by applying strategies. He then provides examples of how these eight different types of learning objectives can be taught in a game context using one or more teaching or training strategies devised from Sweller's Cognitive Load Theory (Sweller, 1988) and Clark's Guided Experiential Learning Design system (Clark, 2004). His chapter is an interesting and creative blend of software engineering, cognitive science, and instructional design techniques.

2. Team Learning

There were four chapters in the Team Learning section. Each chapter was meant to highlight an issue in team training or assessment. These issues were the following: (1) eliciting teamwork skills with a modified COTS game, (2) investigating the use of intelligent agents to replace team members, (3) investigating the role of game experience on motivational outcomes of games, and (4) suggested guidelines for multiplayer games.

2.1. Eliciting and Evaluating Teamwork Within a Multiplayer Game-Based Training Environment

This chapter, by Drs Talib Hussain, Shawn Weil, Tad Brunyé, Jason Sidman, William Ferguson, and Amy Alexander, was principally motivated by a thesis offered by General Paul Gorman (US Army retired). Gorman's gambit, as it was called, is that teamwork skills can be taught effectively to military personnel using COTS multiplayer games, and that there is no need for the game to be realistic with regard to modern military operations to be effective. Thus one could use a multiplayer game whose context was medieval history with some imaginary characters and weapons to train teamwork. If true, such a gambit would allow military trainers to exploit the ability of multiplayer games to engage players (military personnel) in an immersive enjoyable simulated environment to elicit complex behaviors and improve performance on targeted tasks and goals such as teamwork skills.

The study consisted of an 8-month effort involving 40 members of the US Army infantry as study participants. Although the game to be used for the study, *Neverwinter Nights* (Bioware, 2002), intentionally did not contain the level of realism of combat, for example the weaponry used by modern-day infantry, the teamwork elements required for the "siege of Camelot," one of the game's scenarios, require teamwork skills similar to those for many military operations. In order to test Gorman's gambit, study researchers performed three key tasks. First, they had to identify and define what the key teamwork skills were to be assessed. Second, they had to design and build a tutorial system to teach study participants and a communication system to facilitate teamwork. Finally, they had to develop a reliable and valid measurement system.

The teamwork skills they chose to measure were obtained from a survey of the team literature and were: adaptability, monitoring, back-up behaviors, communication, and leadership (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Freeman Diedrich, Haimson, Diller, & Roberts, 2003; Serfaty, Entin, & Johnston, 1998; Sims, Salas, & Burke, 2004; Smith-Jentsch, Johnston, & Payne, 1998; Smith-Jentsch, Zeisig, Acton, & McPherson, 1998). These skills have been demonstrated in the research literature to be linked to superior team performance and are trainable and measurable. Assessment instruments for these teamwork skills consisted of observational scales in game evaluation forms, postsession team debriefing, and postexercise questionnaires, and final After Action Review (AAR) participant feedback. AARs are a traditional form of evaluation used to assess a unit's performance of field training exercises. The intent of their AAR was to collect information to improve the game and not to improve teamwork skills.

One of the most interesting aspects of this chapter is the development process that was used to develop the tutorial system, communications system, and the measurement system. These system components were developed in an iterative fashion where the developers would build a prototype, try it out with participants, collect data on how the participants reacted to the prototype, and modify as needed based on data from participants. The process of adaptation of the *Neverwinter Nights* multiplayer game software to meet the goals of the study was no trivial matter; it was very labor intensive. Unfortunately, due to the nature of their prototype system (e.g., system reliability), the teamwork data were not collected systematically by human observers; thus, the results are suggestive and not definitive regarding Gorman's Gambit.

2.2. The Effectiveness and Efficacy of Intelligent Agents as Team Training Partners in the Acquisition of Complex Skills in a Gaming Environment

This chapter, by Drs Jonathan Whetzel, Winfred Arthur Jr., and Richard Volz, began by asking, can intelligent agents (software programs) replace human partners in team training of complex skill acquisition and learning? They investigated human members' affective reactions to the game and their level of task-specific self-efficacy as part of a team (dyad) where their partner was either an intelligent agent or a person. Team training presents to training managers many challenges in terms of cost, administration, and schedules. For example, training managers have to provide training opportunities for individuals to come together to train as a team which require overcoming scheduling conflicts, administrative issues, and cost constraints. These constraints often result in missing human team members and thus no opportunity for training for the other team members. One solution to these challenges is to substitute human team members with artificial ones. In this study intelligent agents were developed using machine language techniques and neural networks. The intelligent agent learned the game, its strategy, and game play by observing a human expert play.

The authors developed a dyadic training protocol in which participants were either paired with an intelligent agent or a human partner. Trainees in this study were engaged in two consecutive days of 2 h of training on *Space Fortress* (Donchin, 1989). This research-based computer game was designed to simulate a complex and dynamic aviation environment. Donchin and his colleagues carefully analyzed the skills involved in flying a high-performance fighter jet and then designed *Space Fortress* to teach those skills. Research using *Space Fortress* has a relatively long history as a game used to examine team training issues (Day, Arthur, Bell, Edwards, Bennett, Mendoza, & Tubre, 2005). The authors used Kirkpatrick's (Kirkpatrick, 1976) four-level model of training evaluation to access the comparative effectiveness of the intelligent agents vs. human partners. Evaluation criteria were limited to attitudinal (e.g., self-efficacy) and behavioral/performance learning measures.

The results of their study indicated no significant differences on either self-efficacy or performance measures between participants who trained with a human partner vs. an intelligent agent. However, they did observe a positive difference between pre- and posttest performance scores indicating that both groups learned from playing the game. The authors concluded that individuals may be trained in a team context by replacing a human partner with an intelligent agent as a partner since there appears to be no decrement in individual performance or in self-efficacy. Although arguing from the null case, these findings if replicated have large practical consequences. One could conduct team training for an individual and simulate the other team members without loss of effectiveness and a reduction in cost.

2.3. The Influence of Trainee Gaming Experience on Affective and Motivational Learner Outcomes of Videogame-Based Training Environments

The chapter by Drs Karin Orvis, Kara Orvis, James Belanich, and Laura Mullin focuses on the individual learner's characteristics that influence videogame-based performance. Learner characteristics that were studied included prior videogame experience and motivational variables. The authors stated that individual characteristics of the trainee play an important role in predicting games learning outcomes. The basis for this belief comes from their review of the e-learning literature. They cite examples of studies from this literature that have found learner characteristics such as personality traits and prior computer experience to predict e-learning outcomes. They expect such findings would also generalize to the game area. The authors hypothesized that previous videogame experience would predict four learner outcomes of interest: time on task, training satisfaction, ease in the use of the game interface, and team cohesion. It should be noted that they did not have any cognitive outcomes per se.

The research reported on in this chapter included participants who were first-year US Military Academy cadets. The experimental game platform used in this experiment was *America's Army* (United States Army, 2004). Prior to training in the game, participants were administered a questionnaire that asked for their general game and specific game experience. Learner outcomes were assessed using a questionnaire that contained items that assessed satisfaction with training and an estimate of the ease of using the game's user interface. Participants were also asked to assess perceived cohesion and time on task.

This game has sections that can either be played as single person or multiplayer. The study focused on the multiplayer version. Participants were given game mechanics training that consisted of playing a single-player section where they learned the rules and how to play the game. Then each individual played the basic training phase of the game which contained four sections: (a) marksmanship training, (b) an obstacle course, (c) weapons familiarization, and (d) a military operations in urban terrain training mission. Upon completion of the basic training section, participants played in a multiplayer mode that consisted of participants placed in small teams and engaged in several collaborative missions. Each mission consisted of either attacking or defending a radio tower.

The results indicated that a general measure of videogame experience predicted positively learner outcomes of satisfaction, ease in use of the game interface, perceived team cohesion, and time on task. Prior experience with playing *America's Army* also predicted all four learner outcomes. However, prior game playing experience with games that did not share the same characteristics and features of *America's Army* did not predict the four learning outcomes. The authors conclude based on the results of their study that prior game experience whether it is in playing specific games or in general is an important variable to control in assessing the effectiveness of video games. This well-designed study suggested other research avenues to explore these issues in the context of a COTS game. They also used a wrap-around assessment strategy.

One conclusion that one can draw from this study that has a practical implication for training managers is that trainees who have prior videogame experience with games that have similar features and characteristics will feel more at ease with the game's user interface and may require less practice to mastery.

2.4. Utilizing Multiplayer Games for Team Training: Some Guidelines

The chapter by Drs Michael Curtis, Michelle Harper-Sciarini, Deborah DiazGranados, Eduardo Salas, and Florian Jentsch examined the utility of multiplayer video games to develop team skills and then suggested guidelines for team training. They reviewed a number of multiplayer COTS and serious games. Serious games are defined as games that have been developed with a specific training purpose, in contrast with other games where the purpose is entertainment. The chapter had two goals: to develop a set of theory-based guidelines for selecting or developing games that target the development of teamwork skills; and second, to describe an experimental platform for which empirical research on the use of video games for team training could be designed and executed. Perhaps the second goal is the most important for future researchers as they propose an ideal research platform where training opportunities can be created and assessed.

They identify three critical constructs to the success of teams and the development of teamwork. These three critical teamwork constructs were (1) communication, (2) coordination, and (3) team leadership. Such constructs can be trained and assessed in serious multiplayer games. The two most prominent characteristics of multiplayer games that set them apart from single player games are the modes of communication and the elements of coordinated behavior to accomplish game tasks.

In order for games to ideally teach team tasks they must have voice-supported communication that enables individual players to communicate with each other. This feature was used to teach team skills by other authors (see Hussain et al. in this volume). They also list examples of multiplayer games that have this communication feature. The second feature that is essential for success in team training is coordination. The game must have the capability to ensure that coordination of behaviors is targeted and that there are game scenarios that support task-interdependent tasks. This would involve assignment of different roles with specific capabilities to each player and that the task could only be accomplished by the players coordinating their activities.

The last section contains a list of 11 guidelines that research and expert opinion have demonstrated have utility in training teamwork skills. These guidelines are intended to be used by trainers and training developers in selecting, evaluating, or developing games to teach teamwork skills. For example, to ensure that the game has the capability for team members to engage in information exchange (i.e., communicate), the game feature that must be included is communication mode. They also provide a list of potential video games that can be used for team training.

The authors acknowledge that there is a paucity of research on the empirical effectiveness of their guidelines and there is a need for further research before their proposed guidelines can be taken as more than advisory. In particular, they suggest the need to determine the degree of transfer of team skills acquired in a game to an actual task and job. Research is also needed to determine the types of skills that are needed to complete a team task within a game context that includes both taskwork as well as teamwork variables.

3. Individual Learning

There were six chapters in the Individual Learning section. The length of the section reflects the empirical reality that most games research is on individual learners. There was also a focus on COTS games.

3.1. The Effects of Changing Resources on Game Performance, Subjective Workload, and Strategies

This chapter, by Dr Adrienne Lee and Aaron Perez, describes research on individual subjective workload in the context of a gaming environment. It is an excellent example of investigating psychological variables in a COTS game context. The experimental platform used in this research is the game *Real War: Rogue States* (Rival Interactive, 2002) and is based on a Joint Chiefs of Staff training simulation for the United States military. In this game, the objective for the player is to build up an army with an allotted amount of financial resources, maintain the army, and find and eliminate enemy targets.

The main focus of this research is measuring changes in game performance when participants either received a change in the amount of resources (from high to low or low to high) or no change at all (high for both tests, low for both tests) in a dynamic and unpredictable gaming environment. The assumption was that level of workload can be manipulated and affected by varying the amount of resources available. This reasoning was based on earlier research that demonstrated that when individuals are given fewer resources, perceived workload is increased. The experiment was designed to simulate an environment where people are assigned different amounts of resources, and tested on whether changing the amounts of resources, actual workloads, affects either performance, perceived subjective workload, and the strategy used. The authors make the distinction between actual and perceived workload. Subjective work load is defined as "an individual's personal view as to how much effort he/she is expending or encountering at a particular point in time" (p. 166). Actual workload "was determined by the amount of resources available for allocation by participants" (p. 168).

The results of this study demonstrated that perceived workload could be manipulated via resources. For example, participants who had high resources initially had difficulty transferring to a condition where they had fewer resources vs. participants who started out with low resources and transferred to low resources. The rationale for these findings was those participants with high resources not only had to be concerned with the game (low workload), but those subjects in the low resources condition had to be concerned with developing a strategy to deal with the game with a limited amount of resources. However, they did find that reducing workload may not result in better learning. It is clear from their data that a relationship does exist between perceived workload, actual workload, and performance. The directions of that relationship await further study.

An implication for the training community is that the level of perceived workload during training must be accounted for; if it is too low it may not facilitate performance on the job where levels of workload may vary.

3.2. Role of Worked Examples to Stimulate Learning in a Game

The chapter by Drs Chun-Yi Shen and Harold F. O'Neil described an effort to improve learning in a game-based environment using an effective instructional strategy, i.e., worked examples. Worked examples are defined by Sweller and his colleagues (Sweller, 1989, 1990; Ward & Sweller, 1990) as procedures that focus on the problem states and associated operators (i.e., solution steps) that enable students to induce generalized solutions or schemas. Such worked examples as procedures enable trainees to successfully solve the problem task. Worked examples are hypothesized to manipulate aspects of cognitive load theory. Such effective instructional strategies are lacking in most COTS games (e.g., Chen & O'Neil, this volume). The authors also provide an excellent comprehensive review of the literature on the effectiveness of worked examples.

Their purpose in this chapter was to investigate the effectiveness of worked examples on problem solving in a game-based environment. A model of problem solving developed by the Center for Research on Evaluation, Standards, and Student Testing (CRESST) was used to assess problem solving of the participants. The model has three components: (a) content understanding, (b) problem solving strategies, and (c) self-regulation.

The research consisted of a pilot and main study. The pilot was to study the feasibility of experimental procedures, measurement instruments, and the design of the instructional materials. The main study was used to evaluate the effectiveness of worked examples in a game-based (i.e., *SafeCracker*, Daydream Interactive, Inc., 1995/2001) problem-solving task. College students were randomly assigned to two conditions, experimental with worked examples and a control group without worked examples. The COTS game *SafeCracker* is a computer puzzle game that requires players to find clues, apply these clues, tools, and knowledge, and solve a puzzle in order to open a series of safes in a mansion. The experiment consisted of two game-playing sessions. In the first session participants were asked to fill out the self-regulation questionnaire, then knowledge maps, and then a problem solving strategy measure. In the second game-playing session the procedure was the same except participants in the experimental conditions studied the worked examples.

The results provided evidence that worked examples enhanced problem solving in a game-based environment. Participants in the experimental condition improved significantly on the knowledge maps and on the problem solving retention tests and transfer than the controls. The results of this study provided additional support to the thesis that adding a wrap-around instructional strategy and assessments to a COTS game will improve learning.

3.3. Training Visual Attention with Video Games: Not All Games Are Created Equal

This study by Drs Julia Cohen, C. Shawn Green, and Daphne Bavelier examined the effects of playing video games on visual attention. Although visual attention has not been viewed by many as a higher order cognitive process, Bavelier and her colleagues have provided evidence from the research literature and their own work that suggests that visual attention is not simply a bottom-up process involving only the perceptual sensory system but rather an interaction between perceptual processes and executive control. This chapter focused on increasing the temporal resolution of visual selective attention and the number of moving objects that can be attended to simultaneously over a period of several seconds by specific game playing. These researchers demonstrate the positive impact of playing a video game on visual attention with only 12 h of training on a very fast-paced game.

An interesting feature of the work was the selection of appropriate control groups using specific games with specific features. The games they used in their research were the following: *America's Army* (United States Army, 2004); *Harry Potter, Quidditch World Cup* (Electronic Arts, 2003); *Tetris* (Pajitnov, 1985); *Interactive Metronome* (Interactive Metronome, Inc., 1993–2004); and a baseline group that played card games. The experimental game was *Unreal Tournament* (Epic Games, 2000), a very fast-paced action game. Each of the other games were labeled controls and were selected because of their different features. For example *America's Army* is like the experimental game because it is a first-person shooter with a similar interface. However, the pace of the game was generally slower and it placed stronger emphasis on strategy and teamwork. *Interactive Metronome* provided rhythmicity training. This program requires extremely accurate timing and has been linked to improvement in attentional and cognitive skills, including vigilance, temporal sequencing, and motor planning. Second, a baseline group (card playing) was included to control for the effects of practice due to test–retest improvement on the attentional blink or multiple object tracking tasks.

These researchers argue that in order to show that the video training had an effect the retest-retest improvement had to be larger than that of the baseline group. The visual attention tasks were presented to participants before and after game playing sessions. The posttest was administered to participants irrespective of condition at least 24 h after their final game training session. This was done to eliminate the interpretation that any improvement on the dependent measures was due to increased arousal. Participants were college students. Students who had reported playing action or sport games for more than 1 h were not included in the study.

Results indicated that three groups did improve significantly between pre- and posttesting. These groups were the experimental and two controls (*Interactive Metronome* and *Tetris*). However when compared to the baseline group only the experimental group showed improvement on visual attentional skills. The *Unreal Tournament* action game was the only training that produced significant improvement in measures of temporal resolution of visual attention and the ability to track multiple objects over time. The results of this study point to one direction for future research on the impact of games – it is now feasible to select or manipulate very specific aspects of games, such as the level of challenge or pace of the game, and observe their impact on specific aspects of learning (visual attention).

3.4. The Role of Visual Maps to Stimulate Learning in a Computer Game

This chapter, by Drs Richard Wainess and Harold F. O'Neil, reports on research that examines the use of navigational maps in facilitating learning in games. The authors begin their chapter by a review of earlier research on the effects of games on learning. For example, O'Neil et al. (2005) reviewed several thousand articles on games and simulation studies with adult subjects published in peer-reviewed journals over the last 15 years. Of the articles reviewed only 19 were found to be empirical and of those 19 studies, only two compared game-based training to other methods of instruction. One of these comparison studies using game-based training found positive results on retention tests (Ricci, Salas, & Cannon-Bowers, 1996) while the other found no differences on a retention test and negative effects on a cognitive skills transfer test (Parchman, Ellis, Christinaz, & Vogel, 2000). A third study found that certain game skills lead to better performance on a far transfer task (flying a jet fighter) as compared to those that did not play the game (Gopher, Weil, & Bareket, 1994). The remaining 16 studies compared game performance of a variety of games that had embedded in their design an instructional strategy or which examined the effects of individual differences on game performance (see the Cohen et al. chapter in this book for an example). Few found that instructional methods affected game performance. The lack of positive effects of games maybe that most of the games do not have embedded in their design an effective instructional strategy.

Based on this logic these researchers examined the effects of adding an effective instructional strategy (navigational maps) to increase the effectiveness of a game to teach problem solving. Navigational maps are a type of graphical scaffold that has been shown to be effective for navigating two- and three-dimensional virtual environments.

The explanation by the authors for how navigation maps work is based on John Sweller's cognitive load theory (Sweller, 1988). Cognitive load refers to the amount of load placed on the individual's working memory while learning a new task. Since working memory limits the amount of information one can attend to, the more cognitive load the less likely that the knowledge/skills will be acquired. These researchers argue that navigation maps reduce the cognitive load by distributing some of the load that is normally placed on working memory to an external aid, navigational maps. Thus, reducing cognitive load should lead to increased learning.

The study reported in the chapter examined the use of navigation maps to assist problem solving in a video game, *SafeCracker* (Daydream Interactive, Inc., 1995/2001). *SafeCracker* is a nonviolent PC-based game where the object is for players to find and open safes through the use of clues and objects. College students were recruited as participants and randomly assigned to one of two groups, a control group playing the game without the use of navigational maps, and a treatment group where they played the game using navigational maps. Another feature of this study was that participants played the game in two sessions, followed by the researchers offering the participants to continue playing the game for an extra half hour. This latter feature was designed to examine the continuing motivation of the participants to play the game when the "experiment" was over. The results indicated that the use of the navigational maps did not affect problem solving or continuing motivation.

The message learned from this study is that although previous research demonstrated that these navigational maps facilitate learning in some two- and three-dimensional virtual environments, when applied to learning in a three-dimensional game environment there was no observed differential increase in learning.

3.5. Inquiry on the Role of Contemporary Chess Software to Enrich Human Learning and Cognition

This chapter, by Drs William Bart, Saahoon Hong, and Tae Seob Shin, described two studies that examined the relationship between chess instruction learning and higher reasoning skills. Chess is considered by many as an excellent example of problem solving. Further, there are many COTS chess games. The first study examined the role of chess playing on learning among young at-risk students. The authors' thesis was that chess playing regardless of chess skill level provided numerous opportunities in which players could practice higher order thinking skills, and further, that practicing higher order thinking skills would have an impact on academic performance. The second study with college students was designed to investigate novice chess players' ability to solve simple chess problems and transfer that ability to other chess problems.

Participants in the first study were students, ages 8–12, enrolled in an afterschool program for students identified as being at risk. Students were identified at risk by criteria using the Basic Skill Test (BST) developed by the Korean Ministry of Education. These students were assessed to have "poor" math, reading, and writing skills. Although the authors of this study acknowledge that playing chess did not teach directly these specific skills (reading and writing skills) it is their belief that the types of cognitive skills that are acquired by students learning about and playing chess would transfer to their school work. Two groups were formed by assigning students to either a control or experimental group. The experimental condition consisted of participants receiving a 90-min chess lesson once a week for over 3 months; the control attended regular school activities after school. The 12 chess lessons included reviewing, lecturing, and chess playing. Their results indicated no differences between groups. The authors did not report any data on the impact of playing chess on math, reading, or writing.

The second study involved teaching college students who had little or no experience with chess. These novices were asked to solve two chess problems. The participants were first taught to solve a chess problem that involved checkmating the opponent's King with two rooks. Once the participant completed the first game they were asked to play a second game that was a mirror image of the first. If they had been black, they were white. The hypothesis was that the participants who solved the first problem were likely to solve the second because it involved the same strategy. The results indicated that only 22 of the 98 participants were able to solve the first problem. Only 8 of the 22 were able to solve the second problem leading the authors to conclude that the novices who were trained to solve one chess problem successfully were not able to solve a second similar problem, i.e., no transfer. Such results are not uncommon in that transfer has to be directly taught.

A next step in this research would be to carefully analyze the performance of the participants to identify which specific cognitive skills are being learned while playing COTS chess games such as planning, pattern recognition, and monitoring the effectiveness of such strategies.

3.6. Would You Like to Play a Game? Experience and Expectation in Game-Based Learning Environments

The chapter by Drs Jacquelyn Ford Morie, Rebecca Tortell, and Josh Williams presents the results of a series of experiments that examine in a game-based environment the role of prior experience and immediate priming effects on players' arousal state, performance, and memory. This chapter begins with a brief history of the use of games to train military leaders. For example, the military training community is currently interested in using computer games. This chapter, like other chapters in this book, attempts to take into account what factors influence how and why people play games. The authors argue that understanding how these factors influence game players will help to identify which design factors are crucial to the success of a game. For example, rather than using a two-dimensional game, they chose a game to be played within a virtual environment (three dimensions) as their experimental platform. In these game-based environments a game player's engagement is observable and tangible. The learner's actions drive what is to be presented and therefore what is taught. One of the crucial factors in virtual environments (VR) is whether it can create in the individual a sense of "being there," a suspension of disbelief generally referred to as presence (Lee, 2004). Two types of presence were defined as being important, i.e., spatial where a person feels physically being within a virtual space, and social where interaction with others (virtual or real) in the environment provides a strong sense of connection.

The developmental goal of their research was to create a VR environment that evoked an emotional response in the participants, and felt cognitively real. To accomplish this goal they performed a series of studies. For example, priming was manipulated by providing participants with either an instructional set that told them they would be participating in a serious military mission or a cool role-playing game. In the game, the participants played the role of a military scout and she/he was to go into a rural area to observe and to determine if the evidence suggested that rebels had control, and mark their headquarters with a small GPS locator device so an air strike force could target and destroy the hideout. There were two groups of participants, university students and soldiers (Army Rangers). Participants' arousal was measured by recording their heart rate and skin conductance while they were playing the game.

Their results were mixed. They did not find that priming worked for civilians but did have an effect on the soldiers. Soldier performance was superior to civilians in recalling more of the games' goals and characters. Soldiers reported also to be less immersed and had higher arousal levels than civilians. Prior game experience had a positive effect on what the participants learned from the game. This chapter is interesting in that the variables manipulated, i.e., prior experience and priming, are inexpensive to add to VR game training programs.

4. Conclusions

Several of the authors in this volume cite the results of their review of the literature on the effects of games on learning. This section of the chapter melds both results of the empirical game literature in general and the chapters in this book specifically. In general authors conclude that there are few empirical studies, of the empirical studies comparing how effective games are the results are mixed, and that many of the studies reviewed failed to find positive effects of game playing instructional strategies on learning. The lack of empirical studies speaks both to the gaming culture (e.g., if customers buy it, the game works) and to the funding agencies (lack of vision to fund such empirical studies). In addition to the theoretical frameworks and reviews of game literature, the contribution of the authors has been to document the results of nine empirical studies to the literature. These empirical studies were reviewed and lessons learned developed.

The explanations for failures to find positive effects of game playing for adults are many and are addressed in this book. These are (1) The research on games in general lack empirical research designs; (2) There is a need for a conceptual framework to guide inquiry; (3) There is a need to take into account individual difference of potential players; (4) There is a need to develop a conceptual framework for the assessment of learning in games; (5) There is a need to trade off entertainment value vs. training value; (6) There is a lack of attention to the role of transfer in games; (7) There is a lack of a pedagogical base to guide game design, e.g., effective instructional strategies embedded in the game; and (8) There is a need to trade off costs of developing a game for research purposes vs. use of COTS. Each of these issues will be discussed with respect to the issues surrounding them in the remainder of this chapter.

1. Lack of research designs: Reviews of the literature cited in several of the chapters point to two important aspects of this issue. First, most of the studies in the game literature failed to implement an empirical design to test their hypotheses that game playing leads to learning. There is a lack of both qualitative and quantitative designs. For example, there is a lack of the gold standard, i.e., randomized control and experimental groups. Second, much of the game research has used COTS games which seldom have embedded an effective instructional strategy in their games, nor do they have embedded reliable and valid measures of learning outcomes.
2. The lack of a conceptual framework. This lack has led to research that has not been useful in the development of effective training games. There is a need to formulate a conceptual design for research as well as development. A conceptual design would serve to identify learning outcomes and the characteristics of entertainment games that influence game players' motivation and learning. For example, a game that is suitable for individuals may not be suitable for promoting teamwork and task work skills. Currently there is little to no consensus on what game features or characteristics support learning.
3. Need to take into account individual difference of players. According to several authors in this volume, previous research has neglected individual differences of game players. These differences include their prior experience with games and the prior experience with specific type of games played. These differences appear to have an effect on learning. Another individual difference is their style of play (Edwards, 2004), e.g., gamists, narrativists, and simulationists, which has no empirical studies as yet.
4. Need to develop conceptual framework for the assessment of learning in games. Currently, there is no consensus on the development of assessment requirements for games. A framework is needed to address issues as design requirements, outcome and process measures, validity and reliability, wrap-around approaches to evaluating existing games (COTS), and the use of games for assessment when the initial training intervention is not a game. The identification of the right metrics tied to learning objectives is also needed.
5. The need to trade off entertainment value vs. training value. Although serious game developers have modest success in developing games that address learning (more for kids than adults), they have been less successful in implementing features of games that make them entertaining. The entertainment value of a game according to Munro (this volume) is the many hours of diversion provided to game players. The game industry requires that the goals of the game not be achieved too early, so that the consumers' expenditure on the game provides the value of many hours of diversion. How one quantifies the variables that make up fun or entertainment value is still unknown.

6. Lack of attention to the role of transfer in games. The evidence indicates that in the literature very few studies that teach for or measure transfer, i.e., to what extent does the game develop and sustain the transfer of skills, strategies, and other social behaviors to other games and other settings, e.g., on the job. Further, of the empirical studies in this volume, only a few measured transfer (e.g., Shen & O'Neil) but none measured transfer to the job site. One very intriguing idea suggested by Baker and Delacruz (this volume) is the use of games to assess when the game itself embodies the goals of learning, or when the game serves as a proxy for typical tests, such as an on-the-job performance test.
7. Lack of a pedagogical foundation based on the science of learning to guide game design, i.e., effective instructional strategies embedded in the game. Some of the authors in this volume have added to COTS games effective instructional strategies (e.g., worked examples) that have been tried in other media (e.g., computer-based instruction), arguing that it is not the media that teaches but the instructional design embedded in the game that accounts for learning. The challenge for the research, design, and development community has been to collaborate and produce relevant and effective game technologies (e.g., guidelines) that are portable, engaging, extensible, and easy to use in the gaming area. Curtis et al. (this volume) have provided a set of guidelines that can be used to match games and their features to team training objectives. Although these guidelines are not based on empirical research they do provide the trainer/educator and researcher with guidance.
8. Need to trade off costs of developing a game for research purposes vs. use of COTS games. For example, a game specifically designed for research purposes was *Space Fortress* (Donchin, 1989). Donchin and his colleagues carefully analyzed the skills involved in flying a high-performance fighter jet and then designed *Space Fortress* to teach those skills. However, to develop a *Space Fortress*-type game to commercial standards would be very expensive and time-consuming. The advantage would be both various instructional and assessment strategies could be built in. Alternatively the selection and use of a COTS game is inexpensive but limits the researcher to wrap-around strategies as source code to implant embedded instructional or assessment strategies is not available. The lesson to be learned is that it is unlikely that one will find a COTS game that will exactly satisfy the needs of the educator/trainer or researcher completely. It also suggests that one would also look for the development of authoring tools so that scenarios of the game can be tailored with specific instructional and assessment strategies to meet specific learning objectives.

References

Baker, E. L. (1997). Model-based performance assessment. *Theory Into Practice, 36*, 247–254.

Bioware. (2002). Neverwinter Nights [Computer software]. New York: Infogrames.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds), *Team effectiveness and decision making in organizations* (pp. 330–380). San Francisco: Jossey-Bass.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*(4), 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21–30.

Clark, R. E. (Ed.). (2001). *Learning from media: arguments, analysis, & evidence*. Greenwich, CT: Information Age.

Clark, R. E. (2004). *Design document for a guided experiential learning course*. Technical Report. Los Angeles: University of Southern California. Retrieved on November 17, 2006 from http://cogtech.usc.edu/publications/clark_gel.pdf.

Day, E. A., Arthur, W., Jr, Bell, S. T., Edwards, B. D., Bennett, W., Jr, Mendoza, J. L., & Tubre, T. C. (2005). Ability-based pairing strategies in the team-based training of a complex skill: Does the intelligence of your training partner matter? *Intelligence, 33*, 39–65.

Daydream Interactive, Inc. (1995/2001). SafeCracker [Computer software]. Toronto, Ontario, Canada: DreamCatcher Interactive.

Donchin, E. (1989). The learning strategies project: Introductory remarks. *Acta Psychologica, 71*, 1–15.

Edwards, R. (2004). *System does matter*. Retrieved July 28, 2005 from http://www.indie-rpgs.com/_articles/system_does_matter.html.

Electronic Arts. (2003). Harry Potter, Quidditch World Cup [Computer software]. Redwood City, CA: Author.

Epic Games. (2000). Unreal Tournament [Computer software]. Cary, NC: Author.

Freeman, J., Diedrich, F. J., Haimson, C., Diller, D. E., & Roberts, B. (2003). Behavioral representations for training tactical communication skills. In *Proceedings of the 12th Conference on Behavior Representation in Modeling and Simulation*. Scottsdale, AZ.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36*, 387–405.

Interactive Metronome, Inc. (1993–2004). Interactive Metronome [Computer software]. Weston, FL: Author.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development*, (2nd ed., pp. 301–319). New York: McGraw-Hill.

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.

Lee, K. M. (2004). Presence, explicated. *Communication Theory, 14*, 27–50.

Mayer, R. E. & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E. & Wittrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds), *Handbook of educational psychology* (pp. 47–62). New York: Simon & Schuster Macmillan.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing simulation-based performance assessment. *Computers in Human Behavior, 15*, 335–374.

Moreno, R. & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165–173.

O'Neil, H. F., Jr, Baker, E. L., & Fisher, J. Y.-C. (2002). *A formative evaluation of ICT games*. Los Angeles: University of Southern California.

O'Neil, H. F., Wainess, R, & Baker, E. L. (2005, December). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455–474.

Pajitnov, A. (1985). Tetris [Computer software]. Moscow, Russia: Academy of Sciences in Moscow.

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*, 73–87.

Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996). Do computer-based games facilitate knowledge acquisition and retention? *Military Psychology, 8*(4), 295–307.

Rival Interactive. (2002). Real War: Rogue States [Computer software]. Springfield, VA: Simon & Schuster Interactive.

Serfaty, D., Entin, E. E., & Johnston, J. H. (1998). Team coordination training. In J.A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 221–245). Washington, DC: American Psychological Association.

Sims, D. E., Salas, E., & Burke, C. S. (2004). Is there a "Big Five" in teamwork? Presented at the *19th Annual Conference of the Society for Industrial and Organizational Psychology*. Chicago, IL.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In J. A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 271–295). Washington, DC: American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., & McPherson, J. A. (1998). Team dimensional training: A strategy for guided team self-correction. In J. A. Cannon-Bowers & E. Salas (Eds), *Making decisions under stress: Implications for individual and team training* (pp. 271–295). Washington, DC: American Psychological Association.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81*, 457–466.

Sweller, J. (1990). Cognitive Processes and Instruction Procedures. Australian *Journal of Education, 34*(2), 125–130.

Tennyson, R. D. & Breuer, K. (1997). Psychological foundations for instructional design theory. In R. D. Tennyson, F. Schott, N. Seel, & S. Dijkstra (Eds), *Instructional Design: International Perspective, Volume 1: Theory, Research, and Models*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tennyson, R. D. & Breuer, K. (2002). Improving problem solving and creativity through use of complex-dynamic simulations. *Computers in Human Behavior, 18*, 650–668.

United States Army. (2004). America's Army: Special Forces (Vanguard) (v2.2.0) [Computer software]. Retrieved from http://www.americasarmy.com

Ward, M. & Sweller, J. (1990). Structuring effective worked examples. *Cognition & Instruction, 7*(1), 1–39.

Wolfe, J. (1997, December). The effectiveness of business games in strategic management course work [Electronic Version]. *Simulation & Gaming Special Issue: Teaching Strategic Management, 28*(4), 360–376.

AUTHOR INDEX

This page intentionally left blank

SUBJECT INDEX