University of Central Florida
STARS

Electronic Theses and Dissertations, 2004-2019

2013

Explicit Feedback Within Game-based Training: Examining The Influence Of Source Modality Effects On Interaction

Benjamin Goldberg
University of Central Florida

Part of the Psychology Commons
Find similar works at: https://stars.library.ucf.edu/etd
University of Central Florida Libraries http://library.ucf.edu

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation
Goldberg, Benjamin, "Explicit Feedback Within Game-based Training: Examining The Influence Of Source Modality Effects On Interaction" (2013). *Electronic Theses and Dissertations, 2004-2019*. 2631.
https://stars.library.ucf.edu/etd/2631

Showcase of Text, Archives, Research & Scholarship

EXPLICIT FEEDBACK WITHIN GAME-BASED TRAINING: EXAMINING THE INFLUENCE OF SOURCE MODALITY EFFECTS ON INTERACTION

by

BENJAMIN S. GOLDBERG
B.A. University of Florida, 2005
M.S. University of Central Florida, 2010

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Modeling and Simulation
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2013

Major Professors: Jan Cannon-Bowers & Clint Bowers

© 2013 Benjamin S. Goldberg

ABSTRACT

This research aims to enhance Simulation-Based Training (SBT) applications to support training events in the absence of live instruction. The overarching purpose is to explore available tools for integrating intelligent tutoring communications in game-based learning platforms and to examine theory-based techniques for delivering explicit feedback in such environments. The primary tool influencing the design of this research was the Generalized Intelligent Framework for Tutoring (GIFT), a modular domain-independent architecture that provides the tools and methods to author, deliver, and evaluate intelligent tutoring technologies within any training platform. Influenced by research surrounding Social Cognitive Theory and Cognitive Load Theory, the resulting experiment tested varying approaches for utilizing an Embodied Pedagogical Agent (EPA) to function as a tutor during interaction in a game-based environment. Conditions were authored to assess the tradeoffs between embedding an EPA directly in a game, embedding an EPA in GIFT's browser-based Tutor-User Interface (TUI), or using audio prompts alone with no social grounding.

The resulting data supports the application of using an EPA embedded in GIFT's TUI to provide explicit feedback during a game-based learning event. Analyses revealed conditions with an EPA situated in the TUI to be as effective as embedding the agent directly in the game environment. This inference is based on evidence showing reliable differences across conditions on the metrics of performance and self-reported mental demand and feedback usefulness items. This research provides source modality tradeoffs linked to tactics for relaying training relevant explicit information to a user based on real-time performance in a game.

ACKNOWLEDGMENTS

I underestimated the difficulty in acknowledging all of those worthy of recognition, as there are too many to list. It’s been a wild, fun and extremely stressful ride over the past few years and there are a number of individuals to thank. I apologize if I miss you.

First and foremost, I would like to thank my advisor Dr. Jan Cannon-Bowers, my co-chair Dr. Clint Bowers, and the rest of my committee: Dr. Robert Sottilare, Dr. Rudy McDaniel, and Dr. Peter Kincaid. I couldn’t have asked for a better group of professionals to share this experience with. In particular, I would like to thank Jan and Clint for allowing me to tackle this mountain in my own way, and I appreciate the guidance and support throughout. Your expertise on game-based learning and experimental design was invaluable. Jan, you were instrumental right from the start. Your smarts and experience helped me through the difficult phases, and your coaching was just what I needed to get through the initial gauntlet of developing a topic. I’m thankful for the friendship we developed during this process.

I would also like to thank a number of individuals from the U.S. Army Research Laboratory (ARL) and the Simulation and Training Technology Center (STTC). For starters, I would like to thank my fellow LITE Labbers for being an awesome group of individuals to work with on a daily basis. Dr. Heather Holden, Janice Connor, and Dr. Anne Sinatra, thanks for all the help and for being a listening ear when I needed one. I owe a special thanks to Dr. Keith Brawner. You were as critical as ever when we traveled to West Point for data collection. I would also like to thank John Hart, Chris Metevier and Chuck Amburn for helping me stay motivated throughout the entire process. Last but not least, I have a large debt of gratitude to Dr. Robert Sottilare. Bob served an integral role in my development over the past three years, and I would not have been able to accomplish this feat without his support and guidance. Thanks for being a dedicated leader and for mentoring me on both a professional and personal level.

Next, I would like to thank my family and friends for all the love and support during my stint as a graduate student; especially for putting up with me during the rough times. My father Stephen and my mother Elaine were integral in keeping me sane through this process. Thanks for keeping me grounded and for always helping me keep my eye on the prize. A special thanks goes to my brother Jordan and my sister in-law Shelley. You gave me a place to escape the insanity and you were the perfect target allowing me to rant and vent about all the pains and pleasures of working through a dissertation.

Lastly, I owe a big thank you to a few notable individuals from the United States Military Academy at West Point. Dr. Mike Matthews, you're the man! I can't express how great it's been to work with you across multiple projects over the past few years. I consider myself lucky to have gotten to know you so well. You're support and determination was essential to this study becoming a reality. I am indebted to you for your true professionalism and your gracious hospitality. I would also like to thank Colonel Jim Ness, Bess Georgoulas, and Dr. Elizabeth Velilla. You were essential pieces to my data collection, and your support and motivation for the research being conducted made the whole experience that much more enjoyable.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AAR	After Action Review
AI	Artificial Intelligence
ALC 2015	Army Learning Concept 2015
ANOVA	Analysis Of Variance
ANCOVA	Analysis of Co-Variance
API	Agent Persona Instrument
CBT	Computer-Based Training
CLT	Cognitive Load Theory
CLS	Combat Lifesaver
CUF	Care Under Fire
CV	Co-Variate
DRE	Delay Retention Effect
DV	Dependent Variable
ECD	Evidence Centered Design
EPA	Embodied Pedagogical Agent
FiST	Fire Support Team
GIFT	Generalized Intelligent Framework for Tutoring
ITQ	Immersive Tendencies Questionnaire
ITS	Intelligent Tutoring System
IV	Independent Variable
KSAs	Knowledge, Skills, and Abilities
LTM	Long Term Memory
MANOVA	Multivariate Analysis of Variance
MD	Mental Demand
MIMIC	Multiple Intelligent Mentors Instructing Collaboratively
MeTERS	MEdical Training Evaluation Review System
MRT	Multiple Resource Theory
NASA-TLX	National Aeronautics and Space Administration – Task Load Index
NPCs	Non Player Characters
SBT	Simulation-Based Training
SCT	Social Cognitive Theory
SIMILE	Student Information Models in Intelligent Learning Environments
STEM	Science Technology Engineering Math
SWAT	Subjective Workload Assessment Technique
TC3	Tactical Combat Casualty Care
TC3Sim	Tactical Combat Casualty Care Simulation
TFC	Tactical Field Care

TPI	Temple Presence Inventory
TRADOC	Training and Doctrine Command
TUI	Tutor-User Interface
USMA	United States Military Academy
VGE	Videogame Experience
VoG	Voice of God
WL	Workload
WM	Working Memory
WP	Workload Profile
ZPD	Zone of Proximal Development

CHAPTER 1: INTRODUCTION

Today more than ever training and education communities are incorporating technology-driven learning platforms as tools to expand instruction beyond the boundaries of traditional schoolhouse environments. This maturation of technology-based training is important as academic and military communities are pushing for an accelerated, self-directed culture of learning. This is achieved through (among other things) the promotion of active, hands-on learning experiences, and also by making learning materials and exercises available at a place and time convenient to the user. Considerable research is being focused on identifying tools and methods that enable computers to compliment the learning process in the absence of live instructors. The intention is for computers to support the development of realistic and immersive learning experiences designed to promote knowledge and skill acquisition.

To accomplish the development of effective self-directed educational platforms, research needs to examine standardized approaches for monitoring student activity and identifying innovative and creative ways to integrate feedback and pedagogy into technology-based platforms no matter the domain being instructed. The goal of the current effort is to investigate approaches for enhancing game-based training applications through the incorporation of performance-based feedback functions. Specifically, this work examines methods for embedding feedback delivery mechanisms within game environments and assesses the influence variations in the source and delivery of feedback have on learning outcomes and self-reported measures of cognitive load, immersion and flow. The results can be used to inform requirements for future simulation-based training (SBT) and intelligent tutoring system (ITS) design that is aimed at integrating real-time explicit task feedback within an open game environment.

The sections that follow highlight the current state of training technologies and the gaps in knowledge that are the basis for this research. A summary of what is known about SBT is presented first. Next, the role and importance of feedback in SBT is presented, with an emphasis on integrating ITS functions into simulation-based applications. The introduction concludes with a statement of the research problem being addressed along with a purpose of the study description including a listing of research questions motivating this effort.

Simulation-Based Training

A major research thrust in technology-based training is to enhance systems to provide hands-on learning experiences with embedded pedagogical support functions. SBT is one such approach that provides attractive options for education, training and rehearsal. For the context of this research, SBT is defined as "a type of training that depends on the simulation to provide essential cues to trigger appropriate behaviors" (Cannon-Bowers & Bowers, 2008a, p. 317). Under this depiction, simulations are characterized by modeled representations of reality that can abstract, simplify, or accelerate process models associated with real-world phenomena (Galvao, Martins, & Gomes, 2000).

The benefit associated with SBT platforms is they provide realistic environments that allow individuals to master complex material and learn and apply new information through execution of simulated tasks (Menaker, Coleman, Collins, & Murawski, 2006). The learning process is influenced by student-centered teaching methods prompted by theories of 'discovery' (Bruner, 1966; Hermann, 1969) and 'active' (Johnson, Johnson, & Smith, 1991) learning. They incorporate interacting elements of logic, memory, visualization, and problem solving that cater to elements required for learning; engagement, interaction, and satisfaction (Amory, Naicker, Vincent, & Adams, 1999). This is achieved by replacing traditional instructional techniques with methods of role-playing, simulations, self-regulated exercises, and other types of problems requiring creative and critical thinking skills (Greitzer, Kuchar, & Huston, 2007). Research has demonstrated these strategies are an effective alternative to traditional classroom instruction because they assist learners in creating and adjusting mental models for newly acquired information (Cuevas, Fiore, Bowers, & Salas, 2004). These environments also provide a forum for learners to actively participate with learning material and to view the effect varying actions have on outcomes.

These approaches provide a new means for educators and trainers to deliver domain content, as well as new mechanisms for the practice and assessment of relevant instructional objectives. This enables engaging activities that assist individuals with learning and applying the knowledge, skills, and abilities (KSAs) associated with a given domain. These strategies also relieve the associated costs and limitations of live instruction, and reduce the risk of damaging costly equipment or endangering lives (Bratt, 2009). As a result, many professional fields and domains including: military, law enforcement, medical, and emergency-management organizations apply SBT because of their coupled benefits. (Salas & Cannon-Bowers, 2001).

This application across professional fields occurs because SBT allows for the development of authentic scenarios that facilitate learning and cognitive development. In the military context, interactions in a simulated environment enable visualization and practice of task execution. As a result of practice in SBT, learners come to their first live performance experience with an advantage (Waldman, 2009). In addition, SBT enables Soldiers to interact with multiple scenarios in a short timeframe. This allows for rapid exposure to variations in task conditions that build task relevant experience, which would require drastically more time, manpower, and resources to achieve from live training exercises (Pine, 2009).

A primary goal of many modern military training systems is to provide the learner with strategies that aid in the development of higher-order thinking skills and enable them to adapt decision-making tactics under variable missions and conditions (Wisher, Macpherson, Abramson, Thorton, & Dees, 2001). In today's combat environment, tasks are executed under a multitude of complex, stressful, and ambiguous settings where decisions must be quick and actions must be executed in a timely manner (Salas, Priest, Wilson, & Burke, 2006). Therefore, training aims to foster successful task execution and the values associated with making reasonable decisions under difficult circumstances (Bratt, 2009). SBT fosters this type of learning by applying principles of instructional design through the processes of development, application, and evaluation of task relevant KSAs in realistic situations (Oser, Cannon-Bowers, Salas, & Dwyer, 1999; Salas, Rosen, Held, & Weissmuller, 2009b).

Simulating a task in a virtual environment and providing the ability for an individual to practice does not on its own increase expertise (Ericsson & Ward, 2007). SBT simply replicates a real-world representation of a problem space where KSAs can be applied within bounded realistic conditions that aid in skills training (e.g. time pressure, stress). Expertise development in SBT platforms is not practical without apt pedagogical support (Ericsson & Ward, 2007). This is a recognized gap because too often simulations are fielded without pedagogical components and functions. Simulations intended for education and training provide a means for practicing KSAs, but as mentioned above often lack elements of pedagogy that guide the learning process

(Nicholson, Fidopiastis, Davis, Schmorrow, & Stanney, 2007). Currently, simulations in the military are utilized as supplemental tools to instruction and require instructors to monitor interaction for identifying deficiencies and correcting erroneous actions. This limits their applicability as effective training tools outside the schoolhouse when experienced trainers may not be present due to the frequent absence of feedback mechanisms that foster the understanding performance outcomes.

In recognition of these limitations, the Army Learning Model 2015 was developed to highlight a new learning model to drive development of future training systems (TRADOC, 2011). The document outlines the Army's strategy and motivation to steer away from traditional instructor-led courses that are executed in a lock-step approach. One Army Training 2015 requirement is for dramatic reductions in instructor-led training through the incorporation of a blended learning environment of simulations, gaming environments, and other technology-driven platforms (TRADOC, 2011). A secondary goal is to synchronize and tailor training to meet the needs of the individual learner (Durlach & Ray, 2011).

These goals are not attainable given the current state of knowledge regarding SBT. Specifically, further research is required to develop SBT systems that are easily accessible, have mechanisms for monitoring training performance for the purpose of tailoring training on the individual level, and to provide feedback automatically so that direct intervention by a human instructor is not required. The sections that follow provide insight into the nature of required research to achieve Army Learning Model 2015 goals.

Feedback in Simulation-Based Training

It is understood that feedback is an essential element to learning. It is used in a number of ways and for a number of reasons. Feedback serves a multitude of functions in the instructional process and is viewed as a fundamental element in all theories of learning and instruction (Koedinger, Anderson, Hadley, & Mark, 1997; VanLehn et al., 2005). In fact, most scholars commonly agree that *learning cannot occur without some source of feedback* (Bransford, Brown, & Cocking, 2000; Magerko, Stensrud, & Holt, 2006). Whether to inform a learner of an incorrect step or misconception, to increase motivation by acknowledging successful performance, or by promoting reflection through prompts and questioning, feedback is critical in learning from errors and improving KSAs no matter the domain. It allows an individual to compare inconsistencies of their own performance with the desired goals of a given task (Kluger & DeNisi, 1996). This is important because it can increase motivation by identifying discrepancies in performance, reduce uncertainty for how an individual is performing, and assist someone in correcting errors found in execution (Davis, Carson, Ammerter, & Treadway, 2005).

In SBT, feedback often results from environmental changes in a scenario based on actions taken by a player. Narciss (2008) describes this type as *implicit feedback*, in that it occurs naturally within the virtual environment and is tied directly to the context of the task decision and outcome. This allows for the forming of mental connections between actions taken and resulting outcomes and environmental changes (Billings, 2010). In comparison, guidance given outside of a specific task context where information is relayed to link performance with overarching training objectives is termed *explicit feedback* (Kluger & DeNisi, 1996). This feedback is commonly delivered from an external source to the simulation. In most cases it is provided by instructors monitoring how scenarios are performed. Typically, this form of feedback provides information confirming or correcting actions taken, and is used to highlight errors and strengthen response to correct decisions (Kulhavy & Stock, 1989; Mory, 2004). Currently, linking scenario outcomes to training objectives through explicit feedback is most often left to instructors. Current research is focused on identifying tools and methods for embedding autonomous explicit feedback functions that have an implicit feel within the game environment.

In current forms of SBT, feedback can be autonomously generated. It plays an implicit role and typically takes two forms: (1) short-term feedback that is immediate and signifies progress, natural consequences of interactions and task completion, and (2) holistic feedback that comes with player development and progression in the story narrative (Murphy, 2011). Holistic feedback is most noticeable in story-based scenarios, in that scenarios link task events to a common storyline and narrative for providing a long-term feedback metric on performance and progression. The missing piece is how performance within SBT environments mesh with desired training objectives, and how autonomously generated explicit feedback can improve performance and reduce/remove the burden on instructors monitoring task execution. Identifying techniques for embedding explicit (computer-generated) feedback in SBT is the focus of this research. This requires systems to accurately monitor and link performance with specific training objectives, as well as having triggers to carry out interventions when guidance is deemed necessary.

Intelligent Tutoring Systems

One limiting factor associated with computer games and simulations in the educational and training domain is their lack of credible feedback mechanisms in the absence of human intervention. Before their use in a training context is made prevalent, there are a number of faculties these systems must be able to perform prior to reaching their full potential. To enable these systems to produce effective outcomes on their own, SBT requires capabilities for tracking performance and presenting feedback in real-time. Because of the desire to pursue a more self-regulated learning paradigm where instruction outside of the classroom is conventional, mechanisms in SBT need to be implemented that facilitate the corrective strategies and actions executed by instructors.

Solutions developed to meet this need are termed Intelligent Tutoring Systems (ITSs). The heart of this line of research is for the development of tools that enable computer-based training systems to emulate instructional strategies used by human tutors during one-on-one instruction (Person & Graesser, 2003). This involves knowing about the domain being instructed, knowing about the individual being instructed, and knowing how to instruct based on the domain and individual. ITSs accomplish this by providing personalized training experiences through the monitoring of user interactions with a system and using AI methods to assess progress and trigger adaptive interventions (Goldberg, Holden, Brawner, & Sottilare, 2011). The role of the ITS is to mediate training sessions by providing feedback when appropriate and adjusting difficulty levels to maintain desired challenge.

With the identified divergence between SBT and ITSs, research is needed to merge the benefits of adaptive instruction provided by ITSs and the applied experience provided by SBT. A large gap in this arena is empirical research examining how to optimally deliver feedback and adaptation based on individual differences derived from model outputs. Though there is a recent growth of empirical research supporting adaptive SBT applications (Mangos & Johnston, 2009), the literature does not make clear the distinctions of pedagogical strategies found to be most effective (Billings, 2010). A lot of work in this field over the past decade has focused on the modeling and data mining component of task interaction in SBT to determine when and why errors are present, and to predict cognitive and affective state trends during learning events (Woolf, 2009). The gap addressed by the current work is understanding the role feedback plays in these types of learning events and how to relay information without removing the individual from the simulated experience. This includes investigating what to present/adapt, when to present/adapt, and how to present/adapt when actions are deemed to warrant feedback. The driving force of this research is to examine options for answering the 'how' question for presenting feedback within simulated game-based training environments.

Statement of the Problem

SBT designed within synthetic virtual worlds provide the environments for 'practicing' the application of acquired skills, but often lack instructional guidance essential for effective training to occur (Nicholson et al., 2007). It is the goal of this research to explore the synchronization of technology with the learning sciences to foster tailored and guided game-based training. Specifically, this work aims to address a crucial research gap related to how game and instructional designers can leverage the functionality of SBT for the purpose of delivering tailored learning experiences in the absence of an instructor. It is consistent with an evolving thrust in the ITS research domain; namely, how to embed pedagogy and feedback within game-based instructional environments. Through the integration of AI and ITS technologies, SBT applications have the ability to support pedagogical interventions intended to maintain training progression.

A fundamental problem in this area is the lack of empirical evidence supporting the usefulness of instructional components and explicit feedback mechanisms in SBT events. This, in part, is due to recent advancements in gaming technologies that afford this capability, and a lack of understanding on how to deliver feedback within an interactive virtual world environment. A common trend is incorporating ITS functions in SBT just because they are now possible rather than because there is evidence of their training effectiveness (Sweller, 2008). Empirical evidence is required to identify optimal approaches for delivering training relevant feedback in a SBT environment.

Given the central role feedback plays in the learning process, research is needed to address the impact variations in feedback delivery have on learning outcomes and system acceptance within game-based training environments. This involves examining elements available in the game world and existing ITS tools and methods that can be leveraged specifically for the delivery of explicit feedback, including Non-Player Characters (NPCs). The principal goal in this study is to evaluate the feasibility of embedding NPCs in SBT as mechanisms for guiding instruction and delivering feedback content. The specific focus is to introduce instructor qualities into game play that aids in the prevention of erroneous task execution, while maintaining optimal affective and cognitive learning states.

This is important due to work highlighting embodied agents as effective tools in training applications (Cassell, Sullivan, Prevost, & Churchhill, 2000; Yee, Bailenson, & Rickertsen, 2007). These entities within a virtual world are digitally modeled actors with interactions that are determined through predefined algorithms (Bailenson, Yee, Merget, & Schroeder, 2006). With this in mind, an objective of the current study is to investigate the effect varying implementations of Embodied Pedagogical Agents (EPA) have on performance within, and motivation for interacting with, a training system. How an EPA is situated in the learning environment will be examined across multiple conditions, with approaches including characters embedded within the game environment and characters present in a tutor interface external to the virtual world.

In the context of feedback, this requires the evaluation of components that will inform the source modality to present feedback in when conditions exist that call for an intervention. Hence, two subordinate questions will be specifically addressed: *1) what effect does the source modality of explicit feedback have on performance; and 2) what effect does the source modality of explicit feedback have on subsequent interaction and acceptance?* Components include variables that are derived from both learning theory and game design principles. In the context of this work, *source modality* refers to the tools and methods applied for delivering explicit feedback within SBT. A secondary research objective this work addresses is: *does feedback delivered externally to the environment as in-game dialogues affect performance/learning outcomes, subjective ratings of workload, and sense of flow and immersion within a virtual world?* The goal is to determine if there are approaches to embed explicit feedback functions in a game so that it has an implicit feel to its delivery. The notion is to use a character defined implicitly in the environment to deliver information delivered from an explicit external source. If developed properly, the explicit feedback would be viewed as implicit to the environment based on the source of its delivery.

In addition, attributes and characteristics associated with EPA design will be explored to determine how an agent is characterized in the learning event influences interaction from the user. It is with this thought that variations among an NPC's knowledge base, assigned role, and experience level can facilitate multiple functions of instructional support within a game-based training system. For example, comparing two conditions where participants are assigned to EPAs given distinctly different backgrounds; in one case, an EPA with a decorated career as a Soldier and trainer versus an EPA who is a peer and team-member. With research backing from Social Cognitive Theory, an additional objective of this research is: *does an agent's defined background influence their perceived competency and usefulness across learners when there are no differences in interaction?*

Purpose of the Study

This research aims to enhance SBT applications to support training events in the absence of live instruction. The overarching purpose is to determine how EPAs can be utilized as guidance functions in a virtual world environment. Specifically, this research will assess whether explicit feedback delivered by EPAs present in a scenario has a significant effect on performance or subjective ratings when compared to external feedback source modalities. Consistent with this, the secondary purpose of this study is determining how defined attributes and characteristics of an NPC impacts user interactions, and whether this profile has a direct effect on performance, motivation to use the system, and system acceptance.

Specifically, this study will examine theory-based techniques for delivering explicit training relevant feedback and their effect on performance and sense of presence within a game-based training application developed for military use. Data will be collected across multiple conditions where source of feedback is manipulated, while content presented is held constant. This is important because the intent of this work does not focus on testing the effect variations in feedback content have on learning within game-based training. In particular, this study will investigate delivery methods involving both visual and auditory feedback approaches during a game-based training event. The results will inform the ITS, SBT and serious game communities whether there is a benefit to embedding feedback delivery through embodied agents interacting within SBT scenario environment. The goal is to provide empirical support that source of feedback has an effect on training effectiveness and perceived value of application.

To accomplish this, questions will be examined that determine the value associated with a source modality type through comparative evaluations across methodologies. This is to determine if in-game EPA delivery has a noticeable impact on reported presence within the scenario storyline. Games are designed around principles intended to induce a state of flow through immersive and engaging interactions (Murphy, 2011). The notion is that delivering feedback through in-game sources will assist in maintaining immersion and will improve the effectiveness of the system. Individuals' workload will also be assessed to determine if variations in feedback source produce variations in reported scores. This will be evaluated through the implementation of multiple feedback source conditions. Theories source modality of feedback are designed around include Social Cognitive Theory (SCT), Working Memory (WM) and Cognitive Load Theory (CLT). The intent is to identify pedagogical tactics that relay training relevant information efficiently based on real-time performance and are found to be most cognitively effective. The tactics are intended to promote presence within the game world and reduce cognitive load in perceiving and interpreting explicit feedback.

Research Questions

This work investigates the effect variations in the source of real-time feedback within a scenario-based training event has on subsequent task performance; the effect the source of feedback has on post-training learning outcomes; and whether variations in feedback source produces reliable differences in trainee self-reported measures of cognitive load and flow. The study will go deeper by exploring the impact of delivering feedback through NPCs defined as EPAs, and to assess the effect varying agent delivery modalities have on trainee performance and game acceptance. Specifically, this research will examine whether there is a significant benefit to embedding EPAs directly into the task environment versus an EPA interacting with the user from an interface external to the game world. It is expected that feedback delivered by embedded EPAs will produce a higher sense of trainee presence and lower extraneous cognitive load when interpreting feedback, resulting in larger learning gains and greater motivation to interact with game-based training systems. The goal is to identify heuristics associated with how to deliver feedback in a game-based trainer, and how attributes of a character delivering feedback can be modified to compensate for an individual’s strengths/weaknesses in a given domain.

CHAPTER 2: BACKGROUND

Chapter Two Summary

Videogames are one product that supports the application of new and innovative SBT delivery approaches and reinforces concepts identified in the Army Learning Model (ALM) 2015 (TRADOC, 2011). The ALM2015's learner-centric model identifies the role of computer- and simulation-based training systems as essential components to the future of military training. The report also highlights the need for the integration of adaptive functions in such systems that can provide task-relevant feedback and adjust training in real-time based on the desire to supply effective training solutions that can be accessed from anywhere and at anytime. This requires the ability of the Army to develop digitized learning products with embedded AI in order to adapt and tailor training to the experience and knowledge levels of the individual Soldier (Durlach & Ray, 2011). From an instructional design perspective, this calls for training systems to have tools and methods for performing functions of the instructor that are natural in occurrence and do not hamper performance and retention outcomes.

The question this work aims to address is how to best integrate feedback within game-based training events, and to determine how information delivered by EPAs in serious game environments (i.e., how the content is delivered) affects a user's performance and motivation/intention for future usage? Specifically, this research seeks to identify if embedding pedagogical agents for delivering feedback directly in a game-based environment improves training outcomes, reduces cognitive load required for interpreting information, and maintains a user's sense of flow and presence in the virtual environment. With ITSs offering external interfaces for delivering feedback during game-based training, this work aims to determine if the time, money, and effort to integrate EPAs in the game-world has a distinct benefit over more simplistic avenues of relaying information to the user.

Chapter two reviews existing literature on concepts applied to SBT, specifically looking at games for training (a.k.a. serious games) and the functions feedback play in these environments. First, there will be an introduction to serious games along with a representative sample of current applications in use. Principles will be presented focusing on similarities between instructional design and game design, with an emphasis on flow and the role feedback plays in this construct. In the subsequent section, an introduction to ITS literature will be presented defining the specific features required for providing real-time feedback in game environments. An emphasis on the role pedagogy plays within serious game events will be described, and how ITS technologies can be integrated to support those functions. This will include current research on the integration of intelligent tutoring technologies within serious games and SBT applications, and the pursuit of domain-independent tools for authoring ITS components that integrate with game-based applications across multiple platforms.

Serious Games

Games intended to facilitate learning are termed ‘Serious Games’, as they are carefully designed with pedagogy around the overarching objectives of its intended use, and are hypothesized to attend to both the affective and cognitive dimensions of learning (O'Neil, Mainess, & Baker, 2005). For the context of this study, serious games are referred to as SBT applications that operate on standard desktop computing systems and incorporate components commonly seen in entertainment industry games. The term was first coined by Clark Apt in his

1970 book ‘Serious Games’ (Apt, 1970). Apt describes this genre as involving an explicit and carefully thought-out design process with an educational intent where game-play has a primary purpose other than providing entertainment (see Table 1).

Table 1. Differences between Entertainment Games and Serious Games (Susi, Johannesson, & Backlund, 2007)

	Serious Games	**Entertainment Games**
Task vs. Rich Experience	Problem Solving Focus	Rich Experiences Preferred
Focus	Important Elements of Learning	To Have Fun
Simulations	Assumptions Necessary for Workable Simulations	Simplified Simulation Processes
Communication	Should Reflect Natural (i.e., non-perfect) communication	Communication is Often Perfect

Though there are many opinions of what serious games are explicitly designed to do, there lacks a common designation among practitioners (Susi et al., 2007). In essence, “there is no one single definition of the term ‘serious games’, although it is widely accepted that they are games ‘with a purpose’. In other words, they move beyond entertainment per se to deliver engaging interactive media to support learning in the broadest sense” (Stone, 2008, p. 9). While describing the intent and purpose of serious games is rather straightforward, designing and developing such applications is challenging and involves numerous disciplines.

The effectiveness of a serious game is first dependent on the ability of the simulation to replicate specific features associated with executing a task in the real-world operational environment (Salas, Rosen, Held, & Weissmuller, 2009a). This requires constructing a synthetic representation of task environment through psychological fidelity of processes, constructs, and performance (Kozlowski & Bell, 2007; Susi et al., 2007). A common goal of SBT interaction is to promote and develop the application of higher order thinking skills and improve human performance essential for safety, effectiveness, and survival practices among domains within the military, medicine, business, and aviation communities (Salas et al., 2009a).

One early example highlighting the effectiveness of serious games involved business school students working with a series of simulations focused on finance practices (Estes, 1979). Following interaction, students commonly reported truly understanding the modeling and analysis concepts they had only previously studied theoretically, and attributed this deep understanding to the simulation (Estes, 1979). This is achieved through individuals building and verifying mental models of new information as it pertains to the simulated environment (Cuevas et al., 2004). For this purpose, the target objectives associated with a scenario must be accurately modeled so that skills attained in the game environment effectively transfer to the task environment. In this context, a serious game designer must understand both the science of learning, available simulation approaches, and how and why games work (see Figure 1) (Murphy, 2011). Ignoring these components can result in a game that is the worst of both worlds, a dull game informed by ineffective teaching methods (Bowers, 2007).

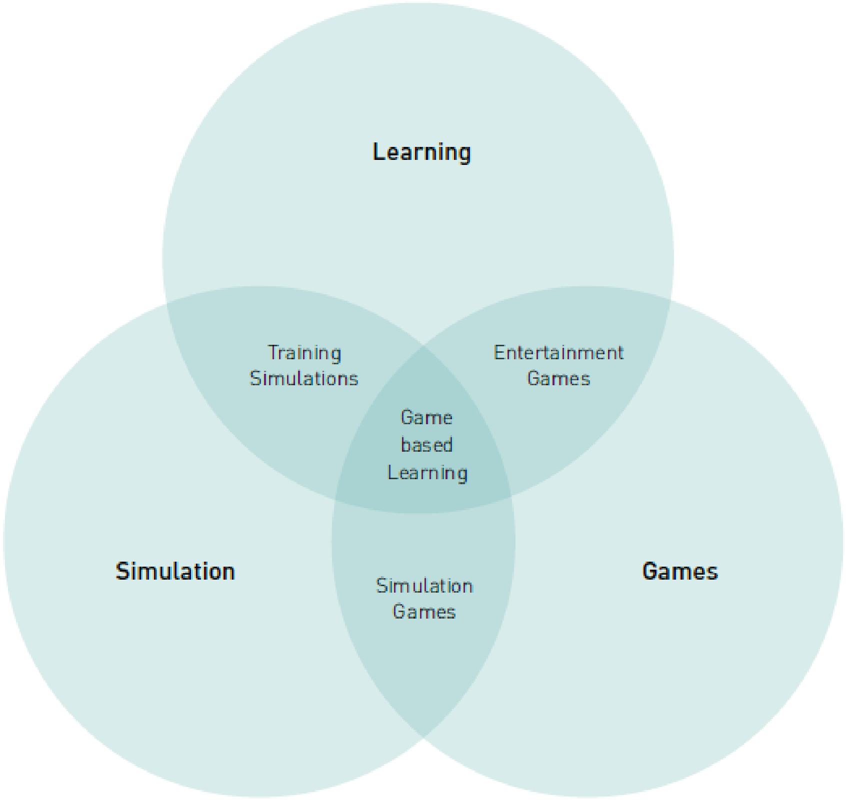


Figure 1. Interplay of Disciplines for Serious Game Design (from Martens, Diener, & Malo, 2008)

Current State of Serious Games

Current use of serious games in education and training communities range from highly interactive, open-world discovery environments to web-based static 2-D environments running on discrete user inputs. The intention, no matter the level of interactivity, is to develop applications that enable a ‘learning-by-doing’ philosophy, where users can observe outcomes and effects of decisions/actions within a safe controlled environment (Bell, Kanar, & Kozlowski, 2008; Cannon-Bowers & Bowers, 2008b). In the domains of military and medical training, the incorporation of game-based systems is for the purpose of job oriented training through scenario-based events that mimic actual operational environment and task conditions (Hartog, 2009; van der Hulst, Muller, Besselink, Coetsier, & Roos, 2008). They provide feasible and affordable solutions to training skills and tactics that are performed under circumstances that are difficult to replicate in live exercises (Bratt, 2009; Susi et al., 2007). In a study conducted by Leemkuil, Jong, and Ootes (2000), 66 articles were reviewed examining the use of games in an instructional environment. The most conclusive findings were: (1) games are most effective when handling specific subject matters with targeted objectives, (2) games produce greater retention over time when compared to conventional classroom techniques, and (3) learners reported higher interest and motivation levels.

In recent years, the use of serious games as tools for learning has seen wide application in the training and vocational fields when compared to standard education (JISC, 2007); though there has been a substantial rise in academically geared games over the past few years as a result of the STEM (Science Technology Engineering Math) Serious Games Challenge, which was started in 2011. Michael and Chen (2005) categorize the common markets serious games are utilized within to six main domains: military games, government games, educational games, corporate games, healthcare games, and political, religious and art games. In the military domain, several serious games have been used over the past decade to train Soldiers on a variety of KSAs. Examples include *Full Spectrum Warrior* to train urban warfare tactics to squad leaders (Reuters, 2003), *America's Army* to train future officers at West Point (Roth, 2003), *ELECT Bi-Lat* to train Soldiers bilateral negotiation tactics and how to practice cultural customs (Kim et al., 2009), and *UrbanSim* for practicing mission command in counterinsurgency and stability operations (McAlinden, Gordon, Lane, & Pynadath, 2009). In addition, various commercial-off-the-shelf game engines (e.g., Unity, UnReal Engine, Virtual Battle Space 2, Ogre, etc.) have been utilized by various armed forces to train military relevant tasks that are difficult and expensive to replicate in the real world (Fong, 2004; Topolski et al., 2010; Zyda, 2005).

While the use of serious games is on the rise, an important question becomes: what evidence is out there to signify these applications actually work and promote efficient learning? In a review based on instructional gaming literature, Hays (2005) finds empirical research on the effectiveness of serious games to be fragmented across different domains, age groups, and levels of interactivity, resulting in experimental confounds that make it difficult to draw valid inferences on learning efficiency (Topolski et al., 2010). However, Prensky (2007) states the military has embraced game-based training because games work and they have been shown to be effective across several training problem spaces. Experimentation has shown specific skills, such as spatial ability (Sims & Mayer, 2002) and critical thinking (McAlinden et al., 2009) to be successfully trained in a game-based environment, while other research has shown serious games to proficiently teach more generalized skills like trouble shooting and visual attention (Topolski et al., 2010). The extent to which games are effective training tools is based on the technology used and the design principles applied in its development. This includes understanding the cognitive processes associated with game interaction, and how these interactions can be leveraged to provide pedagogical function intended to promote knowledge and skill acquisition.

Design Principles in Serious Games

The training benefits associated with serious games include those introduced in SBT along with the addition of pedagogy (i.e., interactions that instruct or educate, with the intent of imparting knowledge and skill; Susi et al., 2007) in the three main elements of entertainment games: software, art, and story (Zyda, 2005). With pedagogical heuristics guiding art and story, the strength of a game lies in its ability to engage a user through practices that promote flow, motivation, and fun (Murphy, Chertoff, Guerrero, & Moffitt, 2011). However, Zyda (2005) emphasizes that pedagogy must remain subordinate to the story and that entertainment value comes first (Susi et al., 2007). With these additional elements, it is important to understand how games work and why they make effective training tools.

Interestingly, the principles applied to designing effective games follow many of the same guidelines applied to the design of effective instruction; they incorporate mechanisms to facilitate practice, feedback, choice/involvement, positive feelings, emotion, and intensity (Murphy, 2011). The goal is for game-based training systems to promote transfer of acquired knowledge and skills to the operational environment, with Alexander, Brunyé, Sidman, and Weil (2005) attributing transfer to four factors: fidelity, immersion, presence, and operator buy-in. What needs to be considered are techniques to reduce the amount of time to reach efficient knowledge transfer. This is achieved by embedding functions of feedback that links what happens in a game to instructional objectives and the inclusion of support mechanisms that assist users in interfacing with the game to promote focused attention on the task relevant information (Hays, 2005). In essence, these principles are applied to promote flow within experiences, leveraging elements to motivate individuals to fully engage in interaction (Murphy et al., 2011).

Flow in Serious Games

Flow is described by Csikszenmihayi (1990) as a state where involvement in a task takes complete precedence over all else, and this experience is a driver for learning new challenges and skills (Csikszenmihayi, 1997). This is the power of videogames. They capture the perceptual resources of an individual with the result of inducing complete focus on an activity where time becomes distorted (Csikszenmihayi, 1997). This immersion and concentrated effort is sought after from both game developers and instructional designers. The pursuit of flow is to balance challenge and skill to create an environment highly conducive to learning through the regulation of arousal (see Figure 2). This is achieved by matching difficulty and challenge in accord with an individual's skill level, negating the effects of boredom when something is too easy and anxiety when something is too hard. This definition is further decomposed by Csikszenmihayi into seven core components associated with an induced state of flow.

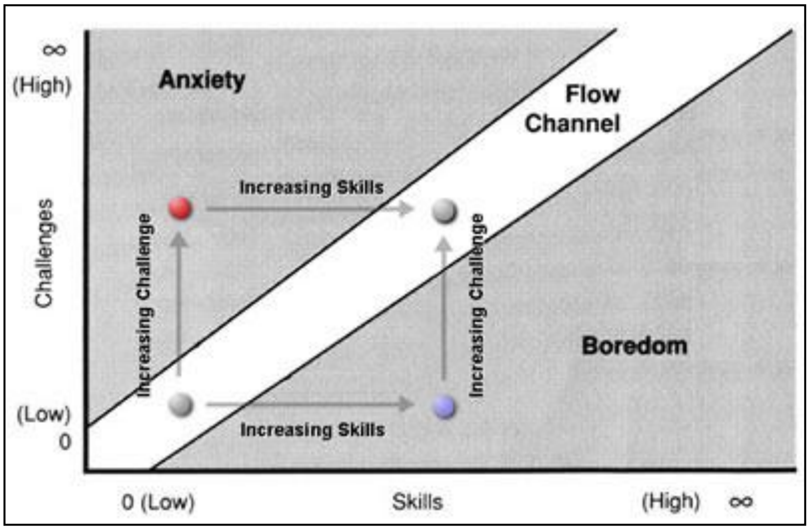


Figure 2. Anxiety, Boredom, and Flow (Csikszenmihayi, 1990 – Dots and Text Added: van Gorp, 2006)

The seven components are broken up into characteristics present when experiencing a state of flow during task interaction and conditions that must be established for an individual to enter a state of flow. Characteristics include: control, diminished awareness of self, and an altered sense of time. Each of these variables is attributable to an individual being immersed in the interaction through cognitive engagement. In a serious game, engaged interaction is important because it keeps learners focused on the material being trained (Murphy et al., 2011), with studies showing engagement to strongly correlate with academic performance (Baker, Corbett, & Koedinger, 2004; Dorneich, Whitlow, Ververs, Carciofini, & Creaser, 2004; Fredricks, Blumenfeld, & Paris, 2004). These characteristics stem from complete focus on a task through platforms affording the ability to have direct control of actions on outcomes, which can create a distorted perception of time where seconds feel like minutes; yet, time passes quickly going unnoticed (Murphy et al., 2011). Because of this induced trance, well-designed serious games provide great potential for immersing individuals in a synthetic learning environment that combines elements of technology and learning science to create a setting for achieving optimal learning (Cannon-Bowers & Bowers, 2008a).

In comparison, the conditions of flow are based on elements that must be in place for an individual to become fully immersed in the experience. These include: clearly defined tasks, attainable/balanced objectives, and feedback (Csikszenmihayi, 1990; Murphy et al., 2011). Having clearly defined tasks with balanced objectives ensures a learner is aware of what they must do, along with the confidence that they have the skill to do it. The other condition required, which is of most interest to this research, is the inclusion of feedback functions that relay the impact of moment-to-moment decisions and actions on outcomes (Murphy, 2011). This incorporates both implicit and explicit modalities, where explicit feedback plays an integral role in managing challenge by defining the causes of error when difficulty is just beyond an individual's level of skill.

Feedback, in terms of flow, associates similarly with Vygotsky's (1987) theory on the Zone of Proximal Development (ZDP). ZPD is defined as the distance between an individual's actual level of performance and the level of potential performance as deemed achievable through guided assistance. ZPD is based on the concept that learning occurs best when individuals are challenged just beyond their capability with socially guided instruction progressing their competency development. From this perspective, Vygotsky (1978) argues that the learning process is a naturally social practice where those capable aid in the development of skill and knowledge by providing feedback to assist in achieving task goals. For the context of serious games, there lacks a true social interaction highlighted in ZPD that facilitates the development of skill through explicit feedback strategies. In recognition of this limitation, this research is focused on applying ITS practices into serious game implementation. In addition, a research question as a result of this focus is: what effect does incorporating ITS function in a game have on an individual's flow within the environment? And, does the inclusion of an EPA for delivering feedback content affect reported levels? A common approach used to gauge the state of flow in mediated learning environments is through self-report instruments, which will be described in detail in the following chapter.

Components of Intelligent Tutoring Systems

The overarching theme of ITSs is to enable computer-based training applications to tailor and personalize instruction to better serve the individual needs and abilities associated with a given learner (Heylen, Nijholt, R., & Vissers, 2003; Loftin, Mastaglio, & Kenney, 2004). The goal is to achieve performance benefits within computer-based instruction as seen in Bloom's

1984 study “the 2-Sigma Problem” (see Figure 3). Though there is recent controversy on the validity of these results (VanLehn, 2011), this classic experiment showed that individuals receiving one-on-one instruction with an expert tutor outperformed their fellow classmates in a traditional one-to-many condition by an average of two standard deviations (see Figure 3). The success of this interaction is in the ability of the instructor to tailor the learning experience to the needs of the individual. Interaction is based on the knowledge level of the learner as well as their performance and reaction (i.e., cognitive and affective response) to subsequent problems and communications (Porayska-Pomsta, Mavrikis, & Pain, 2008).

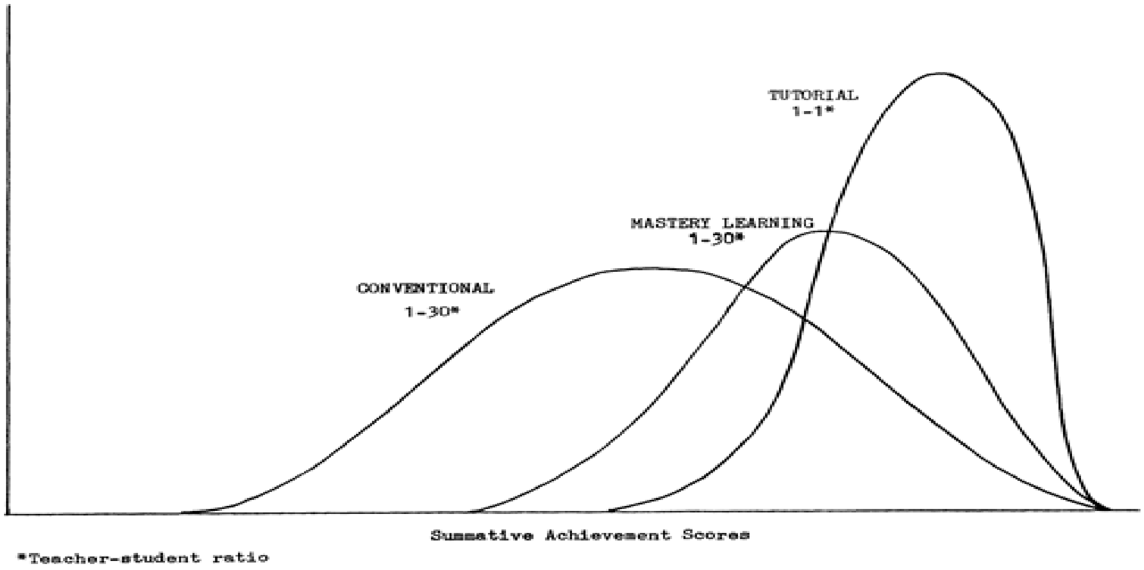


Figure 3. Bloom’s 2-Sigma Problem (1984)

In addition to correcting errors and misconceptions, the power of human tutoring is in peoples’ ability to read and interpret subtle cues from the learner that signify affective response to instruction and is used for applying strategies to maintain motivation. Based on these responses, an effective instructor knows when to intervene and then selects optimal instructional tactics to address learner deficiencies (Goldberg et al., 2012; Porayska-Pomsta et al., 2008). The notion associated with this approach is that information about the learner, both historical and in real-time, can be used to modify learning experiences as to aid in performance and retention (Beck, Stern, & Haugsjaa, 1996).

In general, all ITSs are designed around the concept of the ZPD (Murray & Arroyo, 2002). They function on a cognitive level by managing challenge to make sure material is not too difficult or easy and on an affective level by applying strategies to avoid the extremes of being bored or confused, though it is accepted that some level of cognitive dissonance is necessary (Murray & Arroyo, 2002). Poorly managed interaction that does not account for the relationships of ZPD can lead to distraction, frustration, and a lack of motivation to further pursue objectives (Murray & Arroyo, 2003).

This management of instruction is carried out by four common components to all ITSs (Woolf, 1992): a learner model, a domain knowledge model, a pedagogical model, and a communication model. In the interpretation by Beck et al. (1996), an expert model of performance is included, which is contained within the domain knowledge for the purpose of this description (Woolf, 1992). The information and processes contained within these models are derived from research looking at how effective tutors interact with learners and the information streams they use to base their decisions (Woolf, 2009). In reviewing the various model components in an ITS, it is important to understand how feedback is triggered and the flow of information between models informing these interventions. The functions reported for each model will be applied within this study to monitor performance variables and trigger feedback based on production rules defined within the expert model.

Learner Modeling

A learner model is a system’s representation of an individual’s current knowledge state within a domain, and is used to inform adaptations to better address the strengths and weaknesses of a user (Corbett, Koedinger, & Anderson, 1997; Kassim, Kazi, & Ranganath, 2004). Learner models are designed to serve as the assessment engine within ITSs and are used to determine deficiencies in performance that need attention. Current implementations monitor both performance and affective states to adapt content based on progress towards objectives as well as emotional reactions to training (Ammar, Neji, Alimi, & Gouarderes, 2010).

In terms of applying explicit feedback, learner models must account for specific performance objectives that can be tracked in real-time. They can be designed to outline and recognize learner solution paths to a problem (Conati, Gertner, VanLehn, & Druzdzel, 1997); evaluate performance and problem-solving capacity (Katz, Lesgold, Eggan, & Gordin, 1992); and diagnose misconceptions and constraints associated with a problem space (González, Burguillo, & Llamas, 2006). In conjunction with a defined expert model present within the domain knowledge (e.g., model of desired performance), a learner model can determine gaps in performance for the purpose of selecting focused instructional guidance. In essence, the learner model derives a state of performance by monitoring activity and predicting knowledge levels based on task behaviors (Kelly & Tangney, 2002; Roll, Baker, Aleven, McLaren, & Koedinger, 2005). This information is fed to the pedagogical model for determining strategies to execute (Beck et al., 1996). In the context of game-based applications, linking behavior in a virtual environment to associated training objectives is a challenging task. This requires domain modeling techniques that can translate game state messages into performance metrics on overarching scenario objectives.

Domain Modeling

The domain model contains all relevant information linked to a task or subject (Beck et al., 1996). It represents the knowledge structure of a domain and is accessed by the learner and pedagogical models to manage interventions. In addition to storing domain-dependent materials and content, the domain model also houses representations of expert performance to compare learner interaction against. Rather than just a representation of domain data, the expert model organizes data on how someone skilled in a specific domain represents the associated knowledge (Beck et al., 1996). Interaction is monitored and performance states are communicated to the learner model for determining if a pedagogical intervention is deemed appropriate.

Development of expert models is critical for effective implementation of ITSs in game-based environments. They can be used to compare real-time learner performance versus desired progress (Beck et al., 1996), and are based on detailed descriptions of behaviors and mental activities, task conditions and standards, and other factors leading to successful performance (Sottilare & Gilbert, 2011). A limiting factor associated with expert model authoring is they are often labor intensive and require extensive task analyses to capture all the data to inform assessment practices (Sottilare & Gilbert, 2011).

In addition, game-based applications offer new challenges to expert modeling. Dependent on the application being used, a system must incorporate performance assessments as they relate to a specific game engine’s messaging protocol (Sottilare & Gilbert, 2011). Concepts associated with a system (i.e., inputs, processes, and outputs) differ between platforms, with no standardized approach for interpreting learner interactions (Shute, Masduki, & Donmez, 2010). For explicit feedback to be relevant in context, performance modeling techniques must be addressed in SBT and game-based applications.

Performance Modeling in a Serious Game

No matter the game or the set of tasks to be executed, there should always be sound instructional design practices applied to map objectives and performance criteria associated with achieving levels of proficient execution (Ulicsak & Wright, 2010). In SBT, it is essential to define the root objectives the simulation is designed to train. These objectives influence requirements for what functions and mechanisms must be realistically simulated for the purpose of supporting transfer into the real-world (Shute et al., 2010; Shute, Ventura, Bauer, & Zapata-Rivera, 2009). This is achieved by defining what the system needs to measure and what constitutes successful performance. This information is used to design system interactions and interfaces dependent to performance requirements based on determinations of what constitutes proficient behavior. Shute et al. (2010) refers to this approach as evidence-centered design (ECD), where behavior and performance found to demonstrate knowledge, skills, and abilities associated with a domain are identified (Messick, 1994).

The goal of serious games and SBT is to instill higher-order thinking skills and increase human performance among tasks that are difficult to replicate in live simulation exercises that are often too expensive and too resource extensive to implement on a routine basis. With this functional requirement, it is imperative to accurately monitor performance to recognize error and determine cause. With this information, feedback can be tailored to focus specifically on the identified deficiency and how it impacts task performance. Furthermore, it is important to understand theory and practices applied to monitoring human performance in learning events so as to design games that incorporate sound techniques to gauge successful/unsuccessful interaction.

Human performance is based on associated behaviors required for completing a task and is monitored to assess where an individual falls on the spectrum of novice to expert. According to Rasmussen (1983) three interrelating levels of human performance exist: knowledge-, rule-, and skill-based performance. In the context of learning, these categories differentiate behaviors associated with comprehension and skill, and training systems must recognize the level of behavior they are intended to train so as to identify appropriate objectives to measure (Shute et al., 2010). These performance measurement determinations should be influenced and guided by learning theory, which in turn should influence the methods selected for assessing competencies (Cannon-Bowers et al., 1989). Furthermore, the resulting performance outcomes should then be used to determine deficiencies and misconceptions for the purpose of providing explicit task relevant feedback.

However, many serious games currently developed ignore performance in terms of training effectiveness and provide metrics on implicit objectives associated with a scenario or storyline. Typical approaches to assessing objective-oriented performance (i.e., questions to test declarative and procedural knowledge) in game-based environments requires pausing action, which can be disruptive to an individual’s flow (Shute et al., 2009). Real-time assessment of performance as they relate to training objectives is critical to providing timely and appropriate feedback to assist in the learning process, and tools for accomplishing this must be integrated into game-based architectures. As stated by Shute et al. (2009), this requires serious games to have embedded assessment capabilities to monitor learning for the purpose of maintaining flow, also known as stealth assessments. This is based on ECD in that behaviors elicited during task execution can be used to gauge comprehension and skill associated with the KSAs of a particular domain (Shute et al., 2010).

New tools, such as Student Information Models in Intelligent Learning Environments (SIMILE), are being produced to alleviate this gap (ECS, 2012a). SIMILE is a product co-funded by the Army Research Laboratory (ARL), Naval Air Warfare Center Training Systems Division (NAWC-TSD), and the Joint Advanced Distributed Learning (ADL) Co-Lab for the purpose of providing a generic, adaptable, and standardized mechanism for authoring and performing learner assessments in virtual environments. It works by tracking a learner's progress through a training event and generates a set of performance metrics that determines if a defined objective has been satisfactorily met. Performance metrics are based on associated game messages present when a user interacts with the system. Associated messages are structured as rule-based procedures. This approach is an example of applying stealth assessments where they are seamlessly integrated within the learning environment (Shute et al., 2009). These assessment approaches support learning by maintaining flow through uninterrupted scenario interaction and by removing test anxiety associated with traditional assessment techniques (Shute, Hansen, & Almond, 2008; Shute et al., 2010). With mechanisms for capturing performance in real-time, an ITS must apply pedagogical rules for managing instruction based on progression towards objectives.

Pedagogical Modeling

Pedagogical modeling is associated with the application of learning theory based on variables empirically proven to influence outcomes (Mayes & Freitas, 2004). According to Beal and Lee (2005) the role of a pedagogical model is to balance the level of guidance during a learning event so as to maintain engagement and motivation. Traditionally in ITSs, pedagogical reasoning is informed by an individual's performance within a problem space, and feedback and adaptation strategies are executed when errors in performance are detected (Goldberg et al., 2012). As mentioned above, the learner model is the input source for pedagogical decisions and provides information pertaining to both knowledge and affective states as they relate to the context of the learning event (Beck et al., 1996; Sottilare, 2009). In terms of real-time guidance, the pedagogical model is the driver of explicit feedback selection, with the modality being dependent to the type of application used for training. Explicit feedback is aimed to serve as a facilitator, and content within these messages must be appropriate to the ability level of the learner (Beck et al., 1996).

Typically, strategies for delivering explicit feedback in ITSs are derived from research examining tactics of instruction in a one-to-one learning setting and cognitive theories associated with how individuals transfer information to memory. Consequently, there are a number of studies that have evaluated the tactics and strategies used by tutors as well as the effect varying types of feedback have on training outcomes for the purpose of informing ITS design (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Lepper, Drake, & O'Donnell-Johnson, 1997; Person & Graesser, 2003). Regardless of the approach, all interaction is geared towards aiding the student in solving a problem on their own accord. The timing and specificity of feedback is tailored to the experience and competency level of the individual, with research showing feedback to effectively reduce the load on cognitive resources among novice learners (Sweller, Merrienboer, & Paas, 1998). Interventions in the learning process impacts performance and retention outcomes by providing information on task strategies, procedural errors, and misconceptions linked with learning content (Hembree, 1988; VanLehn et al., 2005). The strategies and methods used by a tutor vary from individual to individual and are adapted to compliment the immediate needs of the learner. However, strategies observed among human tutor studies are difficult to replicate in game-based systems. Currently, there is little empirical research investigating the integration of adaptive pedagogical capabilities into serious games that operate in open virtual environments. Games equipped with programming interfaces that enable the controlling of environmental and character actions offer unique approaches for examining new explicit feedback modalities.

Communication Module

The communication module controls interactions with the learner through determinations of how to present information in the most effective way (Sottilare, 2009). From an explicit feedback point of view, the pedagogical model determines the specific content to present to a user, while the communication model controls the delivery of that information. This current work is focused on assessing communication modalities managed by an ITS in game-based training, and which approach has the greatest return on investment in terms of time and effort required for implementation.

Domain Independency and Current Limitations

Though ITSs have proven to be successful in multiple domains, research on their extension into SBT and experiential learning has been limited (Billings, 2010). Most ITSs are embedded within static environments where actions are discrete events and can easily be tracked for assessment purposes. Many of the successful implementations of these systems are within well-defined problem spaces that involve specific solution paths for satisfying problem objectives (VanLehn, 2011). Such domains include algebra, physics, and calculus. Feedback and manipulation of problem difficulty are provided when errors are present; the most successful of such applications produce an average 1.0 Sigma performance increase over conventional classroom instruction (Koedinger et al., 1997; VanLehn, 2011; VanLehn et al., 2005). This highlights the limitations of existing ITSs in that they train declarative knowledge and principles well, but lack mechanisms for tracking the dynamics and applied experience found in SBT (Nicholson et al., 2007; Nicholson, Fiore, Vogel-Walcutt, & Schatz, 2009). Specifically, the majority of games used for training lack an explicit and formative feedback component (Cannon-Bowers & Bowers, 2008a), which is essential for inducing flow while executing and gaining experience from a training task.

In addition to limited research in game-based ITSs, current systems are commonly developed as one-fit solutions to the domain they instruct, with components being inextensible to other problem spaces (VanLehn, 2011). With a goal to ease the authoring of adaptive functions in common training applications, researchers are working towards the implementation of a domain-agnostic framework that applies standardized modeling techniques for applying intelligent tutoring to any computer-based training application (Goldberg et al., 2012). An effort influencing the questions associated with this research is the Generalized Intelligent Framework for Tutoring (GIFT) (see Figure 4), a modular approach to a domain-independent ITS (Sottilare, Holden, Brawner, & Goldberg, 2011). GIFT consists of all working parts common to intelligent tutors, with additional functions to accommodate application across multiple training systems.

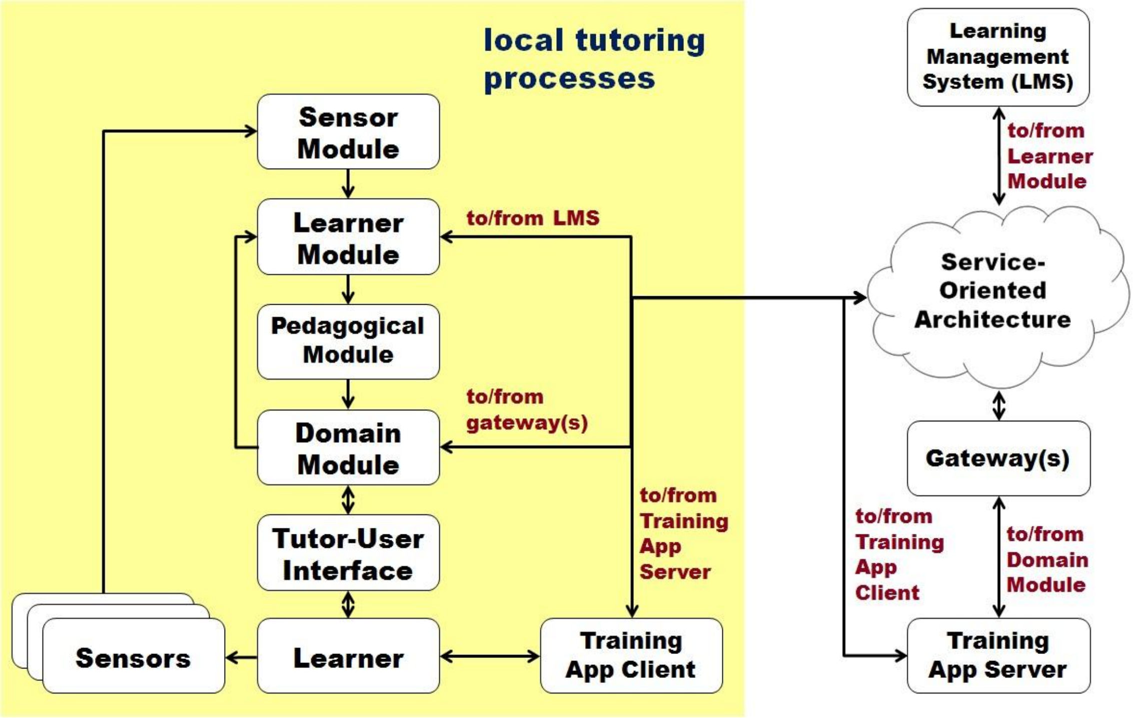


Figure 4. The Generalized Intelligent Framework for Tutoring (GIFT)

A functional component unique to GIFT is the Tutor-User Interface (TUI). The TUI is a browser-based interface designed for collecting inputs and relaying information back to the user. In terms of providing real-time guided instruction, the TUI can be used as a tool for delivering explicit feedback content. It supports multimedia applications and the presence of virtual entities acting as defined tutors. In terms of serious games, the current research is designed to address how the TUI affects interaction and determine its effectiveness versus more labor intensive approaches to embedding real-time feedback. A limitation associated with the TUI during game-based training is that it requires a windowed display of the interfacing game, which may take away from the level of immersion users feel during interaction. As a potential driver for interfacing with a learner, research is required to evaluate feedback delivery in the TUI and assess its effectiveness in relation to other source modality variations. The following chapter will review previous research covering the theories and approaches linked to feedback modalities, and will act as the foundation for the experimental design.

CHAPTER 3: LITERATURE REVIEW

Chapter 3 Summary

Literature on feedback within computer-based learning environments will be assessed, highlighting theory-based support and a review of available methods for content delivery. This will be followed by work centered on feedback specifically delivered by EPAs within Computer-Based Training (CBT) applications. Theory and empirical evidence to support the effectiveness of EPAs within computer-based learning environments will be examined. Gaps and limitations of current practices will be identified, and a foundation for this dissertation work will be highlighted. Following this, work related to measuring the state of flow will be reviewed, highlighting potential trade-offs of embedding ITS functions in game environments on a person's sense of immersion.

Because serious games incorporate a number of interacting elements and entities, their virtual environments offer new avenues for delivering feedback that have not previously been explored in traditional CBT interfaces. Consequently, the overarching goal of this work is to examine the utility of embedding pedagogical function within virtual human characters present in an interacting game environment. Experimentation will be conducted to determine the effect an EPA delivering feedback has on training outcomes and how EPAs internal to the game world affect outcomes when compared to EPAs present in an external interface.

The Role of Feedback in Learning

Everyday tasks are dependent on feedback to determine progress towards objectives. For example, when driving an automobile an operator will use feedback from the vehicle to determine when they have reached their desired speed, when gas is required, and if a door is left ajar. In this instance, feedback is based on the cybernetic definition where output of a system is relayed back to the operator as an input signal to assist in determining next steps to take based on associated goals (Narciss, 2008). This definition is derived from Thorndike’s (1913) law of effect that states the consequence of behavior may influence the application of that behavior in future situations. Feedback is the essential information source that links consequence to behavior.

In the context of instruction and training, feedback is credited as a fundamental principle to efficient knowledge transfer (Andre, 1997; Bilodeau, 1969; Bloom, 1976; Fitts, 1962). The definition varies from that previously stated, in that feedback “is all post-response information that is provided to a learner to inform the learner of his or her actual state of learning or performance” (Narciss, 2008). According to Narciss, the differentiating factor between the two definitions is that feedback in the learning context is provided by an external source of information not directly perceivable during task execution (i.e., internal feedback), and is used as a means for comparing performance outcomes with desired end states. This facilitation is useful for multiple purposes. Feedback: (1) can often motivate higher levels of effort based on current performance compared to desired performance (Locke & Latham, 1990); (2) reduces uncertainty of how well an individual is performing on a task (Ashford, 1986); and (3) is useful for correcting misconceptions and errors when executing inappropriate strategies (Davis et al., 2005; Ilgen, Fisher, & Taylor, 1979).

With an understanding that learning does not take place without feedback, there has been considerable attention over the last six decades in the education research community determining the mechanisms and practices that make feedback most effective. With an emphasis of feedback in a learning context, this review is concerned with principles associated with external/explicit feedback that functions in a confirmatory, corrective, or affective capacity (Billings, 2010). As described in chapter one, implicit feedback is the result of inputs from a user in a training environment and the resulting effect on interacting variables. This feedback is critical in determining progress towards objectives within a scenario, but does not fit within the learning context definition of feedback previously stated. For this effort, *the goal is to determine the benefit of embedding explicit feedback functions implicitly in the training environment*. To this end, the following subsections will review previous research on external/explicit feedback in CBT environments.

Variability of Explicit Feedback

Explicit feedback can take many forms, with a high level classification being content presented to a learner containing either verification of information, elaboration of information, or a combination of both as their performance pertains to a problem space (Billings, 2010). According to Shute (2007), verification feedback incorporates information as it pertains to the correctness of an answer (i.e., outcome), while elaboration references information to assist an individual towards desired levels of deep conceptual understanding. Feedback that incorporates elements of verification and elaboration is termed formative feedback, and is intended to increase the understanding of knowledge and skills as they relate to a content area or general skill (Kulhavy & Stock, 1989; Shute, 2007). Theory surrounds the benefit associated with both forms during the learning process.

Verification of information incorporates corrective feedback for the purpose of amending errors and confirmatory feedback for reinforcing responses and correct actions (Mory, 2004). Confirmatory feedback is commonly applied to strengthen the response to stimuli so it is performed consistently over time, while serving to increase motivation and morale (Kulhavy & Stock, 1989; Mory, 2004). Corrective feedback, in comparison, provides guidance to assist individuals in identifying mistakes and correcting misconceptions as they relate to a problem space (Mory, 2004). Based on the KSAs of an individual, feedback often serves different functions within the learning process. For instance, corrective feedback has been found to be especially beneficial to novices who rarely perform a task successfully on the first try (Billings, 2010).

However, verification of information by itself during task execution cannot fully support the learning process (Billings, 2010). This is argued by Kluger and DeNisi (1996) in which they state if outcome feedback is not supplemented with information used to reject misconceptions, the outcome feedback alone can generate a multitude of hypotheses for why an individual's performance was erroneous. A classic study run by Gilman (1969) found that individuals who received more elaborate information than outcome feedback during science-related tasks netted increases in performance when compared to subjects interacting in competing conditions. All experimental conditions included:: (1) no feedback, (2) feedback explicitly stating if an answer was 'right' or 'wrong', (3) feedback with the correct answer, (4) feedback specific to a user's response, and (5) a combination of all feedback types. Results from the study conveyed that participants in the most detailed feedback condition displayed the best retention of information and exhibited the best performance (Gilman, 1969). A more recent experiment conducted by

Astwood, Van Buskirk, Cornejo, and Dalton (2008) produced results supporting this claim, where participants who received process-oriented feedback (i.e., step-by-step instructions for performing a task) during a simulated Fire Support Team (FiST) exercise significantly outperformed participants who received outcome feedback (i.e., percentage of correct moves), normative feedback (i.e., your performance in relation to everyone else), or no feedback at all. Additional studies have shown outcome feedback when administered as the sole source of performance output to have limited positive effects on learning outcomes (Gonzalez, 2005; Mory, 2004; Shute, 2007).

In terms of generating optimal elaboration of information for feedback delivery, there are a number of components that have been researched over the years. The general approach to formative feedback research is determining the level of specificity contained within the content of a feedback message and determining when best to intervene based on ability levels of the learner. Billings (2010) describes formative feedback as ranging from detailed descriptions telling an individual exactly how to execute a problem to very general and conceptual suggestions (i.e., hints) aimed at guiding a learner towards the correct solution path (Shute, 2007). Research has found that the authoring of formative feedback should account for individual differences among learners, with an individual’s KSAs dictating the level of guidance system interventions are intended to provide. This requires changing the level of specificity of feedback content as learners progress from novice to expert, with theory suggesting the right kind of feedback delivered at the right time is likely to lead to increased performance and learning outcomes (Reiser, 2004). This adaptive approach to guided instruction has been explored in numerous studies, with initial experimentation showing tutored students who interacted with domain experts to benefit most in terms of learning outcomes; as instruction was tailored to individual ability levels (Bloom, 1984; Burke, 1983).

The remaining question is how best to replicate these relationships in CBT environments? As mentioned earlier, majority of instructional strategy research conducted in the ITS community is identifying approaches to model how expert tutors interact with learners (Boulay & Luckin, 2001; Chi et al., 2001; Person & Graesser, 2003). This is evident in the number of experiments examining how manipulations to specificity and timing of feedback impact learning performance in technology-based environments. For the purpose of this literature review, previous empirical work looking at manipulations of feedback will inform the strategies applied within this study. Based on consensus among analyses, strategies will be selected for how best to apply feedback for novice learners performing well-defined, yet complex procedural skills. As the focus of this effort is to examine source modality approaches to feedback, the determined feedback content approach will remain constant across conditions.

Feedback Specificity and Learning

Feedback serves various levels of the learning process (i.e., cognitive, metacognitive, and motivational), resulting in multiple functions for the purpose of regulating interaction through a reinforcing function, an informing function, and/or a guiding or steering function (Butler & Winne, 1995; Narciss, 2008). In terms of elaboration of information, Narciss (2008) defines the simplistic components of elaborated feedback as: (1) knowledge on task constraints, (2) knowledge about concepts, (3) knowledge about mistakes, (4) knowledge on how to proceed, and (5) knowledge on metacognitive strategy. The component to incorporate in feedback content is dependent to the task being conducted, and should take into account the individual differences of the learner as it pertains ability level. Distinctions between the associated functions are dependent on what mechanism of the learning process feedback is intended to address.

For novice learners, research has consistently shown beginners to benefit from more detailed feedback when learning a new subject or skill (Kalyuga, 2009; Moreno, 2004; Reiser, 2004; Shute, 2007). For example, in two experiments assessing the influence of different types of feedback in a discovery learning environment, Moreno (2004) found learners receiving explanatory feedback to score higher on transfer tests when compared to individuals receiving outcome information alone. There are multiple empirical studies dating back to the 1960's backing this assumption, showing more elaborate feedback to produce greater learner outcomes among novice learners (Gilman, 1969; Hanna, 1976). In addition, there a number of meta-analyses documenting a multitude of experiments yielding positive results associated with more detailed feedback in computer-based learning environments (Azevedo & Bernard, 1995; Kluger & DeNisi, 1996; Kulhavy & Stock, 1989; Mason & Bruning, 2001; Mory, 2004; Shute, 2007).

As the associated benefit of detailed feedback is well established in the literature, it is important to note that not all studies support the notion that more detailed feedback is the optimal approach with novices. For instance, Hays et al. (2009) found specific feedback to result in worse transfer task performance in a bilateral negotiations trainer. Delgado (2005) reported in her experiment that individuals receiving process-oriented feedback showed no increase in performance, while those who received outcome feedback alone led to the worse eventual performance. And Pridemore and Klein (1995) showed specific feedback to be no more beneficial than no feedback at all, as performance related to computer-aided instruction for learning how to operate a microscope. However, these findings are contradictory to the findings of many other studies.

Based on the impact feedback specificity has on performance outcomes, certain assumptions will be applied in the experimental design. To test the effect of feedback modality on game-based training interaction, feedback strategies will be applied and held constant for all participants. As the population of interest will be deemed novices in the domain of interest, explicit feedback will be authored based on principles of formative content including verification and elaboration of information. Content will be presented as it pertains to task objectives and how actions in the game link to knowledge components to be tested following scenario completion.

Feedback Timing and Learning

Another avenue of feedback research is examining how people learn and perform when feedback is delivered at varying times during the learning process (Billings, 2010). This area of the literature is primarily concerned with whether feedback should be delivered immediately or delayed following problem solving and system interaction, with studies ranging back to the early 1970’s (Shute, 2007). Mory (2004) defines immediate feedback as corrective content given to a learner as quickly as the system’s hardware and software will allow during interaction, while delayed feedback consists of corrective content given to a learner after a previously specified programming delay interval during instruction. Not surprising, research assessing the effect of feedback timing shows mixed results. For instance, some argue immediate feedback prevents errors from being encoded in memory, while others posit delayed feedback reduces proactive interference by allowing initial errors to be forgotten (Kulhavy & Anderson, 1972; Shute, 2007).

In support of immediate feedback, researchers theorize retention of knowledge and skills is most efficient when corrective information is provided directly following erroneous action or execution (Phye & Andre, 1989; Shute, 2007). There are a number of studies demonstrating this relationship across the acquisition of verbal materials, procedural skills, and physical motor skills (Anderson, Magill, & Sekiya, 2001; Corbett & Anderson, 2001; Kulik & Kulik, 1988). In a meta-analysis conducted by Azevedo and Bernard (1995) looking at feedback during computer-based instruction across 31 experiments, the authors found immediate feedback to produce significantly larger effect sizes in terms of learning gains (Mean Weighted Effect Sizes: Immediate = 0.80; Delayed = 0.35).

Those in favor of delayed feedback response argue in terms of the delay-retention effect (DRE) (Brackbill, Bravos, & Starr, 1962). DRE poses immediate feedback as conflictive in the learning process due to response interference as a result of the inclusion of additional distracters. This is based on the assumption that early encountered errors do not compete with to-be-learned correct actions when delivery of formative feedback is delayed (Shute, 2007). Many of the empirical studies supporting this effect are based on multiple-choice testing, where initial errors are believed to be forgotten over time (Kulhavy & Anderson, 1972; Kulik & Kulik, 1988; Mory, 2004). However, Kulik and Kulik (1988) dispute many of the results found in these studies. Over a meta-analysis of 53 studies, they found a variety of results associated with feedback timing research. What they found was studies using actual classroom assessment and testing materials in their design usually concluded that immediate feedback had a more positive effect on outcomes than delayed. Kulik and Kulik (1988) attribute studies supporting delayed feedback as being most effective to when assessments are contrived by the experimental performance measure techniques, such as list learning (Mory, 2004).

In comparing the benefits of immediate versus delayed feedback, Shute (2007) states "delayed feedback may be more superior for promoting transfer of learning , especially in relation to concept formation tasks, while immediate feedback may be more efficient, particularly in short run and procedural skills." Furthermore, task difficulty has been found to influence when the timing of feedback is most beneficial. Clariana (1999) describes this as, immediate feedback is most beneficial during difficult tasks, while delayed feedback is preferred when a task is at or below an individual's skill level. Because of these distinctions in the literature, the feedback implemented in this study will be immediate, including both verification and elaboration of information.

As made evident from the feedback literature, majority of research in the field is dominated by investigating theory as it applies to specificity and timing variables. What remains a relatively under-studied area is how best to deliver feedback in computer-based learning environments. As explained in Chapters 1 and 2, advancements in technology are providing new means for integrating real-time feedback in SBT and serious game platforms. With ITSs able to embed pedagogical functions across multiple learning environments, it is important to investigate optimal modalities of feedback as they relate to the source of information. This research is motivated by the effort vs. impact tradeoffs associated with delivery approaches. Based on the cost and time required to apply a feedback modality, as well the strengths associated with SBT applications, this study will determine the feedback source that produces the greatest learning outcomes, while maintaining an individual's presence within the learning experience. Because of this aim, it is necessary to investigate research as it applies to feedback modality design and implementation.

Research Surrounding Feedback Modality

In designing feedback source modality approaches it is important to understand how individuals cognitively apply explicit streams of information and the effect the source of this information has on performance and presence levels within a game-based environment. With the development of domain-agnostic tutoring frameworks that integrate with game-based platforms and operate externally to the training environment, it is imperative to assess the utility of available functions as they pertain to known principles of how people learn and the limitations associated with memory and knowledge transfer in the human brain.

In the context of feedback modality, there are a number of tradeoffs to consider in delivery approach. For instance, what is the best approach for relaying feedback back to the learner? Does the inclusion of an EPA significantly affect outcomes in game-based training? Does an external source remove the user from the game environment, thus reducing presence? Theory commonly applied to guide feedback research and EPA design is Social Cognitive Theory (SCT) and Cognitive Load Theory (CLT), which incorporates elements of Working Memory (WM) and Multiple Resource Theory (MRT) as they apply to the perception and interpretation of information as it relates to task execution.

Social Cognitive Theory

In terms of knowledge and skill development, learning is theorized to be inherently social (Bandura, 1986; Piaget & Smith, 1995; Vygotsky, 1987). Social interaction has been found to increase motivation, increase comfort with tasks, enhance flow of information, and improve achievement in terms of memory, problem solving, and understanding during learning events (Bandura, 2011; Gulz, 2004). Social Cognitive Theory postulates that behavioral consequences during task execution serve as sources of motivation and information (Bandura, 1986), rather than as response strengtheners as theorized by reinforcement learning (Gulz, 2004; Skinner, 1953). In terms of cognitive skill learning, social cognitive theory bases skill development through the application of strategies incorporating vicarious learning, and through activities focused on practice and feedback (Schunk, 2001). Because of this, research among the training and education communities is emphasizing the incorporation of social dimensions within SBT and serious game platforms as a form to promote states of flow and presence, thus producing greater learning outcomes (Gulz, 2004).

The incorporation of virtual entities as EPAs is an approach receiving a lot of attention as a mechanism for embedding social cognitive dimensions in computer-based training (Graesser & McNamara, 2010; Kim & Baylor, 2006a; Kim & Baylor, 2006c). Nonetheless, the use of EPAs in technology-based training is not a new concept. The intent is for an agent-learner relationship to mimic Vygotsky (1978) social theory in that more capable others facilitate the development of an individual’s KSAs (Moreno, Mayer, Spires, & Lester, 2001). With learning theory driving the use of EPAs, there have been a number of empirical investigations looking at variables associated with agent design and the resulting effect on metrics of performance, motivation, and immersion. However, there is a lack of extensive empirical research looking at the effect of where an EPA is situated during a learning event and the application of various interfacing capabilities. In addition, there has been little work examining EPAs in SBT environments where users have free control over interaction in a virtual space.

In examining previous empirical literature assessing the effectiveness of EPAs in learning environments, it becomes apparent that social interaction facilitated by a virtual entity has a direct benefit. In initial experimentation incorporating EPAs, studies were designed to determine if they had positive effects on learning achievement levels in terms of problem solving, memory, and deep understanding (Gulz, 2004). An example is a set of two studies conducted by Moreno et al. (2001) where students interacted with multimedia courseware to learn how to design roots, stems, and leaves of plants across multiple climates. In the two experiments (Experiment 1: College Students; Experiment 2: 7th Graders), hypotheses were focused on whether the presence of a pedagogical agent was enough to produce increases in achievement. Participants either interacted with an EPA who spoke and presented information or participants received information through graphics and explanations as on-screen prompts. Results show those in the EPA condition scored significantly higher on transfer tests and interest ratings, but not on retention tests (Moreno et al., 2001). In a similar study Graesser, VanLehn, Rosé, Jordan, and Harter (2001) used the program AutoTutor to assess the impact a conversational agent has on teaching computer literacy when compared against a control condition of students learning the material through assigned readings. Analysis showed AutoTutor to produce an effect size of 0.5 (about half a letter grade). The caveat with this study is that AutoTutor incorporates natural language dialog, which is a confounding factor in determining the impact the presence of the EPA had. No study has been conducted using a condition where there is no present agent, but the conversational dialogue functions remain.

In comparison, a number of studies present evidence where EPAs produce no effect on performance and retention in CBT environments. Van Mulken and André (1998) assessed the influence EPAs had on objective measures of training for technical and non-technical information domains, with results conveying neither a positive or negative implication across both subjects. Höök, Persson, and Sjölinder (2000) presented analogous conclusions in a study looking at participant interaction with EPAs across an information space on the web. Results showed that agent interaction encouraged deeper exploration of the information space, but subjects did not learn more about the space based on an administered post-test (Höök et al., 2000). Moundridou and Virvou (2002) showed similar outcomes when evaluating EPAs integrated within the WEAR (WEb-based authoring tool for Algebra Related domains) ITS for the purpose of delivering feedback. Statistical tests showed that the presence of an agent had no direct effect on short-term performance outcomes when compared against those who received the same feedback without an agent. However, analysis showed interaction with an EPA to produce behavior more congruent with attentiveness to system interaction and positive self-report experience ratings (Moundridou & Virvou, 2002).

Though there are contradictions in the literature pertaining to an EPAs presence increasing learning achievement in computer-based learning environments, research has shown EPAs to consistently affect other facets associated with learning effectiveness. Lester et al. (1997b) refers to this interaction as the *persona effect*, where the presence of a lifelike character in an interactive learning environment can have a significant positive effect on the perception of the learning experience. Specifically, the incorporation of social agents based on the persona effect have been found to increase motivation for using a system, as well as stimulate interest in topics across multiple subjects and learning environments (Gulz, 2004). In terms of motivation, a common conclusion from research shows character enhanced systems to report as more entertaining, lively, likeable, or engaging (André & Rist, 2001; Johnson, Rickel, & Lester, 2000; Lester, 2011; Lester et al., 1997b).

The role an agent plays within the environment can affect its perceived usefulness as well as impact performance outcomes. Research has shown that defined agent roles and personas affect interaction in multiple ways producing different benefits in terms of learning, motivation, and experience. For example, Baylor and Kim (2005) examined how three distinct pedagogical roles (Expert, Motivator, and Mentor) impacted learner perception, performance, and motivation across two experiments using college students in the Multiple Intelligent Mentors Instructing Collaboratively (MIMIC) research environment. The role of the agent was operationalized by voice, image, animations, affect, and dialogue and was implemented through Microsoft Agent (Baylor & Kim, 2005). For the initial study, students from a computer literacy course (N = 78) interacted with an abbreviated version of MIMIC and were asked to report on the agent's perceived role alone, while the second study (N = 71) assessed student perception along with impact on learning and motivation during an instructional planning course implemented through MIMIC. For self-reported student perception based on agent role, results suggest that EPAs can be designed to authentically simulate different instructional roles (Baylor & Kim, 2005). For instance, both the motivator and mentor role were perceived as more human like and produced significant increases in learner self-efficacy following interaction; however, the motivational agent displaying affective encouragement and support failed to produce increases in learning performance. In comparison, the agents with domain expertise (expert and mentor role) produced reliable improvements in learning outcomes and knowledge acquisition, with participants perceiving these roles as more facilitative to learning (Baylor & Kim, 2005).

An additional study examined the effect stereotyping has on EPA perception. To test this Veletsianos (2010) conducted an experiment looking at four conditions where agent role was defined as scientist or artist, and tutorial type covered material either on nanotechnology or punk rock. Results show evidence that visible representation of an agent as it relates to the domain of interest influences student expectations, impressions, and overall learning. One interesting outcome from this study is that participants who interacted with the agent represented as an artist scored higher during a recall task across both tutorial types. The authors posited two explanations for this outcome: (1) participants identified better with the artist in comparison to the scientist through association of an agent's image with one's own, as viewed within the Similarity Attraction Hypothesis (Moreno & Flowerday, 2006), and (2) the artist agent was more visually interesting than its counterpart, directing attention toward the task (Veletsianos, 2010).

Implications to Feedback Modality Research

As evident from feedback research testing practices associated with social cognitive theory, the incorporation of EPAs in a learning environment does more good than harm. While their impact on performance-based metrics is mixed, there is little to no evidence that their presence has negative consequences; yet, research consistently shows EPAs to influence affective response to learning. With multiple studies supporting the persona effect and showing defined EPA roles to impact different components of the learning process, it is important to investigate how a background description of an agent's profile will affect perceived credibility and trust in a training environment. Because of this association, the assumption for this study is that using EPAs as feedback mechanisms during SBT and serious game interaction is the best approach. This highlights a research question looking at whether an agent's description prior to interaction will alone influence the source's credibility? This will be assessed by defining the background and intended role of the EPA during the learning event, while having all EPAs deliver the same feedback strings during interaction. Influence on source credibility will be determined through responses collected from the Agent Persona Instrument (Ryu & Baylor, 2005). The instrument incorporates items looking at the dimensions of (1) credibility, (2) perceived facilitation to learning, (3) engagement, and (4) human-likeness. Full descriptions of the methodology and instrument will be provided in chapter four.

In addition to profiling effects on game play, a main research question this work aims to address is how best to integrate EPA functions during game-based learning experiences? This issue remains an open research topic, and incorporates elements of agent and interface design. With a domain-independent tutoring framework (i.e., GIFT) driving this research theme, it is important to identify the various modalities and approaches the system can use to present feedback information to a user. To this effect, it is necessary to incorporate principles and heuristics associated with how people perceive and process information from computer information sources. Theory applied to guiding this type of application is CLT.

Cognitive Load Theory

CLT is derived from how individuals manage cognitive resources during execution of a task and is based around the notion that an individual's WM is restricted in its resources while Long-Term Memory (LTM) is limitless in storage capacity (Kalyuga, 2009). This is emphasized in the research as limited attention and working memory capacity bottlenecks that continually exert load during information processing (Oviatt, 2006). This association is the basis for CLT, and provides a framework for examining cognitive processes for the purpose of informing instructional design (Paas, Renkl, & Sweller, 2003). Before CLT can be described in terms of feedback source modality, it is necessary to review the theoretical foundations that guide work in this field, including WM and MRT. These perspectives provide a basis for both interface and educational design (Oviatt, 2006).

Working Memory

WM is described as the processes required for temporarily storing, interpreting, and integrating information within short-term memory prior to encoding and transfer (Baddeley, 2000). It is essentially a limited capacity system that temporarily stores and manipulates information for the tasks of comprehension, learning and reasoning. Matthews, Davies, Westerman, and Stammers (2000) describes this as internal computation of perceived information, which accounts for the processing shortfalls inherent with the human brain. Many refer to this limitation as the magical number seven, plus or minus two (i.e., 7 ± 2), where the short term memory can process only five to nine items of information at a time; though this theory has been disputed over the years (Jones, 2002). In considering feedback modalities, it is important to design around the capabilities of WM to facilitate the best opportunity for transferring and retaining information in memory.

The processes involved with WM start in sensory memory, often referred to as sensory stores (Matthews et al., 2000). When external stimuli is present in the environment, information is gathered and sensory memory activates stored information in LTM for transfer into WM (Billings, 2010; Kalyuga, 2009). Information is then integrated, where mental representations and schemas are constructed, which are then transferred back to LTM if enough attentional resources are applied (Kalyuga, 2009). The most recognized model outlining components and interactions within WM is credited to Baddeley and Hitch (1974). They originally accounted for three primary components, including two slave systems and a central executive. The central executive is limited in resources for processing information, which includes directing attention to relevant information, suppressing irrelevant information, and coordinating cognitive processes when multiple tasks are performed at the same instance (Billings, 2010). Because of this, the central executive controls short-term retention of information in the early stages of processing by coordinating and storing information in the associated slave systems (Matthews et al., 2000).

The two slave systems consist of the phonological loop and the visuo-spatial sketchpad. The phonological loop accounts for information as it relates to language, and is maintained through sub-vocal rehearsal as phonological representations tend to decay over time (Baddeley, 1990; Matthews et al., 2000). This system is effective for the retention of sequential information, with its function most clearly suited for memory span tasks (Baddeley, 2000). The second slave system, visuo-spatial sketchpad, stores all visual and spatial information and is used to construct and manipulate images, as well as control movement (Kalyuga, 2009), with the functions being similar to the phonological loop (Baddeley, 2004). In 2000, Baddeley extended his WM model to incorporate a fourth component called the episodic buffer (see Figure 5). This new element is assumed to be a limited storage system that integrates information from phonological, visual and spatial information sources, and is controlled by the central executive (Baddeley, 2000; Billings, 2010). The buffer works by storing information in a multi-dimensional code and provides an interface between the two WM slave systems and LTM (Baddeley, 2000).

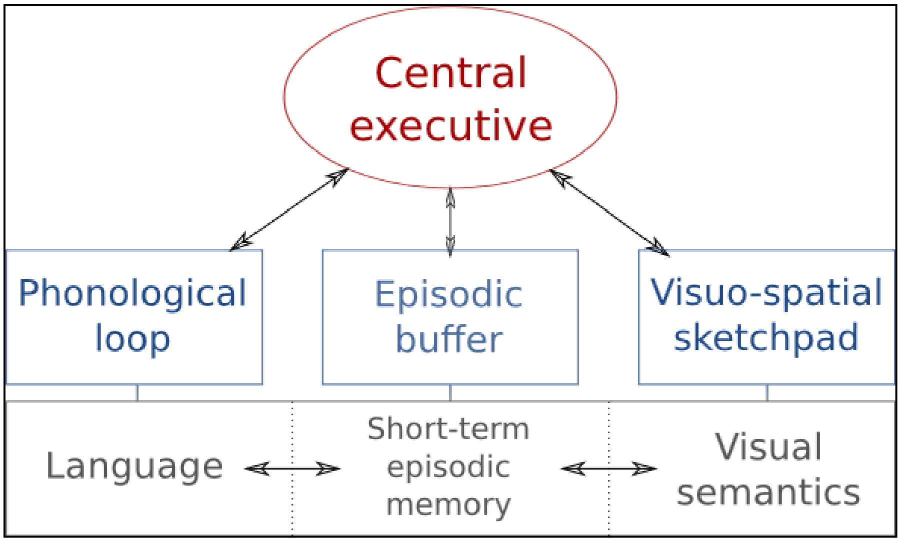


Figure 5. Baddeley's Model of Working Memory

In relation to feedback research, the cognitive processes associated with WM are important to account for in instructional design, and especially interface design for game-based training platforms. With a goal to assess the inclusion of a tutor interface external to the game environment for facilitating social feedback delivery, the next component to consider with cognitive functioning is divided attention and how system design should include principles associated with MRT as it relates to individuals processing and interpreting information while performing complex and/or more than one task at a time.

Multiple Resource Theory

The theoretical foundation of MRT is based on a substantial review of dual task studies and recognizes a competition in cognitive processing between information modalities while executing a task (e.g., physical/verbal user input and auditory/verbal outputs) (Oviatt, 2006; Wickens, Sandry, & Vidulich, 1983). From this Wickens hypothesized that human attentional capacity is conceived as multiple resource pools, with dual-task interference being greatest when task execution requires similar processing resources (Matthews et al., 2000). Wickens (2002) further goes on to postulate that better performance can be attained if information associated with a task is distributed across complimentary modalities to aid in attention and processing. “Broadly, resource theories suppose that attentional resources can be flexibly allocated to more than one task at a time, up to the point that all attention has been allocated” (Matthews et al., 2000). The principles underlying attentional resource models are described by Wickens (1991) in terms of the formula: P = R / D, where performance (P) is a function of resource allocation (R) and task demand (D). Essentially, if more resources are allocated to a task while the difficulty of the task remains the same, performance will increase (Matthews et al., 2000).

In the context of game-based training, the incorporation of explicit feedback in real-time poses an additional task component of efficiently processing new information channels that are not implicit within the game environment. Furthermore, with the incorporation of an additional interface, as proposed within the GIFT architecture for situating an EPA, there is an additional visual element in the training environment that may grab attentional resources, and ultimately create competition between resources and affect training outcomes. Because of this, it is important to incorporate principles associated with MRT to determine best approaches to provide formative feedback while reducing cognitive load associated with interpreting verification and elaboration information.

For predicting performance effects on multiple tasks being concurrently executed, Wickens generated the MRT around a four dimensional model, consisting of the following dichotomies: stages (cognitive vs. response); sensory/perceptual modalities (auditory vs. visual);

processing codes (visual vs. spatial); and visual channels (focal vs. ambient) (Buttimer, 2003; Wickens, 2002). In terms of feedback modality research, the specific dichotomies of interest as they relate to CLT are sensory modalities and visual channels. Stages is concerned with the use of resources as they apply to cognitive activities versus response activities, with Wickens (2002) highlighting research showing the resources applied to cognitive activities are the same as applied to perceptual activities (Isreal, Chesney, Wickens, & Donchin, 1980). With this research being interested in the resources required to perceive and process explicit feedback channels, the stages dichotomy falls outside the scope of this review. In addition, the processing codes dichotomy distinguishes separate processes associated between analogue/spatial and categorical/symbolic (i.e. verbal or linguistic) information (Wickens, 2002). Because the feedback of interest in this study incorporates formative feedback elements, the primary information channel will be categorical/symbolic. If feedback strategies were to incorporate spatial cues (e.g., arrows pointing to extraction point) in the environment, this distinction between verbal and spatial processing of information in WM would be of importance.

In applying MRT to feedback source modality research, the dichotomies of perception and visual coding must be understood to properly design the optimal approaches to reduce conflict among resources. For perceptual modalities, Wickens (2002) describes cross-modal time-sharing as the apparent ability of humans to better divide attention between the eye and the ear than between two auditory channels or two visual channels. As an example, Parkes and Coleman (1990) found that subjects driving a simulated vehicle to perform better when they received discrete route guidance auditorily rather than visually. This is related to visual scanning resources conflicting over the task of driving the vehicle while processing the route information.

From this, it is assumed that visual scanning is enough of a factor that dual-task resource conflict can be reduced by off-loading information channels to an auditory modality (Seagull, Wickens, & Loeb, 2001; Wickens, 2002). Based on this notion, feedback content will be delivered as an auditory channel, with conditions incorporating EPAs for inclusion of social dimensions. To reduce required visual scanning, feedback will not be presented in written text formats. This design decision will be explained in more detail further in the chapter.

When examining the visual channels dichotomy as it applies to feedback modality research, the factors to consider are two components associated with visual processing. These include focal and ambient vision channels as they apply to interacting elements in a learning environment (Wickens, 2002). Focal vision is necessary for interpreting detail and patterns (e.g., reading text, identifying objects), while ambient vision involves peripheral vision for the purpose of sensing orientation. The goal associated with feedback modalities in game-based training incorporating EPAs is to instantiate these interacting elements as ambient visual channels that do not distract individuals from the focal field of the game environment. An example of such a design outside of technology-based learning platforms is aircraft designers identifying several ways to exploit ambient vision for the purpose of providing system feedback, while the focal vision remains loaded on the necessary instruments to maintain flight (Liggett, Reising, & Hartsock, 1999; Wickens, 2002).

Implications to this study include the presence of an additional interface element with a present EPA for feedback delivery versus an EPA embedded directly in the task environment. This distinguishes focal vision from ambient vision that in the latter case the EPA is placed directly in the environment requiring fovial attention. Based on this distinction, it is hypothesized that an EPA in the task environment will reduce cognitive load on the visual channel due to visual scanning being reduced as a result of no extra interface component. However, the inclusion of an extra element in a game-based scenario may require additional focal scanning to distinguish the EPA from the remaining interacting characters, whereas the EPA in the GIFT TUI will remain present, requiring only an ambient channel source to maintain awareness of its communications. This implication will be addressed in the experimental design, with hypotheses being defined specifically around this observation.

Cognitive Load Theory Applied to Instructional Design

With a background on the interacting components of perception and memory, the basis for CLT research on feedback modality is grounded on how best to deliver information that supports WM limitations by reducing competition between resources necessary for processing information. Prior to an evaluation of empirical studies examining cognitive load factors associated with feedback source modality, a review of the underlying principles of CLT will be presented.

In the context of instructional design, CLT is concerned with methods for managing how material is presented to the learner based on limitations of concurrent WM load (Sweller, Merrienboer, et al., 1998). It provides a basis for predictions of performance when considering alternative interface design, with much research devoted to defining design principles and heuristics that effectively manages cognitive load (Oviatt, 2006). These guidelines are based on the assumption that WM has a limited capacity when handling novel information obtained through sensory memory, whereas WM has no known limitations with handling and retrieving information from LTM (Sweller, 2008; Van Merriënboer & Sweller, 2005). This is achieved by

LTM holding information within constructed schemata; schema being defined as the categorical rules individuals apply to make sense of the world around them (Billings, 2010). The development of human expertise is attributed to knowledge stored within schemata where simple ideas are combined or chunked into more complex ones, not through the processing and arrangement of elements unorganized within LTM (Van Merriënboer & Sweller, 2005).

When new information is processed in WM, it can result in the construction of a new schema or is used to modify an existing one (Widmayer, 2005). In terms of knowledge items within WM, schemas work as a single item resulting in less cognitive load when handling familiar situations and the complexity of the schema differs between novices and experts. In the instance where there are too few resources available in WM, cognitive overload occurs, affecting the ability of schema creation and transfer to LTM (Ayres & Gog, 2009). The goal of instruction is to apply methods that promote efficient creation of schema as they relate to new information so as to reduce resource limitations within WM. To this effect, it is necessary to examine the task demands associated with executing an instructional scenario, and the types of cognitive load they produce as it pertains to optimized schemata formation.

The construct of CLT is based around the various forms of load one experiences when interacting with instructional materials, including three distinct types: intrinsic, extraneous, and germane load. Cognitive load as a result of the structure and complexity of the task is called *intrinsic cognitive load* (Brunken, Plass, & Leutner, 2003; Sweller, 1999). Pollock, Chandler, and Sweller (2002) attributes complexity of a task to the level of item interactivity associated with successful performance, and is defined in terms of the amount of information a learner needs to hold in WM to promote comprehension. For novice learners, there is no associated schema and processing domain information will require more resources for construction (Sweller, 2008). This form of load cannot be manipulated by the instructor, but recognizing the inherent difficulty of a task is necessary in determining the flow of instruction, as well as determining the information schemas that should be in place to promote better understanding while reducing the load in doing so. While instructors have no control over the difficulty of a problem or task, they can choose when to apply problems of greater or lesser complexity levels (Sottilare & Goldberg, 2013).

Extraneous cognitive load is a product of instructional design and pertains to how information is presented and how individuals interact with a learning environment (Sweller, Merrienboer, et al., 1998). Essentially, extraneous load is the result of ineffective design or the result of factors associated with interface design that require cognitive processes not inherent to the problem space being instructed. In game-based training design, this is of special importance. The complexity associated with the operating controls as well as the interacting elements in the environment can create high extraneous cognitive load, causing poor schema construction as it pertains to the domain of interest. This is of special importance when considering those individuals who have limited experience with videogames, where the mere task of learning the interface controls may be a challenging task. CLT separates the extraneous complexity associated with an interface from the intrinsic complexity inherent to a learner’s main task because the two forms are additive (Oviatt, 2006; Paas et al., 2003). As a result, domains associated with high element interactivity require design strategies to reduce the extraneous load to promote optimized resource allocation to the primary task (Paas et al., 2003).

The last component of CLT is germane cognitive load. This is an additional type of load influenced by the instructional design, and is considered to enhance learning and schema construction (Sweller, Merrienboer, et al., 1998). Germane load incorporates the processes of learning and is facilitated when WM has enough available resources to process information thoroughly for transfer into LTM (Bannert, 2002). In addition, germane load can be effectively managed by providing explicit feedback on performance to mitigate negative effects from erroneous problem execution. Effective explicit feedback can manage intrinsic load as misconceptions and repeated errors can be recognized and remediated for accurate schema construction. This feedback can often be the missing information a learner needs to confirm knowledge construction or revising already existing schemata.

When it comes to feedback oriented research as it applies to technology-based learning environments, the tenets of CLT drive much of the research questions under investigation. For instance, the research presented earlier in the chapter on authoring explicit feedback is based on CLT assumptions as it pertains to skill level and processes associated with handling novel information in WM. The extent of these studies looked at performance outcomes across conditions applying variations in feedback specificity and timing (Shute, 2007). The conclusions from these studies will guide the feedback content utilized in this experiment, based on novices learning complex procedural skills in a game-based operational environment.

The question this research seeks to answer in terms of CLT is, if explicit feedback provided during a game-based training scenario is held constant for all participants, what effect does the source modality of feedback have on performance metrics and an individual's reported workload? It is believed that variations in the source of feedback will require different cognitive processes depending on the mode it is delivered in and the interface it is presented from. “Combining measures of cognitive workload (i.e., subjective assessments of mental effort) with measures of post-training performance may be more diagnostic of the effectiveness of computer-based training programs in terms of the cognitive costs of instruction beyond what would be found with measures of mental effort or performance in isolation” (Cuevas et al., 2004, p. 12). Sweller (1999) supports this approach based on diminished value of training outcomes if extraneous cognitive load experienced during complex task training is high, even if post-assessments are satisfactory. From this perspective, the evaluation of game-based training systems should focus on how display augmentation techniques affect perceived workload measures during training in relation to performance on post-training assessments (Cuevas et al., 2004; Paas & Van Merriënboer, 1993). For this purpose, it is necessary to define the Independent Variables (IVs) of interest in feedback modality research and available approaches for effectively assessing cognitive load based metrics.

The first variable to consider in feedback source modality is what approach to apply when presenting explicit feedback to the learner. This is determining whether to display feedback in a visual channel (text), an audio channel (spoken words), a spatial channel (e.g., arrows pointing to a rally point), or a hybrid approach combining two or more channels. Much of this research is based on MRT and establishes guidelines for presenting information based on the task environment and the available memory stores in WM. The notion is to exploit alternative modes of feedback presentation (e.g., acoustic, visual, etc.) to avoid cognitive overload due to modality effects encountered when presenting feedback as text (Mayer & Moreno, 2002; Shute, 2007). This is based on the ‘modality principle’, and reflects that individuals learn more deeply when both the visual and/or verbal working memories are not overloaded. When words are presented as onscreen text, it must initially be processed by the visual system, creating a competition for attention with task based elements. This is what Mousavi, Low, and Sweller (1995) refer to as the split-attention effect, which predicts increases in WM load when information is presented in multiple modalities to a subject. It highlights that when words are presented to a person as narration, this information is processed in the verbal channel, freeing visual resources for attending to existing elements in the learning environment and increasing chances for deeper cognitive processing (Mayer & Moreno, 2002).

A meta-analysis conducted by Ginns (2005) reviewed 43 independent effects related to the modality principle based on hypotheses that there are instructional benefits associated with presenting information across modalities. Results from this review support assumptions associated with the modality effect, with two identified moderators influencing the level of cognitive load experienced: level of interactivity with the material and the pacing of instruction. As an illustration, a study conducted by Kalyuga, Chandler, and Sweller (1999) examined the differences in performance between groups receiving text feedback in auditory form, written form, or both while interacting with a mechanics trainer. Analysis found auditory delivery of feedback to be superior over written form, but not when both audio and visual were presented concurrently. In that case, the authors attribute the visual written form of text to be redundant, resulting in cognitive load found to interfere with learning (Kalyuga et al., 1999). Their analysis was based on a training effectiveness score derived from transfer tests associated with the experimental task and self-reported levels of cognitive load. This combination of performance and cognitive load creates an instructional effectiveness metric based on using the Paas and Van

Merriënboer (1993) procedure. The score values are calculated by converting both metrics into *Z*-scores and combining those values into the following formula: $E = (P - R)// SQRT(2)/$ (Kalyuga et al., 1999); where the performance *Z*-score (P) and the cognitive load *Z*-score (R) are represented as a coordinate system to determine training effectiveness (E). The resulting point is measured against the line of zero effectiveness ($E = 0$), and provides a visual representation of the condition effectiveness (see Figure 6). Results put the audio only condition in the area of high-effectiveness while the remaining conditions were located in areas of low-effectiveness showing more cognitive load with lower performance (Kalyuga et al., 1999). Analysis of variance showed significant differences between groups, with results supporting evidence of the modality and redundancy effects associated with cognitive load.

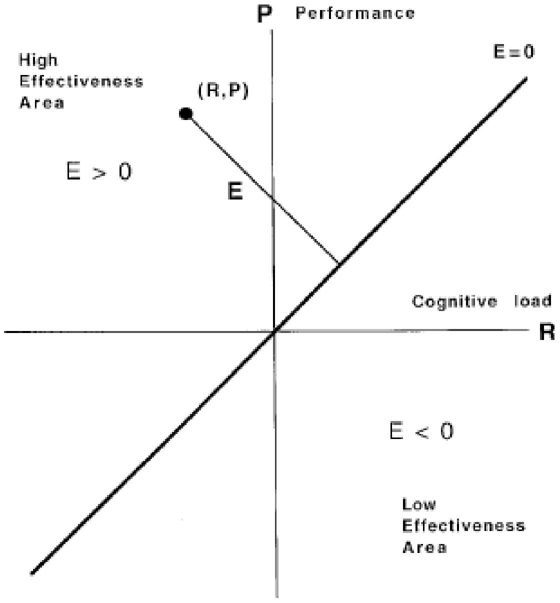


Figure 6. Representation of relative condition effectiveness (Kalyuga et al., 1999)

An additional set of experiments supporting the ‘modality principle’ looked at performance outcomes of students learning from animation and narration and students learning from animation and on-screen text while being presented information about botany (Mayer & Moreno, 1998; Moreno & Mayer, 1999). In all four comparisons, the animation and narration group performed significantly better on problem-solving transfer tasks, with a median effect size of 1.17 (Mayer & Moreno, 2002). A similar study conducted by Mousavi et al. (1995) looked at split-attention and modality effects for teaching geometry on computer-based instruction. Participants interacted with worked examples of geometry problems presented on a diagram along with associated root statements as related to the visual figure. Conditions included a simultaneous group who received visual and auditory proof statements, a group who received only visual statements, and a group who received only auditory. The results showed a mixed auditory and visual mode to be more effective than just a single mode (visual or auditory), which is consistent with work surrounding the modality principle in that the use of dual sensory modes in instructional delivery reduces cognitive load by increasing WM capacity (Mousavi et al., 1995).

An interesting outcome from the Ginns (2005) meta-analysis shows the modality principle to reliably produce positive learning gains when compared against conditions incorporating split-attention tasking. Of the 43 experiments reviewed, only four reported negative learning gains, with the worst outcome displaying an effect size of -0.66 (Tabbers, 2002). In addition, when reviewing the literature on feedback modality, it becomes evident that the majority of studies associate cognitive load effect based on observed differences in performance metric values (Ginns, 2005). Performance metrics associated with cognitive load research include: performance on transfer tests, amount of time to reach a solution, performance gains comparing pre- and post-test, and performance on a primary task when a secondary task is introduced to the scenario (Matthews et al., 2002). What is ignored, and called out by Cuevas et al. (2004), is a lack of research looking at an individual's subjective cognitive load rating in comparison to performance outcomes. For these reasons, it is necessary to include multiple cognitive load measures in any empirical evaluation (Farmer & Brownson, 2003). The following section will review research associated with workload, a common metric used to gauge an individual's cognitive effort, and the available instruments commonly applied today.

Measuring Workload

While dual-task procedures and performance assessment comparisons dominate CLT research in terms of testing the modality effect, it is important to incorporate metrics of cognitive load as it pertains to the amount of mental effort put forth by the learner to accomplish task objectives. A metric regularly utilized to gauge cognitive load during task execution is workload. "Workload is not an inherent property, but rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviours, and perceptions of the operator" (Hart & Staveland, 1988). Common techniques applied to workload assessment are physiological measures and self-report measures.

Physiological measures are based on the assumption that there is a 'physiological' cost for effectively performing cognitively demanding tasks (Matthews et al., 2000). That is, increases in cognitive workload result in physiological change (Farmer & Brownson, 2003). Previous work in this field has been based around the arousal theory assumption that the brain varies its level of activity based on the state of cognitive demand, which can be assessed through central nervous system measures (e.g., Electroencephalogram or EEG) and autonomic nervous system measures such as perspiration and increased heart rate (Matthews et al., 2000). The bodily indicators most often applied that correlate with user load include heart rate variability, pupilometry, galvanic skin response, and functional near infrared imagining; each being studied in lab settings through the use of sensing technologies (Berka et al., 2007). For instance, a learner in a cognitive overload state may experience increased arousal as a result of effort, seen in increases in heart rate and skin conductance (Farmer & Brownson, 2003).

However, EEG is the only signal source found to accurately track subtle shifts in attention and workload on a second-by-second basis (Berka et al., 2007), with empirical evidence supporting this claim (Berka et al., 2004; Brookings, Wilson, & Swain, 1996; Wilson & Eggemeier, 1991). The strength behind all physiological measures is they are not affected by self-report bias, and data reflects real-time indices during task execution rather than reports following completion. The limitation is physiological variables require sensing technologies that are often expensive to acquire and obtrusive to administer.

A more simplistic approach to assessing cognitive demand is by asking the learner performing the task to rate their level of experienced workload. This form of measurement is easy to collect and is minimally invasive when compared to physiological sensing techniques (Matthews et al., 2000). An interesting finding on reported workload as it relates to performance is presented in a study conducted by Eggemeier, Crabtree, and LaPointe (1983). They present a 20-item sequence of three alphabetical letters (e.g., a, c, b, a, a, c, b, b, a …) and ask subjects to count and retain the number of times each letter was presented. The manipulation of task demand was the time interval between displayed letters (1, 2, or 3 seconds). The outcome showed reported workload to increase as the time interval between letter display was decreased, while task performance remained stable across conditions (Eggemeier et al., 1983). This asserts that participants were able to maintain proficient performance in the more difficult conditions by investing more cognitive resources (Matthews et al., 2000). This signifies that individuals can compensate for poorly designed interfacing approaches in instructional design by designating more attentional resources in WM, thus increasing cognitive load (Farmer & Brownson, 2003). In the case of interface design, principles should be applied and research conducted to determine the effect of variations in interacting components on reported levels of workload to determine optimal applications.

Available self-report instruments include unidimensional and multidimensional scales, where multidimensional metrics address individual components of workload giving them a more diagnostic value (Farmer & Brownson, 2003). Associated criteria of effective mental workload metrics include validity, sensitivity, reliability, and diagnosticity (Luximon & Goonetilleke, 2001). Three multidimensional workload instruments commonly used in training analysis found to meet this criteria include the Subjective Workload Assessment Technique (SWAT; Reid & Nygren, 1988), the Workload Profile (WP; Tsang & Velazquez, 1996), and the National Aeronautics and Space Administration – Task Load indeX (NASA-TLX; Hart & Staveland, 1988).

The SWAT applies subjective ratings (i.e., low, medium, or high) across the three dimensions of time load, mental effort load, and psychological stress load; then conjoint measurement and scaling techniques are used to calculate a global rating scale with interval properties (Rubio, Díaz, Martín, & Puente, 2004). The use of the SWAT requires three steps: (1)

participants rank order a set of the 27 possible SWAT combinations creating a scale with interval properties through joint order scaling; (2) an actual rating of workload is reported by the subject based on the task they just completed; and (3) each rating on the three dimensions is converted into a value between 0 and 100 using the scale informed from the first step (Reid & Nygren, 1988; Rubio et al., 2004). The tool has successfully been applied to assess workload variations on memory tasks, manual control tasks and display monitoring across aircraft multitask conditions, nuclear plant simulations, and military tank simulators (Reid & Nygren, 1988; Whitaker, Peters, & Garinther, 1989). However, the SWAT has been criticized for not being very sensitive to low mental workloads and it considered time-consuming due to the rank order step, which can last multiple minutes (Luximon & Goonetilleke, 2001).

The WP is another multidimensional workload instrument established around Wicken's (1983) MRT, and combines elements of secondary task performance-based procedures for the purpose of attaining high diagnosticity (Rubio et al., 2004). The instrument is administered once all associated tasks are completed. All tasks are organized in random order and each participant is asked to rank each task on the workload dimensions highlighted in MRT (stages and codes of processing and input/output dimensions). For each dimension and task, subjects rate their subjective experience between 0 (no resource demand) and 1 (maximum resource allocation required) to represent the proportion of resources used for a particular task condition (Rubio et al., 2004).

The NASA-TLX operates on a six dimension scale to assess subjective perception of workload, and is often regarded as the benchmark tool for self-report measures (Fournier, Montreuil, Brun, Bilodeau, & Villa, 2011; Young, Zavelina, & Hooper, 2008). The six workload-related dimensions include: mental demands, physical demands, temporal demands, own performance, effort and frustration. Participants are asked to rate each factor on a scale from low to high following completion of a task, which is followed by a series of pairwise comparisons to signify the level of importance the subject feels for each of the six workload dimensions (Matthews et al., 2000). An overall workload metric is determined by combining the initial ratings with the associated weights deemed from the comparison series, creating a single workload value (Hart & Staveland, 1988).

Implications to Feedback Modality Research

Empirical evidence supports the ‘modality effect’ of visual and auditory channels as it relates to processing information during task execution, but what happens when an EPA is added to this context? In addition, does this effect apply outside of multimedia systems and into interactive game-based platforms? Moreno, Mayer, and Lester (2000) ran an experiment looking at the role of an EPA’s visual auditory presence in a discovery learning environment. They based hypotheses on CLT’s modality effect and social cognitive theory’s persona effect, predicting students who learn with the voice and image of an agent to remember materials of the lesson better and are more likely to use what they learned to solve problems, thus creating a modality and persona effect on retention and transfer. A third hypothesis was based on social cognitive theory with a prediction that interaction with an EPA will result in higher motivation and likeability of the system. Findings from their analysis showed no positive or negative effect on performance as a result from the visual presence or absence of an EPA, while hypotheses associated with the perosna effect were supported with students consistently reporting the lesson more favorably, they recalled more information, and reported being more motivated and interested in the program (Moreno et al., 2000).

What's interesting from this study is that the mere existence of an EPA through the auditory channel had the greatest impact on learning outcomes. It is in this author's opinion that this is due to the additional demand on visual attention, requiring students to use focal vision resources to scan the learning environment to locate the agent and interpret its interactions. With new technologies being developed that enable a domain-independent tutoring framework to integrate with serious game platforms, new approaches are available for relaying information to the user. Through a Tutor-User Interface (TUI), content can be presented to a user from an external channel to the training environment. Embedding an EPA in the TUI can have one of two effects on game interaction: (1) it provides a grounded base for the visual presence of an EPA, requiring only ambient visual scanning and reducing load for focused attention on the task environment, or (2) the extra interface creates an associated dual-task in the learning environment requiring a user to monitor both the game and TUI equally to maintain appropriate awareness of the interacting elements, thus introducing additional extraneous cognitive load elements. Hypotheses in this effort defined around CLT are based on these assumptions.

Measuring Flow

The final variable of interest in this research, as it pertains to explicit feedback functions in game-based learning experiences, is flow. As reported earlier, when individuals experience a state of flow (i.e., state conducive to learning) in a mediated event, they become immersed in the experience as if they are present in the scenario, resulting in high cognitive effort (Murphy et al., 2011). This is the concept of achieving 'presence', which is loosely defined as a 'sense of being there’ (Lessiter, Freeman, Keogh, & Davidoff, 2001; Witmer & Singer, 1998). In cognitive terms, users become a part of what they are interacting within – devoting all cognitive resources to the elements in the mediated environment (Conkey, 2011). The result is a state of cognitive engagement that reflects processes of information gathering, visual scanning, and periods of sustained attention (Berka et al., 2007). Research into presence and flow has proven important across multiple disciplines and industries, including: Hollywood movies, theme park rides, teleconferencing technologies, and academic and military training (Lombard & Ditton, 1997; Lombard, Ditton, & Weinstein, 2009).

In the education and training context, flow is a critical factor in simulation-based exercises, with research supporting a weak but consistently positive relationship between task performance and an individual’s self-reported level of presence (Witmer & Singer, 1998). The notion behind using computer games as instructional tools is in their ability to immerse individuals in an experience. According to Conkey (2011) the concept of ‘immersion’ comes from situations where technology feeds the human senses with visual, audio, and tactile input through mediated interfaces, creating a perceptual sense of presence within the environment. The perceptual component associated with presence is that interaction invokes response from human senses, human cognition, and affective systems as if the user has a perceptual illusion of non-mediation (Lombard & Ditton, 1997). A causal factor associated with producing and maintaining a state of flow is through promoting a sense of deep involvement within scenario events (Murphy et al., 2011; Witmer & Singer, 1998). A game interaction study conducted by Clarke and Duimering (2006) found players to desire high-sensory experiences, with associated tasks and goals of a scenario influencing how they perceived information in the virtual environment. End-state objectives must be clearly defined prior to interaction so players are more prone to assess elements in the virtual environment as they pertain to reaching defined goal states. It is up to the game developer to combine the game’s sensory experience with targeted objectives for increasing involvement and the chance of becoming immersed in events (Murphy et al., 2011).

To this effect, a serious game’s pedagogical purpose must outweigh entertainment value (Apt, 1970; Susi et al., 2007). Interaction needs to be regulated to promote the proper application of knowledge and skills as they apply to the overall domain and operational environment, not just the scenario. Relying on novices learning solely from implicit feedback provided when errors are made is not enough. Presentation of explicit feedback can make available formative information that corrects misconceptions and affords immediate guidance on context specific problems. In fact, research surrounding flow posits that effective feedback channels are required in game environments so learners can monitor progress towards objectives to assist in reaching objectives when difficulty of a scenario is beyond the abilities of the user. When a novice’s skill begins to advance, the feedback functions in the ITS can be scaffolded back to allow for a more immersive experience, using post-scenario interactions to deliver performance information. Yet, when considering explicit feedback delivery in serious games for novices, one must consider the possible consequences of embedding additional information channels not inherent to game interaction.

With flow being the desired induced state in game design, it is important to determine the effect pedagogical functions delivered during game play have across the associated dimensions that make up the construct. It is hypothesized that incorporating real-time explicit feedback in a game-based event may by itself affect levels of immersion through the processing of information external to game interaction. Adding to that, Lombard and Ditton (1997) warn against the use of conventions that take users out of the story (e.g., voice over narrations), which can ultimately interrupt an individual's sense of flow. From this perspective, explicit feedback provided to a learner reduces the chance of individuals losing themselves in the experience, as if it were real, while providing information to reduce the associated cognitive load of interacting in a problem space. Using an EPA directly in the game environment may alleviate this effect, making the delivered explicit feedback appear as if it were part of the game interaction.

In addition, manipulating the visual field by adding a tutor interface component, as proposed in this study, may grab a user's attention and hamper their ability to attain a presence state within the scenario. For instance, Held & Durlach (1992) would dispute the inclusion of a TUI during game-based training, arguing that a mediated interface should remain low-key and not draw attention to itself and remind the user it's a mediated environment (Conkey, 2011). Based on this stance, one would believe there is a greater chance of inducing a state of flow in users when ITS functions within a serious game appear to come within the interacting environment. This supports the application of EPAs in the game world as social actors, with Heeter (1992) suggesting this approach as an easy way to embed pedagogy and promote presence in virtual environments.

Because of this, the level of effort required to embed EPAs in a game world must be taken into consideration. While technologies such as GIFT provide components to integrate ITS functions in previously developed games, applying game entities as delivery sources for feedback requires additional development within the game itself. Yet, with GIFT's TUI one can incorporate social actors in the training environment with minimal effort. To this effect, research must be conducted to determine the outcome of source modalities, and determine if there is a distinct benefit to flow levels when feedback is delivered from EPAs in the scenario.

For the purpose of this research, there are multiple factors to consider that affect an individual's flow within serious games. In the context of game-based training, flow is a multidimensional concept with each factor facilitating different functions in the learning process. When considering feedback research, the overarching question is whether the information provided that is not implicit within the game environment assists the individual in achieving task goals or if it is distracting enough to remove the user from fully engaging in the experience? In terms of assessing a benefit to learning, it would be deemed a success if user's responded to the provided feedback positively even if immersion and presence levels are affected. While the goal is for feedback to act as a guiding function to promote increases in skill performance while allowing the learner to maintain a sense of presence in the environment, as long as performance was found to increase, the level of immersion someone experiences can be compromised. In terms of effectively delivering explicit feedback without affecting an individual's sense of presence, research needs to identify optimal approaches through empirical evaluations that take into consideration the varying components that come into play.

In terms of immersion and presence, Witmer and Singer (1998) define the dimensions found to influence an individual's subjective rating to be: (1) control factors (e.g., degree of control, mode of control, anticipation, etc.), (2) sensory factors (e.g., environmental richness, multimodal presentation, sensory modality, etc.), (3) distraction factors (e.g., isolation, selective attention, interface awareness, etc.), and (4) realism factors (e.g., scenario fidelity, consistency of information with the objective world, etc.). Based on this construct, there are multiple research questions to reflect on: By adding the TUI to the visual field of the learner, will there be a significant effect on reported levels of presence when compared against an EPA embedded directly in the task environment? Does the mere inclusion of explicit feedback take away from users becoming immersed in the scenario event? Furthermore, adding the TUI to the environment requires a game to run in a windowed-mode on the desktop, which may hamper an individual's ability to become completely immersed in the experience. Because of this, it is necessary to utilize a flow –based metric that takes into account information pertaining to presence levels as induced by a mediated virtual environment.

There are multiple approaches to collecting flow metrics as they relate to interaction within mediated environments. Similar to workload assessment techniques, obtaining metrics on the dimensions of flow has historically been dominated by self-report instruments. However, physiological markers (i.e., heart rate and galvanic skin response) have been analyzed in lab settings to gauge individuals' presence from the body's response to mediated stimuli. Consistent with this approach, results commonly show high stress situations to be most reliable in inducing signals of presence as informed from sensor data (Meehan, Insko, Whitton, & Brooks Jr, 2002; Slater, 2004; Slater & Garau, 2007). If the use of physiological sensors is encouraged, it is important to recognize confounding factors associated with their data. Many of the body's signals believed to correlate with presence are also the same signals that are believed to correlate with workload and affective reaction. For this review, the focus will be on available survey instruments, as these questionnaires expand beyond the dimension of immersion and take into account elements in the game interaction, looking for factors that contribute to the flow state an individual experiences.

A wide array of instruments have been developed and validated over the years for the purpose of collecting subjective levels of flow. These surveys provide a quantified value of users' self-reported experiences, allowing for statistical comparisons across different treatments (Lombard et al., 2009). However, as reported by Procci, Singer, Levy, and Bowers (2012) there is not a reliable tool for measuring an individual's flow state specifically for interaction with a videogame. The conclusion was based on a study examining the applicability of a popular metric of flow, the Dispositional Flow Scale-2 (DFS-2), within game-based evaluations. Based on a literature review and a thorough factor analysis, it was determined that the DFS-2 was not suitable for gamer populations, and that more work was required to refine the measurement techniques for assessing flow in virtual environments (Procci et al., 2012).

As a result, the instrument used for this experiment is a survey currently under development by the Institute for Simulation and Training's Recent and Emerging Technologies Research Organization (RETRO) Lab. The RETRO Flow Scale is constructed from 8 independent scales, which were selected by the criteria of popularity, the type of items making up the scales, and the associated subscales assessed within. The selected instruments include: the DFS-2 and the Flow State Scale-2 (Jackson & Eklund, 2004), the Game Engagement Questionnaire (Brockmyer et al., 2009), the Presence Questionnaire (Witmer & Singer, 1998), the Immersive Tendencies Questionnaire (Witmer & Singer, 1998), the E-Game Flow Scale (Fu, Su, & Yu, 2009), the Response Questionnaire (Tychsen, Newman, Brolund, & Hitchens, 2007), and the Refiana Flow Scale (Refiana, Mizerski, & Murphy, 2005).

All items for each scale was examined based on their factor loadings and correlations from previously run validation studies. If the factor loading or correlation score was at .40 or above, the item was pulled from the list for further analysis. This list was then decomposed by two raters, resulting in a 35-item instrument composed of the following seven subscales: Mastery of Gameplay, Feedback, Concentration, Merging of Action and Awareness, Temporal Dissociation, Loss of Self-Consciousness, Autotelic Experience, and one experimental scale still needing refinement, Visual Quality. The scale distinguishes dimensions linked to flow as either being an antecedent of flow (i.e., an element required to experience a flow state) or being explicitly part of the experience as a result of the interaction. With feedback being considered an antecedent of flow, the scale allows for a granular examination of the subscale to determine if explicit feedback provided contributed to the state reported.

Summary

Based on the associated literature reviewed above, hypotheses have been generated to guide experimental design for the purpose of assessing variations of feedback source modalities in a game-based environment. The specific focus is to evaluate the effect of different implementations of EPAs as feedback delivery mechanisms, and determine the utility of GIFT's external TUI for housing EPA communication during game interaction. Empirical evidence from previous feedback research will be leveraged to design experimental conditions that allow for evaluation of a source modality's influence on metrics associated with performance, workload, the persona effect, and flow.

CHAPTER 4: METHODOLOGY

Participants

Participants for this study were cadets recruited from the United States Military Academy (USMA) at West Point. This was a population of interest because they represent a group of future Army Officers who will potentially interact with training systems embedded with ITS components. Age of cadets at USMA typically range between 18-22, with a small sample of individuals who have previously served prior to enrollment. USMA cadets also account for a standard university population, with results informing system design outside of military application. Participant recruitment was primarily focused on Plebes (i.e. freshman) and Yearlings (i.e. sophomores) enrolled in the introduction to psychology course.

An a priori power analysis was conducted using the G*Power3 application for the purpose of calculating an estimated sample size required to attain statistical power (Faul, Erdfelder, Lang, & Buchner, 2007). The following inputs were used: (1) medium estimated effect size of $f = 0.25$; (2) $\alpha = 0.05$; (3) desired power level = 0.80; (4) numerator df (df = degrees of freedom) = 1; and (5) number of groups = 6. The power analysis inputs, results, and associated graphics are shown in APPENDIX A: POWER ANALYSIS WITH G*POWER3. Based on inputs, the estimated sample size required to achieve a power level of 0.80 is 126 total participants (21 per condition).

Data collection was conducted over a five-day period at USMA where a total of 131 subjects participated. This resulted in 22 participants for each experimental condition minus the control, which totaled at 21 subjects. Across all subjects, 105 were male and 26 were female, and 108 were Plebes (e.g., freshmen) and 23 were Cows (e.g., sophomores). All participants were enrolled in USMA's PL100 Intro to Psychology course, and recruitment was performed through West Point's SONA System. It is important to note that all USMA cadets complete a basic training course (i.e., known as Beast) the first summer they are in West Point, with a small portion dedicated to TC3 related materials. However, when asked to rate their skill in administering first aid procedures, 100 of the subjects reported as being novice, while 31 reported as being experienced. No participants considered themselves as experts in the domain. In addition, questions were administered to gauge an individual's videogame experience (VGE), with majority ranking (95 participants) themselves as having moderately low to no experience, with the remaining subjects (36 participants) ranking themselves as having moderately high to high experience. Based on the variability across this metric, VGE will be considered as a Co-Variate (CV) within statistical analyses linked around game interaction.

In terms of data collection, the lab space was located in USMA's Thayer Hall and was arranged for running six subjects at a time, with two experimental proctors administering informed consents and handling any technical issues that arose during each session. Once a subject logged in, GIFT managed all experimental procedures and sequencing between surveys and training envirronments, allowing the proctors to maintain an experimenter's log for all six machines.

Experimental Testbed

Domain

The domain selected for this experiment was Tactical Combat Casualty Care (TC3). This is defined as pre-hospital care rendered to a casualty in an active combat environment, and focuses primarily on individuals who will die if not treated in a timely manner ((CALL), 2006).

The Army’s mission is to fight and win the nation’s wars. It is a Soldier Medic’s job to provide treatment necessary to sustain the Soldier in support of the mission (Army, 2009). It is a critical role performed under extremely stressful and dynamic circumstances. Practice of complicated and difficult life saving tactics is necessary to attaining skill. Use of simulated training events for practice under variable conditions is desired by the Army’s trainers, but live exercises are often expensive to implement and factors relevant to the domain are hard to replicate. Factors that can affect combat casualty care include: (1) hostile fire preventing treatment, (2) limited medical supplies and equipment, (3) tactical considerations taking precedence over casualty care, (4) and time until evacuation (Sotomayor, 2008). Because of this, combat medics must attain skills not trained in civilian trauma care. The focus is to train medics to effectively perform treatment at time of injury without bringing harm to themselves or others in a unit. This makes TC3 an excellent candidate for game-based training applications to simulate facets of the domain involving critical thinking and on-the-spot decision making (Barad, 2010; Sotomayor, 2008).

Computer-based and serious game applications are developed to assist in skill development and enable practice opportunities by incorporating environmental elements difficult to simulate in live training events. Though game-based trainers lack physical interactions associated with providing hands-on TC3 treatments, their unique benefit is in replicating multiple environments and conditions to expose Soldiers to possible decision points they may face in theater (Fowlkes, Dickinson, & Lazarus, 2010). This type of training should prepare Soldiers for rigors of live field training, making these interactions more focused and beneficial to procedural skill development.

Computer-Based Serious Game

The serious game selected for this study was the Tactical Combat Casualty Care Simulation (TC3Sim), also known as vMedic, a SBT application designed by Engineering and Computer Simulations (ECS), Inc. The serious game is designed to teach and reinforce the tactics, techniques, and procedures required to successfully perform as an Army Combat Medic and Combat Lifesaver (CLS) (ECS, 2012b). The game incorporates story-driven scenarios designed within a game-engine based simulation and uses short, goal-oriented exercises to provide a means to train a closely grouped set of related tasks as they fit within the context of a mission (Fowler, Smith, & Litteral, 2005). Tasks simulated within TC3Sim include assessing casualties, performing triage, providing initial treatments, and preparing a casualty for evacuation under conditions of conflict (ECS, 2012b).

An innovative tool used in conjunction with TC3Sim for the purpose standardized assessment is SIMILE (ECS, 2012a). SIMILE, described in chapter two, was the application used to monitor participant interaction in the game environment and ultimately was used to trigger explicit feedback interventions as deemed by GIFT’s domain knowledge and pedagogical model. For the context of this study, SIMILE will use established rule-based assessment models built within TC3Sim to generate real-time performance metric communication to GIFT. SIMILE monitors game message traffic (i.e., ActiveMQ messaging for this instance) and compares user interaction to pre-established domain expertise. GIFT structures domain expertise by defining training objectives within the domain and learner model based on an ontology schema. As user data from gameplay is collected in SIMILE, specific message types pair with an associated rule that provides evidence determining if the rule has been satisfied; that information is then communicated to GIFT, which establishes if there was a transition in performance on a specific objective defined in the schema. Next, that performance state is passed to the learner model. GIFT interprets SIMILE performance metrics for the purpose of tracking progress as it relates to objectives. When errors in performance are detected, causal information is communicated by SIMILE into GIFT, which then determines the feedback string to deliver. This association enables the system to track individual enabling objectives, giving the diagnosis required to provide relevant explicit feedback within the context of the game action.

To assess the defined objectives outlined in the SIMILE expert models, a scenario has been specifically designed in collaboration with ECS and includes aspects associated with training objectives described below (for description of scenario events and associated SIMILE rules, see APPENDIX M: GIFT CONCEPTS AND SIMILE RULE CONDITIONS FOR TC3SIM SCENARIOS). Task elements were reviewed with a current rising senior at West Point, who directed and verified tactics and procedures in the game as they relate to live training received. All participants interacted with the same scenario, with two conditions including an EPA present in the virtual environment as an NPC. The remaining conditions received feedback from external sources to the game, as outlined below in the experimental design (i.e., from TUI, audio file). It is important to note that participants interacting with the external EPA source condition viewed the TC3Sim training application in a windowed mode to enable presence of the virtual entity.

Training Objectives

Training objectives are the defined standards that denote required competencies for conducting a task, and serve as guidelines for developing game-based training applications and authoring scenarios to train specific competencies and KSAs. Training objectives for TC3 were selected around knowledge and skills associated with hemorrhage control in a combat environment, which served as guiding principles for scenario and assessment design. Objectives were informed by competencies identified in ARL-STTC's Medical Training Evaluation Review System (MeTERS) program, which decomposed applied and technical skills for Combat Medics and CLSs into their associated tasks, conditions, and standards for assessment purposes (Weible, n.d.). In development of the game TC3Sim, the identified competencies were further decomposed into specific learning objectives in terms of enabling learning objectives and terminal learning objectives for each role and task simulated in the game environment. The resulting learning objectives were used to develop validated SIMILE models applied to monitor performance. It is important to note the designed experimental scenario leveraged previously validated learning objectives from MeTERS, but the resulting scenario is for experimental purposes and was not validated by the Army Medical Department Center & School.

As the U.S. Army TC3 domain consists of multiple components, the topic area of hemorrhage control was selected to focus the targeted tasks and skills participants were asked to perform. Selecting a subset of the course allows for more focused assessment on the key skills for controlling bleeding, and reduces the overall session length. This is necessary because West Point cadets have limited time for anything extra in their daily schedule, including participation in experimental sessions. The objectives associated with hemorrhage control consist of treating casualties quickly and appropriately, applying methods to reduce the chance of increasing the casualty count, and evacuating those who need further treatment safely (Army, 2010). Participants received training on two distinct phases of performing hemorrhage control: Care

Under Fire (CUF) and Tactical Field Care (TFC). Each phase incorporates different goals requiring variations in cognitive function.

The distinction between CUF and TFC is based on the level of hostile presence. In CUF, a unit is under direct fire, limiting the amount of care a medic and CLS can provide.

> "Remember, in combat, functioning as a combat lifesaver is your secondary mission. Your combat duties remain your primary mission. Your first priority while under fire is to return fire and kill the enemy. You should render care to injured soldiers only when care does not endanger your primary mission" (Army, 2010, pp. 1-4).

Treatments in this phase are primarily composed of using tourniquets to control bleeding from wounds on the extremities and moving those injured to a safe location (Army, 2010). Bleeding from extremity wounds has been recognized as the number one cause of preventable death with research stressing the necessity of training every Soldier how to apply a tourniquet (Fowlkes et al., 2010). In addition, during CUF the risk of sustaining additional injuries to the unit is extremely high. The major considerations during this phase of treatment is to suppress enemy fire, move the injured to a safe location, and provide immediate treatment to life threatening conditions (Sotomayor, 2008). In this circumstance, care is limited due to engagement with unfriendly forces, which prevents medics and CLSs from performing thorough treatment practices. The benefit with serious games is that specific scenarios can be authored to test the tenets associated with CUF that simulate decision points that require on-the-spot judgments that dictate next actions taken.

Army (2010) outlines the responsibilities and procedures of a combat medic while receiving enemy fire and identifying a wounded unit member. These include: (1) actions under fire (e.g., returning fire, directing casualty to move to cover or engage if able, instruct those with serious bleeding to apply a tourniquet themselves if able, etc.), (2) actions before approaching the casualty (e.g., survey the area for small arms fire and explosive devices, identify route with best cover, request covering fire, etc.), and (3) providing care under fire (e.g., determine casualty responsiveness upon arrival, apply tourniquet immediately over uniform if life-threatening bleeding is determined, move casualty to safe location if possible, etc.). Training materials and scenario interaction will focus on these components of hemorrhage control during CUF.

In comparison, TFC is provided when the individual performing treatment and the casualty receiving treatment are located in an area deemed to be out of harm's way. Enemy fire is currently suppressed and a medic can provide casualty care to the best of their ability (Army, 2010). In this situation, care is directed to conditions that could not be addressed during CUF. With current suppressed fire a medic can execute more thorough examinations, but the risk of enemies reengaging still exists, requiring rapid decisions and treatments on the wounded (Sotomayor, 2008). During this phase of treatment Army (2010) recognizes the following tasks as critical to TFC: (1) reassess tourniquet if appropriate (e.g., expose wound and determine need, apply pressure bandage if not required, reapply a new tourniquet directly on the skin if required, etc.), (2) check casualty for untreated wounds, (3) continue to evaluate and treat, (4) communicate the situation and coordinate extraction, and (5) monitor the casualty. Based on the associated descriptions of CUF and TFC, each phase associates distinctly different goal states coupled with care. These distinctions in associated objectives between the two phases of care were used for authoring production rules that would serve as assessment in the produced SIMILE model.

Experimental Tasks

Tasks associated with this experiment incorporate the common practices applied in a standard training event. Experimental tasks incorporate the introduction of knowledge and procedures associated with conducting a kinetic task, followed by the opportunity to demonstrate application in a simulated virtual environment. Subjects are provided opportunities to practice with guided real-time feedback (based on assigned condition) in a designed training scenario prior to a performance-based assessment. The tasks allow individuals to demonstrate levels of knowledge and understanding for applying the practices coupled with hemorrhage control in the TC3 domain. The experimental procedure is managed by GIFT and consists of the tutoring platform administering a pre-test to measure initial competency levels, presenting training courseware, guiding users through TC3Sim training scenarios, and administering post-tests to determine learning gains. Once logged into the system, the experimental proctor has no interaction with participants.

Participants were presented training material (described below) that introduced the knowledge and procedures associated with hemorrhage control during CUF and TFC. They were then asked to demonstrate the tactics and procedures within TC3Sim. Subjects first interacted with a game tutorial to become familiar with the user interface and available options within the game environment. This task is intended to reduce the extraneous cognitive load associated with learning a new game’s controls, and the task was self-regulated, allowing a participant to spend as much time in the tutorial environment as they wanted. Inputs covered were those required by users to control movement (i.e., keyboard array), point of view (i.e., mouse movement), and character actions (i.e., selection of action wheel items with mouse click). Action items become available based on the area of the body a player is looking (see Figure 7).

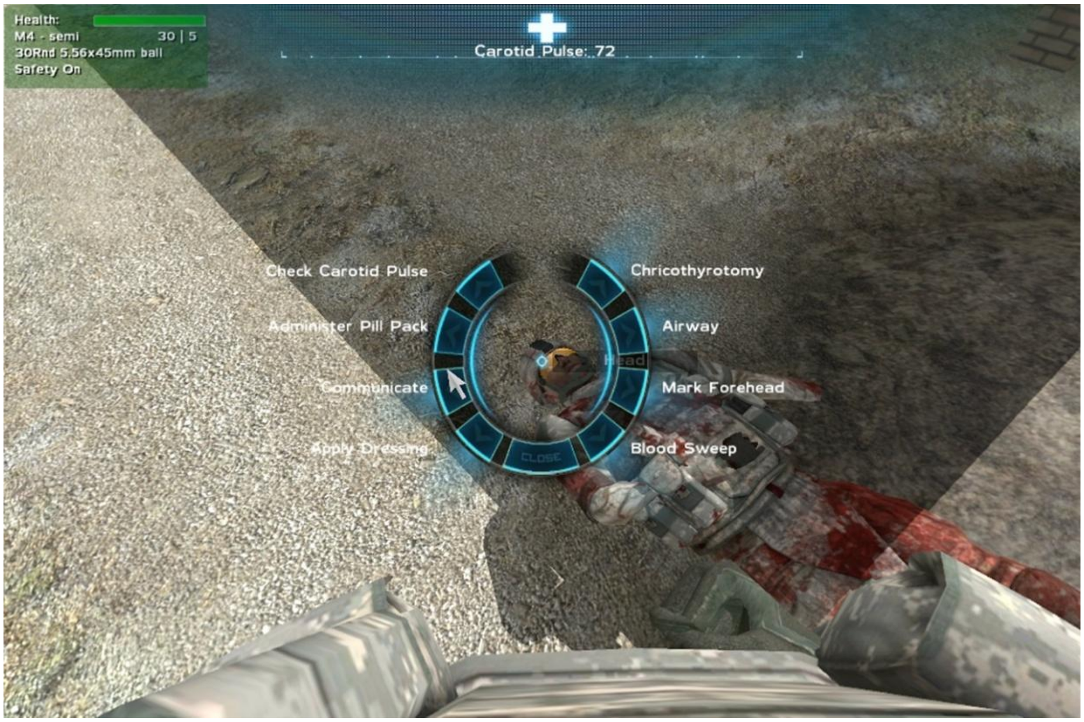


Figure 7. TC3Sim Action Wheel

After completion of the tutorial, participants interacted with two designed scenarios covering the knowledge and skills covered in the courseware, with each scenario including the same task characteristics. Each was designed with multiple casualties in a hostile urban environment and required participants to react and make tactical decisions based on scenario conditions. The first scenario included ITS support facilitated by GIFT and incorporated the feedback source modality manipulations. The second scenario was used as a skill assessment metric, where subjects' performance was assessed by expert models generated in SIMILE. Once interaction with TC3Sim was complete, participants completed a post-test on the training objectives covered.

Experimental Design

The design for this experiment is a 2 x 2 counter-balanced mixed design with two independent variables (IV). The first IV, source of feedback, has two levels and refers to the interfacing component that relays feedback information to the user. In the context of this experiment, source conditions are described as being internal or external to the training simulation being applied (see Figure 8). Source incorporates an EPA as an interacting character in training events, and is present either in the game environment as an entity part of the scenario or is present in the GIFT TUI external to the game. All tutor interactions are managed by GIFT's pedagogical model. Performance is delivered in real-time via SIMILE, relaying information in terms of task execution within the defined scenario assessment models. The feedback logic incorporated corrective responses when errors were present, positive praise when actions were properly executed and reflective prompts as they relate to training objectives.

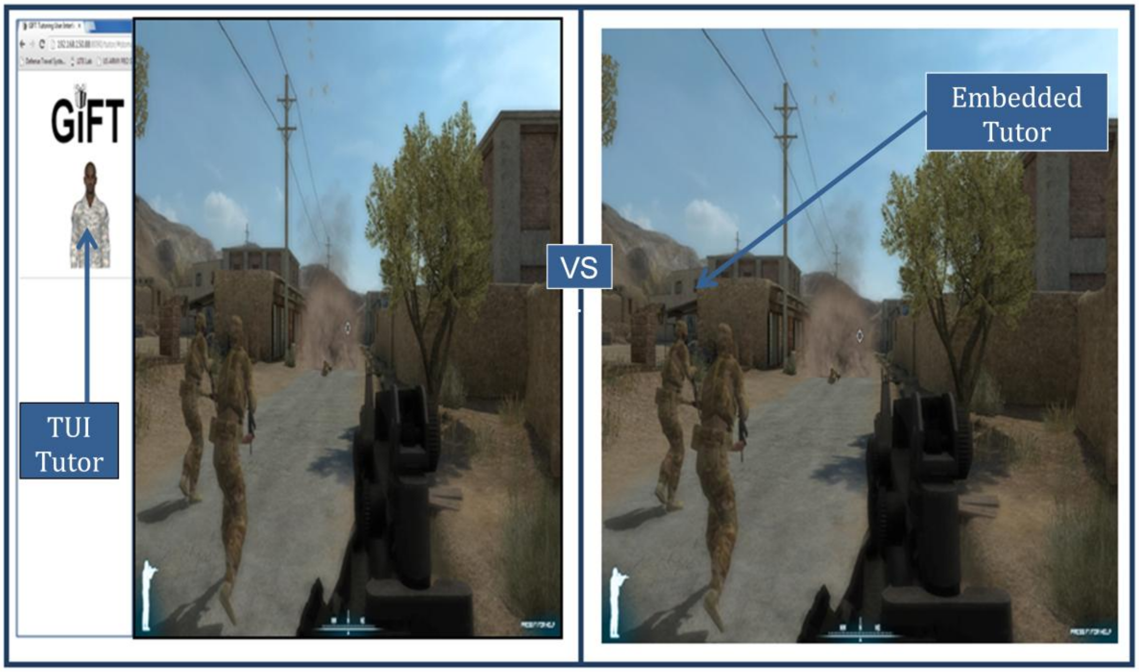


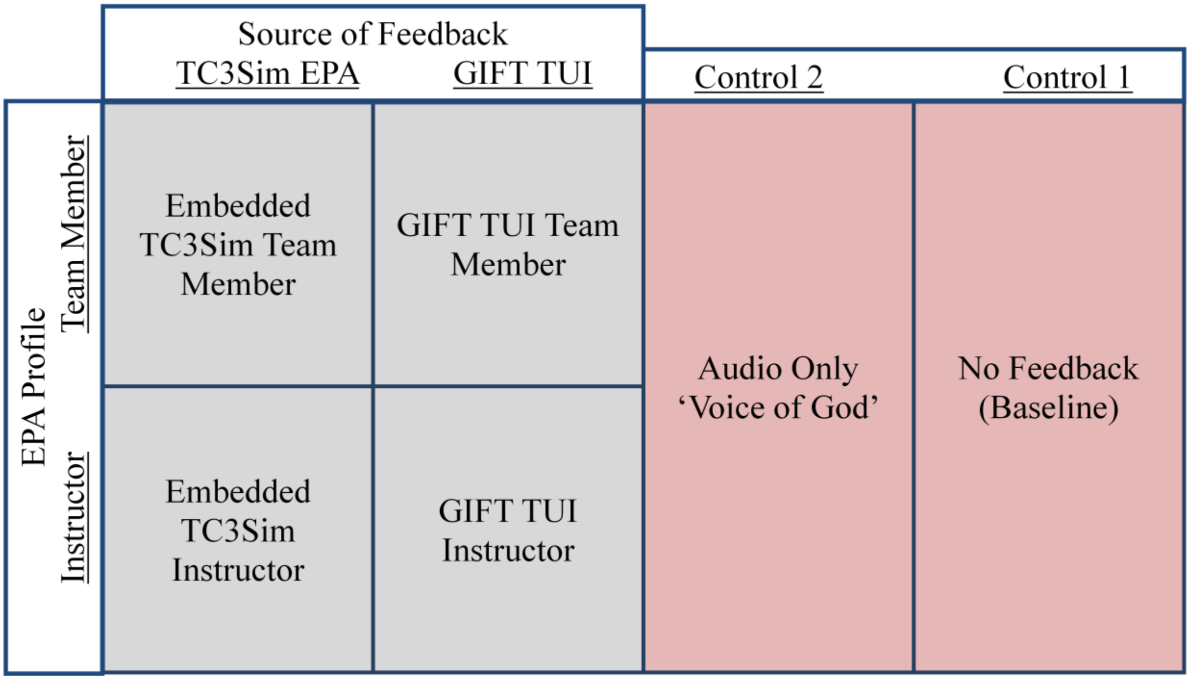
Figure 8. Variable Source Conditions

The second IV, character profile, was based on an associated description of the EPA's background and role within the scenario, and was centered around research on the SCT's persona effect. There were two defined EPAs: (1) an accredited instructor with extensive experience in TC3 training programs and (2) a current combat medic Soldier filling a peer role within the squad team. The profile presents the EPA's professional experience within the domain of combat medic and combat lifesaving skills. No character background was provided for the 'Voice of God' condition (see APPENDIX L: EPA PROFILE BACKGROUNDS/BIOS).

For the purpose of assessing the effect manipulated variables have on associated dependent measures there is the need for base line conditions to determine effect size. To achieve this, there are two control conditions associated with this experiment. The first control condition involves the initial TC3Sim guided scenario without any tutor interaction or explicit feedback.

This is how TC3Sim is currently implemented, with no real-time interpretation of results and performance is provided within an After-Action Review (AAR) following scenario completion. From this condition it can be determined whether enhancements to the current version of TC3Sim had a significant effect on dependent measures.

The second control condition incorporates the initial TC3Sim guided scenario with feedback provided solely as an audio message. This condition is being termed ‘Voice of God’ (VoG) as there is no direct visual component accompanying the voice message; as if it comes from nowhere. This condition enables the ability to determine if the presence of an EPA effects participant performance and survey responses, as well as if the feedback presented solely as an audio file improves performance when compared to the baseline condition. It is important to note that feedback scripts are consistent across conditions. This results in six total conditions (see Figure 9).



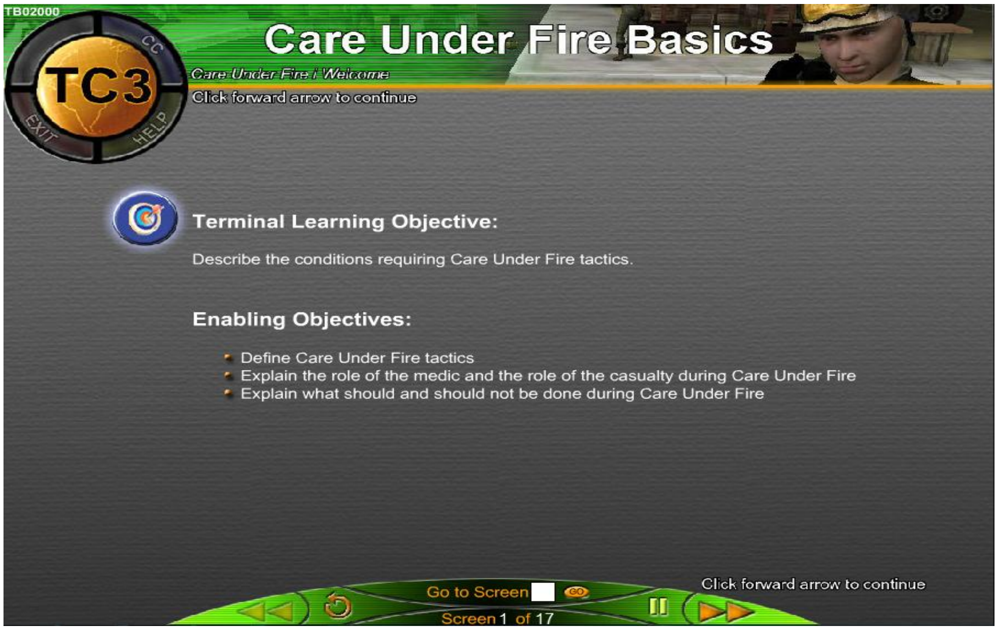
		Source of Feedback			
		TC3Sim EPA	GIFT TUI	Control 2	Control 1
EPA Profile	Team Member	Embedded TC3Sim Team Member	GIFT TUI Team Member	Audio Only ‘Voice of God’	No Feedback (Baseline)
	Instructor	Embedded TC3Sim Instructor	GIFT TUI Instructor		

Figure 9. Experimental Conditions

Equipment and Materials

Training Materials

Participants will first interact with combat lifesaving skill courseware (see Figure 10) designed to teach the declarative and procedural knowledge associated with hemorrhage control. Content is pulled from previously developed courseware versions of the TC3Sim training program. The content is cut down to focus on procedures for hemorrhage control during CUF and TFC to reduce the overall session runtimes. Training was presented through multimedia power point slides that include text, audio, and video presentations. This portion of interaction is self-regulated and did not include any feedback or tutor interventions.



Care Under Fire Basics

Care Under Fire : Welcome

Click forward arrow to continue

Terminal Learning Objective:

Describe the conditions requiring Care Under Fire tactics.

Enabling Objectives:

- Define Care Under Fire tactics
- Explain the role of the medic and the role of the casualty during Care Under Fire
- Explain what should and should not be done during Care Under Fire

Click forward arrow to continue

Figure 10. Courseware Interface

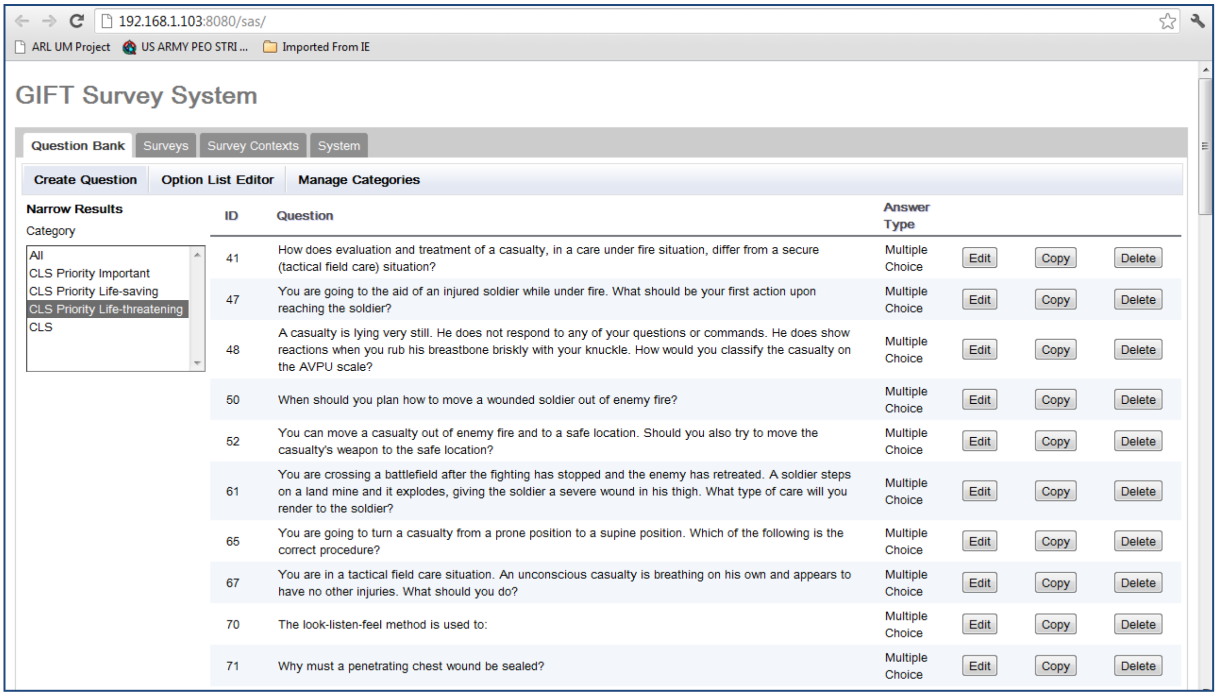
Following completion of the courseware participants began familiarization training with the TC3Sim game controls. This introduction scenario reviews interface components and allows participants to interact with game elements prior to the start of the scenario-based training event. Next, participants proceeded into a TC3Sim scenario requiring the application of associated knowledge and skills presented in the initial courseware. The scenario performance was monitored in real-time for the purpose of providing explicit feedback as environment actions relate to defined training objectives. This is the only portion of the experiment that incorporates the experimental manipulations. This was the final interaction with training materials prior to the game- and test-based performance assessments.

Surveys

Several survey instruments were used in this study. Participants completed questionnaires prior to system exposure, following completion of the TC3Sim scenarios, and following experiment completion. Detailed descriptions of each survey are provided below in the ‘measures’ section. Upon arrival participants first completed a battery of surveys to collect demographic data (age, sex, education level, computer game experience etc.) and individual differences across an immersive tendencies instrument. The demographics questionnaire incorporates associated items previously used in experiments conducted by ARL (see APPENDIX D: DEMOGRAPHICS SURVEY)(Carroll et al., 2011; Goldberg, Sottilare, Brawner, & Holden, 2011). The immersive tendencies survey (Witmer & Singer, 1994) was administered following demographics and collects information on a participants tendency to experience a sense of presence while interacting with a mediated environment (Conkey, 2011)(see APPENDIX G: IMMERSIVE TENDENCIES QUESTIONNAIRE). The collection of this data is purely for exploratory purposes to identify if responses across this instrument predict outcomes on dependent measures collected following exposure to the game.

Another set of surveys was presented to each participant following completion of the TC3Sim training scenario. The instruments selected for this reporting session include the RETRO Flow Scale (see APPENDIX H: RETRO-FLOW S), the NASA-TLX (see APPENDIX E: NASA-TLX INSTRUMENT) (Hart & Staveland, 1988), and the Agent Persona Instrument (API; see APPENDIX F: AGENT PERSONA INSTRUMENT) (Ryu & Baylor, 2005). The RETRO Flow Scale collects a participant's reported level of flow and immersion while in the training environment, the NASA-TLX provides metrics on Workload (WL) and Mental Demand (MD) during scenario interaction, and the API allows a participant to rate the assigned tutor on information usefulness and affective interaction. These measures are used as dependent variables to explain identified variance within the IVs of interest.

The last survey collected subjective ratings on usability and ease of use of interfacing with the training environment through the game controls and game directions, and was administered following completion of all post-training assessments. All surveys were authored in GIFT's Survey Authoring System (see Figure 11) and were presented to the user within the TUI browser window. No paper based versions were administered and all data was extracted from log files post-experiment.



ID	Question	Answer Type			
41	How does evaluation and treatment of a casualty, in a care under fire situation, differ from a secure (tactical field care) situation?	Multiple Choice	Edit	Copy	Delete
47	You are going to the aid of an injured soldier while under fire. What should be your first action upon reaching the soldier?	Multiple Choice	Edit	Copy	Delete
48	A casualty is lying very still. He does not respond to any of your questions or commands. He does show reactions when you rub his breastbone briskly with your knuckle. How would you classify the casualty on the AVPU scale?	Multiple Choice	Edit	Copy	Delete
50	When should you plan how to move a wounded soldier out of enemy fire?	Multiple Choice	Edit	Copy	Delete
52	You can move a casualty out of enemy fire and to a safe location. Should you also try to move the casualty's weapon to the safe location?	Multiple Choice	Edit	Copy	Delete
61	You are crossing a battlefield after the fighting has stopped and the enemy has retreated. A soldier steps on a land mine and it explodes, giving the soldier a severe wound in his thigh. What type of care will you render to the soldier?	Multiple Choice	Edit	Copy	Delete
65	You are going to turn a casualty from a prone position to a supine position. Which of the following is the correct procedure?	Multiple Choice	Edit	Copy	Delete
67	You are in a tactical field care situation. An unconscious casualty is breathing on his own and appears to have no other injuries. What should you do?	Multiple Choice	Edit	Copy	Delete
70	The look-listen-feel method is used to:	Multiple Choice	Edit	Copy	Delete
71	Why must a penetrating chest wound be sealed?	Multiple Choice	Edit	Copy	Delete

Figure 11. GIFT Survey Authoring System

Dependent Measures

Multiple data sources were examined to assess the influence and effect of feedback source modality and tutor character profile on TC3 training. The metrics selected were based on findings from the literature review and are influenced by SCT, CLT, and flow and presence in mediated environments research. These metrics are important to define, as they shape the hypotheses associated with the study.

Performance Metrics

Two forms of performance measures were collected. The initial metric, learning gains, was based on performance generated on the administered post-test assessing knowledge levels in hemorrhage control, with a subject's pre-test score being defined as a co-variate to control for the effect the pre-test outcomes has on post-test performance. Both the pre- and post-test items were generated from exam questions associated with the MeTERS effort. Items were based on the instructional categories of technical skills (e.g., basic anatomy, physiology, pathology), tactical skills (e.g., move, shoot, communicate), and clinical skills (e.g., assess, diagnose, treat, evacuate). Each test included 15 multiple choice questions to assess the various knowledge components associated with hemorrhage control (see APPENDIX I: KNOWLEDGE PRE-TEST and APPENDIX J: KNOWLEDGE POST-TEST).

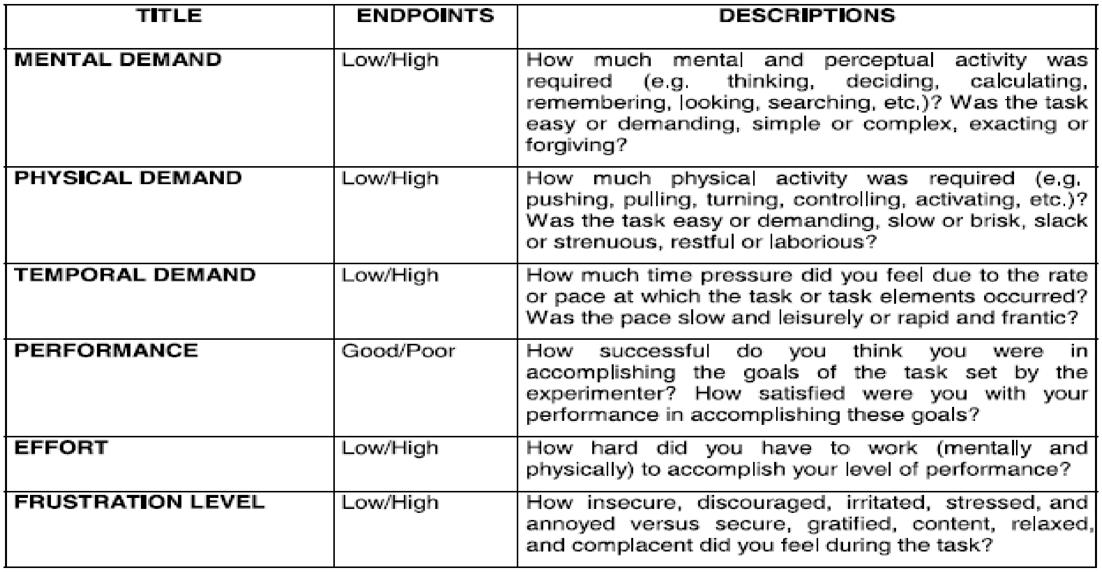
The second performance metric comes directly from the TC3Sim assessment scenario. Interaction was monitored and logged via SIMILE, and player actions were measured against scenario-based expert models. Performance was based on observed procedures during game play, and ‘go’/’no-go’ determinations are marked across all defined critical competency measures (e.g., security sweep, tourniquet application, dress bleed, etc.). The metric output consisted of the number of correct actions taken within the scenario in relation to the full set of competencies being monitored. In accordance with the analysis proposed for the knowledge post-test, the in-game performance analysis will define outcomes on the training scenario with tutor feedback as a CV.

In addition to relative comparisons of performance across conditions, generated assessment metrics are also analyzed in unison with reported MD and feedback measures (described below), as performed by Kalyuga et al. (1999). This approach enables a visual representation of the condition effectiveness by taking into account both performance and associated workload, giving a new metric to base tradeoff analyses from.

Workload and Mental Demand

Measures of an individual's subjective WL and MD were recorded following interaction with the guided TC3Sim scenario. For this purpose, each participant completed the NASA-TLX. A participant's overall workload was determined by a weighted average of responses across six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Definitions of each subscale (see Table 2) were provided to participants to reduce uncertainty associated with the scale meaning. The instrument was selected because it shows good face and construct validity (Cao, Chintamani, Pandya, & Ellis, 2009), and has been found to meet criteria associated with effective workload assessment techniques: sensitivity, diagnostic capabilities, selectivity, low intrusiveness, reliability, and ease of administration (Rubio et al., 2004). Associated reliability of the instrument has been tested with Cronbach's Alpha scoring higher than 0.80 on all factors (Xiao, Wang, Wang, & Lan, 2005). Scores for WL and the independent subscale of MD were assessed individually, providing the data for cognitive load comparisons to determine how the scores were affected by the IVs.

Table 2. NASA-TLX Subscales



TITLE	ENDPOINTS	DESCRIPTIONS
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

Tutor (Source) Credibility

The Agent Persona Instrument (Ryu & Baylor, 2005) was administered to collect metrics on a users perception of the EPA following the feedback-guided scenario. The instrument scores on factors associated with an agent’s role as a knowledgeable instructor facilitating learning and its management of affective, human-like interactions. The Agent Persona Instrument was developed from an item pool of previously used instruments investigating the persona effect. Through experimentation and validation of the resulting survey, a four factor model was produced (i.e., facilitation to learning, credibility, engagement, and human likeness), with high reliability across all subscales: 0.94, 0.92, 0.87, and 0.86 for *facilitating learning* (10 items), *credibility* (5 items), *human- likeness* (5 items), *and engagement* (5 items) (Ryu & Baylor, 2005). The four subscales are used to determine perceptions across two constructs: informational usefulness and social presence of an interacting EPA. Participants rated the 25-items on a 5-point Likert scale, ranging from 1 = Strongly disagree to 5 = Strongly agree. Outcomes from this metric assisted in assessing the overall usefulness of the agent, as well as determining if subject's feel EPAs are a good fit for TC3 game-based training.

Flow and Presence

Immersive Tendencies Questionnaire (ITQ)

The ITQ is an instrument developed by Witmer and Singer (1994) to gauge an individual's propensity to experience presence in mediated environments a priori to system interaction. Participants rate 29-items on a 7-point scale derived from the semantic differential principle (i.e., 1 for never; 7 for always; Dyer, Matthews, Wright, & Yudowitch, 1976). The instrument is intended to identify individual differences across a sample in their ability to immerse themselves in different environmental situations (Witmer & Singer, 1994). The instrument is scored on a single scale, with internal consistency showing satisfactory Chronbach's Alpha ($\alpha = 0.74$). This measure is being collected for exploratory analysis to observe if outcomes on the ITQ influence the recorded outcomes for the dependent measures linked to the experimental procedure.

RETRO Flow Scale

The RETRO Flow Scale is a survey instrument used to assess an individual's perceived state of flow across seven dimensions: Mastery of Gameplay, Feedback, Concentration, Merging of Action and Awareness, Temporal Dissociation, Loss of Self-Consciousness, Autotelic Experience, and Visual Quality. The scale was created around a recognition that not one survey centered around the theories of flow and presence accurately gauges an individual's flow experience across all associated dimensions within a virtual environment (Procci et al., 2012). As a result, the 35-items making up this instrument are based on research investigating optimal approaches for measuring an individual's flow experience specifically within gaming environments. The scale is constructed on a hybrid approach, and uses items from eight instruments previously used to measure flow and immersive experiences (see APPENDIX H: RETRO-FLOW SCALE). The scale was selected due to the granularity the dimensions provide in determining the elements that contribute to an individual's flow state within a game environment. It also distinguishes elements of a game that are required to enter a flow state (i.e., antecedents of flow) from dimensions associated with experiencing a flow state. In terms of feedback research, the antecedents of flow items allow a researcher to observe if the inclusion of feedback promotes a higher sense of antecedents of flow when compared to remaining experimental conditions. In addition, as interface designs play an integral role in the experiment, the dimensions that focus on presence and immersion are beneficial. The scale is currently still under development, and there are no available validation and reliability measures to present.

Experimental Hypotheses

Based on the research questions and existing literature, the following hypotheses were generated for testing in the TC3Sim training environment. Hypotheses are defined around the associated experimental manipulations and their effect on identified dependent measures.

Hypothesis 1

It is hypothesized that the five conditions including real-time explicit feedback (i.e., participants who receive feedback during interaction with TC3Sim) will produce greater learning outcomes in comparison to the baseline condition with only implicit environmental feedback.

Based on formative feedback literature and principles associated with ZPD and CLT, theoretical perspectives suggest that explicit feedback geared towards improving performance is more effective than implicit feedback indicating action outcomes (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1996). In addition, it is hypothesized that experimental conditions with present EPAs will produce higher learning gains over the two control conditions.

Prediction 1

All groups receiving real-time explicit feedback will significantly outperform the baseline condition on performance metrics collected in the training scenario. It is expected that participants receiving explicit feedback will show greater performance during the training scenario. This prediction is linked solely to the training scenario with the tutor, as it is looking to test whether the feedback produced a significant difference in performance for the portion of the game where it was present.

Prediction 2

All conditions with interactive EPAs will produce significantly higher performance metrics when compared to both defined control conditions. This is based on evidence provided by social cognitive theory and persona effect research. It is expected that participants receiving explicit feedback from an EPA will show greater performance during the training scenario and larger learning gains as deemed by transfer assessments of game performance and pre-/post-test scores.

Prediction 3

Performance metrics in conditions with the external EPA present in the TUI will be equal when compared to conditions with an embedded EPA directly in the environment. This is based on all associated conditions receiving the same feedback information regardless of where the tutor is situated. As subjects will have a social agent to ground the delivered information to, it is expected that performance will not be significantly different between these groups.

Hypothesis 2

Based on MRT and CLT in training interface design (Oviatt, 2006), it is hypothesized there will be significant differences in reported WL and MD during TC3Sim interaction across EPA source conditions. Variations in feedback source modality are believed to affect the allocation of cognitive resources based on where the EPA is situated in the learning environment.

Prediction 1

Reported MD and overall WL will be greatest in conditions where the EPA is present in GIFT’s TUI. This is based on users having to allocate visual resources to maintain awareness of the EPAs presence, while managing complex game events. It is also expected that conditions including an EPA will score higher on MD and WL when compared to the VoG treatment, as these subjects will not have the additional visual resources to maintain awareness of.

Prediction 2

This prediction is contradictory to prediction 1. Based on Wickens (2002) description of ambient vision, information perceived through peripheral vision allows individuals to maintain a sense of orientation with that source while maintaining focus on the primary task; as seen in the

Liggett et al. (1999) study. In addition, if an EPA is situated directly in the game environment, does that require extra focal attention to locate the entity among other objects in the scenario? Because the EPA is not in a static location like the TUI, load on the visual resources may increase to maintain orientation of where the agent is. If this is the case, then the prediction is reversed from number one, with expectations of WL and MD scores reporting higher in the internal feedback source condition when compared to the external TUI scores.

Prediction 3

WL and MD will report highest in the control condition with no explicit feedback. This will be due to a lack of information designating performance outside of implicit channels. This can create an element of uncertainty when it comes to the selecting next actions to take, in turn requiring more cognitive load to interpret data in the game environment to determine progress towards objectives.

Prediction 4

There is no expected difference in WL and MD metrics when comparing conditions with the same source modality but having different character profiles associated with the agent. With participants receiving different profile descriptions of their assigned EPA based on condition, this association is not expected to impact a subject's reported score of WL and MD.

Hypothesis 3

Influenced by SCT research on pedagogical agents in learning environments, it is hypothesized that the character profile associated with the EPA condition will significantly affect scores across dimensions of the Agent Persona Instrument (Ryu & Baylor, 2005). With studies investigating the persona effect showing an agent's defined role to impact learning outcomes, character profiles have been established to determine influence on perceived credibility of the tutor agent. This is guided by research looking at stereotypes associated with EPA perception (Veletsianos, 2010). Because feedback in TC3Sim will remain the same regardless of the profile condition, the Agent Persona Instrument will determine the effect character backgrounds have on stereotyping of source credibility, as deemed from interactions with an instructor versus a peer mentor. Dimensions will also be analyzed against source modality to determine if there is an influence on an agent's perception based on where they are located in the environment.

Prediction 1

Based on research from Baylor and Kim (2005) it is hypothesized the 'instructor' background conditions will score significantly lower on the human-likeness dimension and engagement dimension in comparison to the 'peer-mentor' conditions.

Prediction 2

Influenced from stereotype research by Veletsianos (2010) and outcomes from Baylor and Kim (2005) it is hypothesized that participants interacting with the 'instructor' profile will report significantly higher scores on the dimensions of perceived 'facilitation to learning' and 'credibility' for the EPAs when compared to the 'peer-mentor' role.

Prediction 3

Source modality of feedback (i.e., internal vs. external EPA) will affect 'human-likeness' and 'engagement' scores on the Agent Persona Instrument. The 'Internal EPA' condition will report significantly higher ratings across these two dimensions when compared to the ‘External EPA’ condition.

Prediction 4

Source modality of feedback will affect ‘facilitation to learning’ scores on the Agent Persona Instrument. The ‘external’ conditions will rate significantly higher in this dimension when compared to ‘internal conditions’ based on constant visibility of the EPA.

Hypothesis 4

It is hypothesized that the source modality of feedback will significantly influence an individual’s sense of flow, as deemed by the RETRO-Flow Scale, within a TC3Sim environment. Predictions will examine effect of all agent conditions against the controls to determine if the sole presence of an agent affects the dimensions of flow within the game world. The main factor this hypothesis addresses is the impact GIFT’s TUI has on flow levels when interacting with game-based training applications, and the effect this manipulation has on immersion and presence.

Prediction 1

Participants in the ‘Internal EPA’ conditions will report significantly higher scores on dimensions of Flow Experience (e.g., Concentration, Temporal Dissociation, Loss of Self-Consciousness, Autotelic Experience, and Merging of Action and Awareness) when compared against the ‘External EPA’ conditions.

Prediction 2

Participants in the ‘Instructor’ profile conditions will report significantly higher scores on the dimensions of Antecendents of Flow (e.g., Mastery of Gameplay and Feedback) when compared against the ‘Peer’ profile and VoG conditions. This is based on the notion that feedback delivered by an instructor will be perceived as more useful, resulting in a better gameplay experience and higher reported flow scores on those dimensions.

Prediction 3

Due to the absence of explicit feedback channels removing the user from the game experience, it is hypothesized that the control condition with no feedback will score the highest on presence dimensions (e.g., Concentration, Temporal Dissociation, and Loss of Self-Consciousness). This is based on participants having to rely on implicit information within the environment to gauge progress and next actions, resulting in increases of perceived presence.

Procedure

Pre-Test, Surveys, and Training

Upon arrival participants were randomly assigned to an experimental condition. Following, they read and signed the approved informed consent outlining the purpose and risks associated with the study. Next, they began interaction with GIFT by logging in the session based on their assigned participant number. GIFT managed the execution of all experimental procedures once the session was initialized. Instructions and user inputs were established through the TUI, a browser-based interface used for presenting information to the user.

A participant was first prompted to complete a battery of surveys. Instruments included a demographics survey, a videogame experience metric, and the Immersive Tendencies Questionnaire. When complete, the developed pre-test assessing initial knowledge levels was administered. The test included questions assessing all associated training objectives. This initial performance metric was used to determine learning gains following interaction with the training materials.

Upon completion of the initial surveys and pre-test, GIFT directed the participant to interact with a custom courseware developed to deliver TC3 associated content. The course materials were self-guided and included interactive multimedia selected across multiple source applications. All participants interacted with the same courseware, with subjects spending an average of 10-12 minutes with this content.

TC3Sim Exposure

Following training, GIFT initialized the first interaction with the TC3Sim interface environment. Participants performed a short scenario designed to introduce the interfaces and inputs associated with the game. This tutorial session lasted an average of three minutes and took no longer than five minutes. Next, GIFT prepped the subject for the first of two scenarios in TC3Sim. This is where manipulations to the independent variables were introduced. All conditions presented a mission overview highlighting the objectives of the game session (see APPENDIX K: TC3SIM MISSION BRIEFING SCRIPT). Incorporated with this overview was an introduction to the EPA the participant would interact with. A background description associated with the EPA was provided for the purpose of defining the agent’s perceived role (see APPENDIX L: EPA PROFILE BACKGROUNDS/BIOS). This background was the defined second IV and was used to determine if how an EPA is presented to a subject affects their perception of the agent’s usefulness. For participants in the two assigned control conditions, they only received a mission overview before progressing into the game.

The mission overview and EPA background narrative led directly into the first of two scenarios described above to train and test hemorrhage control while performing CUF and TFC. The first scenario incorporated real-time feedback presented through the assigned condition source. During task interaction, SIMILE interpreted user inputs for determining performance and communicated the results to GIFT for executing feedback scripts. Based on the condition, feedback was delivered either as audio only (VoG condition), through an EPA present in GIFT’s TUI, through a character present in the virtual game environment, or no feedback at all. When complete, participants completed survey instruments to collect data on cognitive load (NASA-TLX), flow (RETRO Flow Scale), and source credibility (Agent Persona Instrument) as it solely related to the guided interaction. This led into the second of two scenarios in TC3Sim, which involved similar events to the first session, minus the real-time feedback element. SIMILE monitored interaction and provided outcome results as a source of performance for determining skill at executing trained procedures with no assistance.

Post-Test and Surveys

After interaction with TC3Sim, GIFT presented participants with a post-test in similar fashion to the initial pre-test. A new set of questions was presented and the resulting score was used to gauge learning gains. Next, participants were given the opportunity to record comments as they related to their experience with the experimental procedure.

Participant Debrief

Following the post-test and comments, GIFT completes the session and informs participants to notify the experimental proctor. A debrief form was given to participants and any questions they had were addressed.

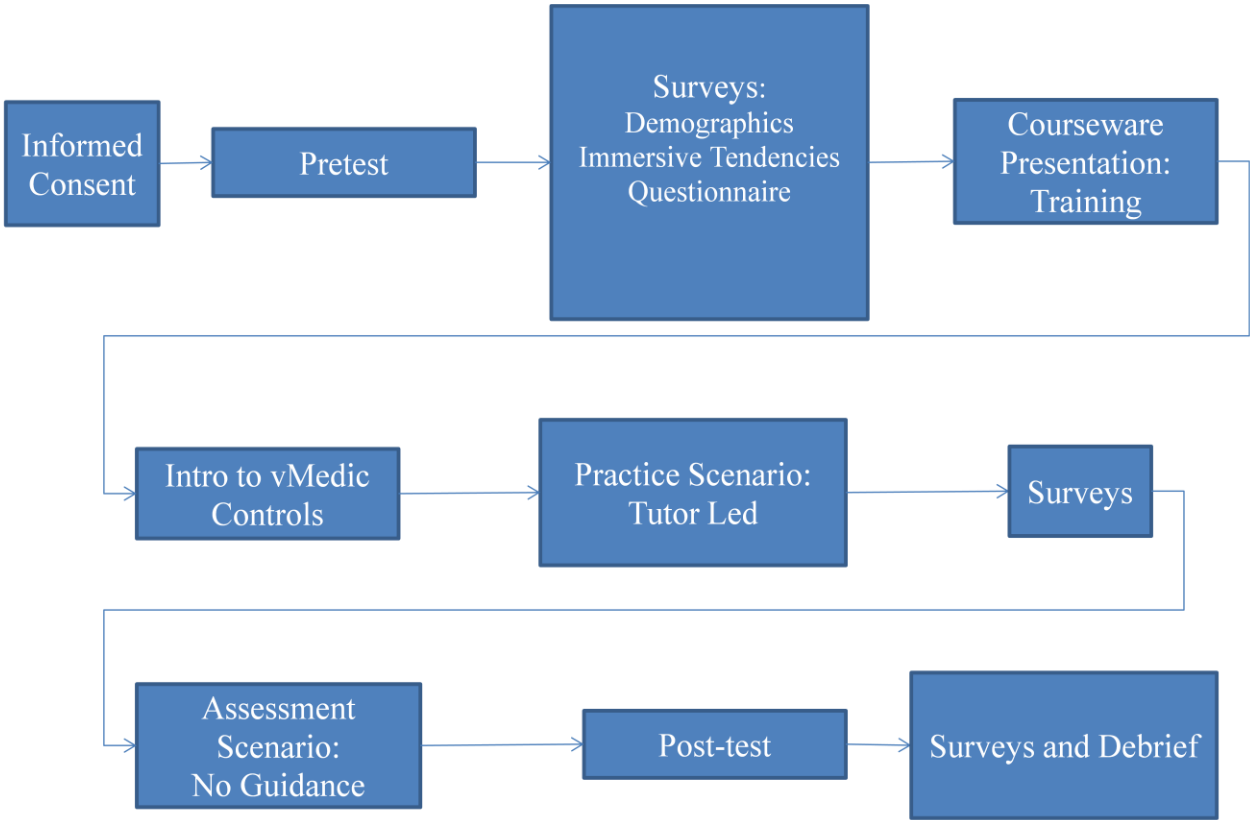


Figure 12. Experimental Procedure

CHAPTER 5: ANALYSIS AND RESULTS

Chapter 5 Summary and Data Analysis Plan

Statistical analyses were performed on the data using IBM SPSS Statistics 19. For indication of statistical significance, an alpha value of .05 was used for all tests, unless explicitly stated otherwise. Prior to conducting hypotheses testing, evaluations were performed on the data to identify potential factors that could affect the output of statistical values and to assert assumptions that influence the statistical approach applied.

First, the experimenter’s log was examined to locate issues within the data collection that warranted the removal of specific interaction values. The most significant factor was the issue of time. Due to data collection restrictions at USMA, the maximum allotted time to complete the procedure for cadets was 60 minutes. As a result, there were four individuals who were unable to complete the TC3Sim capstone scenario. To avoid loss of further data, these specific participants were skipped through this interaction component by the proctors for the purpose of allowing time to complete the post knowledge test and final surveys. No other issues were identified that resulted in the removal of data.

Next, initial testing was conducted for examining the distribution properties of the data across all dependent measures. Because many of the statistical analyses proposed for hypotheses testing run on the assumption that data has a normal distribution, normality was checked using the Kolmogorov-Smirnov test for all associated variables. In analyzing the output, it was determined that there were multiple occurrences where the data associated with dependent measures across the condition groups were not normally distributed (see Table 3). According to

Pallant (2007) this is quite common when dealing with a large sample of data. In addition, Games (1984) highlights the central limit theorem's stance that in big samples the distribution will be normal regardless of assumption testing outputs and that transforming data often reduces the accuracy of *F* (Games & Lucas, 1966). Furthermore, in Billing's (2010) dissertation, she highlights researchers that argue the necessity of running statistical tests such as Analysis of Variance and *F* tests when normality assumptions have not been met. In support of this claim Field (2009) shows from early research that *F* tests run on skewed data performed as they should, while transforming the data both assisted and hindered the accuracy of the resulting statistical output (Games, 1984). Despite the recognized violations, it is important to note that the instance of a variable consistently reporting as not normally distributed across all conditions is not present. For this reason, the *F*-test associated with Analysis of Variance (ANOVA) and Analysis of Co-Variance (ANCOVA) will be used for hypothesis testing.

Table 3. Violations of Normality across Associated Dependent Measures

Source Condition	*Dependent Variable*	*K-S Statistic*	*df*	*Sig.*
TC3Sim-Peer	Post-test score	.222	22	.006
	API Facilitating Learning	.299	22	<.001
	API Credible	.263	22	<.001
	TC3Sim Capstone Scenario	.216	21	.012
TC3Sim-Instr	Workload	.187	22	.044
	API Credible	.201	22	.021
	API Human-like	.206	22	.016
TUI-Peer	Pre-test score	.194	21	.037
	Flow-feedback score	.256	21	.001
	Flow Experience	.210	21	.016
	Mental demand-NASA TLX	.214	21	.013
	API Credible	.277	22	<.001
	TC3Sim Training Scenario	.214	22	.010
TUI_Instr	Pre-test score	.292	22	<.001
	Flow-feedback score	.277	22	<.001
	Mental demand-NASA TLX	.232	22	.003

Source Condition	*Dependent Variable*	*K-S Statistic*	*df*	*Sig.*
	API Facilitating Learning	.218	22	.008
	API Credible	.213	22	.011
	API Engaging	.208	22	.014
VoG	Pre-test score	.258	22	.001
	API Facilitating Learning	.211	22	.026
	API Engaging	.214	22	.010
	TC3Sim Training Scenario	.215	22	.010
	TC3Sim Capstone Scenario	.195	22	.029
No Feedback	Post-test score	.221	21	.009

Hypothesis 1

The first hypothesis examines to what effect the inclusion of feedback within a game-based training event has on performance outcomes in both knowledge- and skill-based assessments. It is hypothesized that individuals receiving explicit feedback aimed at improving performance during game-play will produce higher performance scores for all game interaction as well as achievement on post-test scores. Predictions defined around this hypothesis were focused on three theoretical underpinnings of feedback research; ZPD, CLT, and SCT. Statistical tests were conducted looking at the independent variables (e.g., source of feedback and EPA profile) to determine if they had an effect on performance outcomes (i.e., for a list of all descriptive statistics on associated performance metrics across all six conditions, see Table 4).

Table 4. Experimental Performance Metrics Across All Conditions

Feedback Modality Condition		*TC3Sim Scenario %* *Training*	 *Capstone*	*Knowledge Pre-Test*	*Knowledge Post-Test*
TC3Sim-Peer	*M*	38.48	40.76	63.33	70.91
(N = 21)	*SD*	6.75	6.15	12.30	11.18
TC3Sim-Instr	*M*	36.60	38.00	61.52	65.46
(N = 22)	*SD*	7.14	10.01	12.84	14.35
TUI-Peer	*M*	36.91	39.82	66.36	69.39
(N = 20)	*SD*	6.16	9.27	11.36	11.76
TUI_Instr	*M*	38.10	41.33	65.46	70.30
(N = 21)	*SD*	7.11	8.97	6.71	14.36
VoG	*M*	40.91	39.09	63.03	60.00
(N = 22)	*SD*	4.60	7.60	11.77	13.80
Control	*M*	32.19	35.43	58.73	61.9
(N = 21)	*SD*	6.98	8.03	10.25	18.64

Prediction 1

The first prediction associated with this hypothesis focused on examining the effectiveness of including real-time explicit feedback within a game-based training environment by itself. It is hypothesized that individuals receiving real-time feedback will score better on all performance metrics when compared to the baseline where individuals had to rely on implicit information from the environment to gauge performance. This was carried out by examining performance outcomes within the TC3Sim training scenario, and grouping individuals in the analysis as whether they received or didn’t receive explicit feedback during gameplay. To test this, a Univariate ANCOVA was run comparing the two groups. For this analysis VGE was defined as a CV. Results showed the inclusion of explicit feedback, regardless of the source, to have a significant main effect on training scenario performance, (F (1, 129) = 11.749, p = .001, η_p^2 = .05, power = 0.925; see Figure 13 for a visual representation), with VGE reporting as a significant CV, (F (1, 122) = 5.312, p < .025, η_p^2 = .040, power = 0.628).

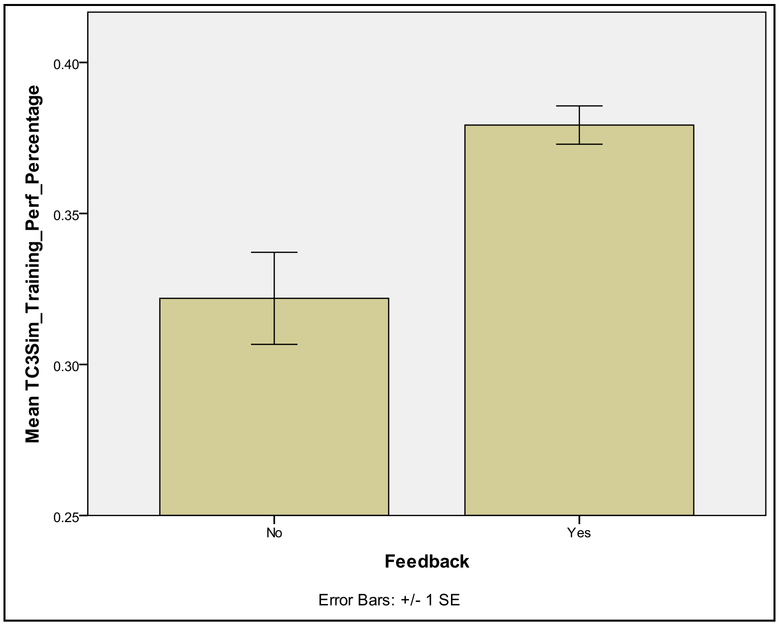


Figure 13. TC3Sim Training Scenario Performance With/With-Out Explicit Feedback

Next, a Univariate ANCOVA was run incorporating the comparisons of training scenario outcomes across all treatments. See Figure 14 for a graphical representation of training scenario performance results. This test identifies if there are reliable differences in the training scenario performance metric for all experimental conditions. The analysis shows a significant main effect of feedback on performance outcomes for the TC3Sim training scenario (F (5, 122) = 3.735, $p <$ .01, η_p^2 = .133, power = 0.925), with video game experience being identified as a significant CV (F (1, 122) = 4.791, $p <$ .025, η_p^2 = .038, power = 1.000). This relationship shows those scoring higher on video game experience produced higher performance during training scenario interaction (Pearson r = .218).

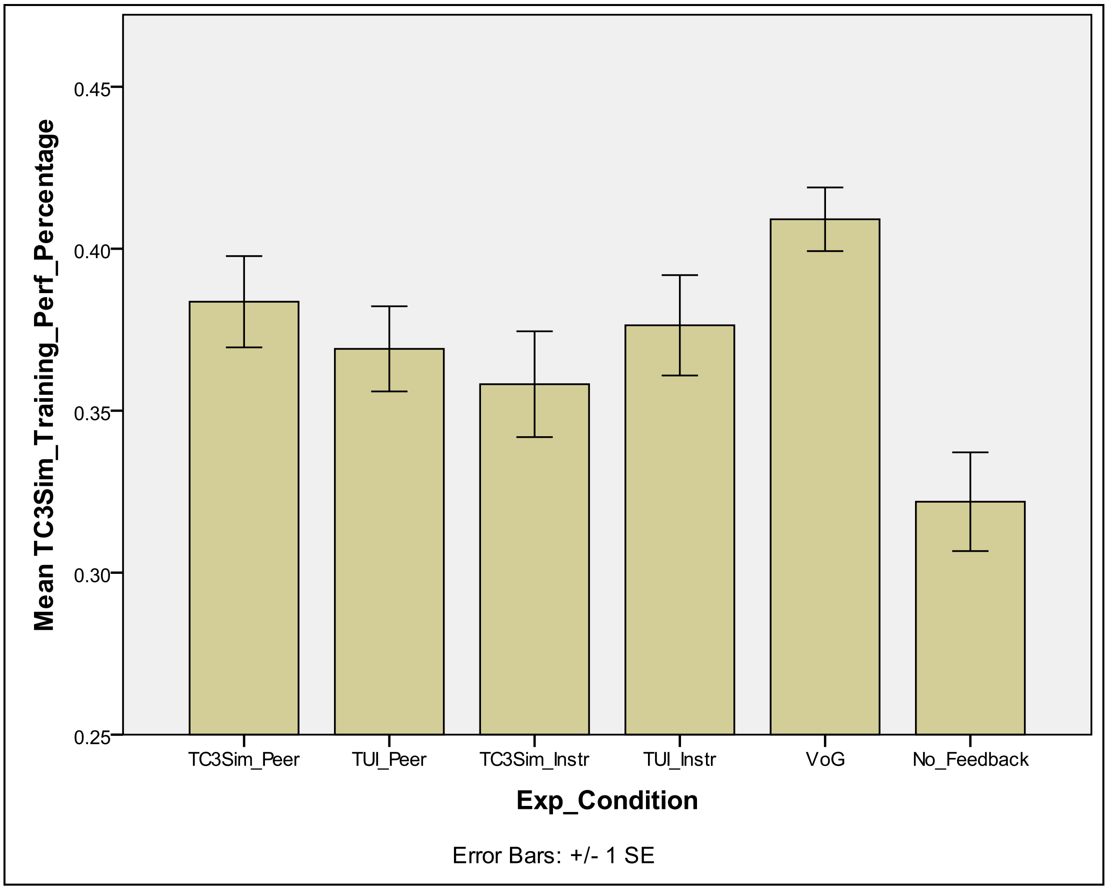


Figure 14. TC3Sim Training Scenario Performance

Based on this finding, planned comparisons between each of the conditions were conducted, with results being summarized in Table 5. The analyses show the mere presence of explicit feedback during game play significantly improved scenario performance outcomes when compared to the baseline version of the game that is currently being used in training houses across the country. Outcomes show all conditions, minus the TC3Sim Instructor, were found to significantly outperform the control. It is interesting to note that the VoG condition, which had no associated EPA, produced the highest overall scores for the training scenario.

Table 5. Planned Comparisons examining each experimental condition versus the control with no feedback.

Condition	*t*	*p*
No Feedback vs. TC3Sim_peer	(41) = -2.987	<.01
No Feedback vs. TUI_peer	(41) = -2.352	<.025
No Feedback vs. TUI_instr	(41) = -2.504	<.025
No Feedback vs. VoG	(41) = -4.854	<.001

Because feedback was provided solely in the training scenario, prediction 1 analyses are focused on this performance metric alone. The effect the IVs have on associated learning gains will be addressed in analyses described below. In assessing the statistical approaches applied to test prediction 1, it is clear that the inclusion of explicit feedback during a game-based training event significantly improved in-game performance metrics across all associated conditions.

Prediction 2

Next, analyses were conducted examining the influence an EPA has on performance scores from the game and knowledge assessments. Prediction 2 states that conditions where participants interacted with an EPA in the game environment would produce significantly better performance scores on both game-based metrics and the associated knowledge post-test. A fundamental component to this prediction is based around SCT and tests if the mere presence of an EPA produces improved performance when compared against conditions with no feedback or with feedback that does not have a grounded source (i.e., feedback delivered from no visible entity in the environment). It was hypothesized that the presence of an EPA will result in better overall learning due to the inclusion of a social element that is inherent to learning new skills, as highlighted in research covering SCT and the Persona Effect. Descriptive statistics for all performance variables as they relate to the EPA Presence breakdown can be seen in Table 6.

Table 6. Descriptive Statistics Comparing Conditions with/without an EPA

EPA, VoG, or No Feedback		*TC3Sim Scenario %*		*Knowledge*	*Knowledge*
		Training	*Capstone*	*Pre-Test*	*Post-Test*
EPA	*M*	37.18	40.00	64.17	69.02
(N = 88)	*SD*	6.89	8.65	11.04	12.95
VoG	*M*	40.91	39.09	63.03	60.00
(N = 22)	*SD*	4.61	7.60	11.77	13.80
No Feedback	*M*	32.19	35.43	58.73	61.90
(N = 21)	*SD*	6.98	8.03	10.25	18.64

The first test performed was to examine the effect an EPA has on performance within the training scenario alone. This differentiates the analysis from above, in that it takes into account the VoG condition to determine if performance between these two design treatments is significantly different. A Univariate ANCOVA was run across the three defined groups, with VGE defined as the CV. The test output shows the conditions relating to interaction with an EPA, VoG, or No Feedback to produce significant differences in performance outcomes, (F (2, 129) = 8.28, $p < .001$, $\eta_p^2 = .117$, power = 0.958), along with VGE reporting as a significant CV, (F (1, 129) = 4.356, $p < .05$, $\eta_p^2 = .034$, power = 0.544). To examine further post-hoc analysis was performed using the Bonferroni test, with results showing both the EPA and VoG groups to score significantly higher than the No Feedback condition (see Table 7). However, no significant difference was found between the EPA and VoG groupings. See Figure 15 for a visual representation of the estimated marginal means of the TC3Sim training scenario performance scores as a result of the ANCOVA.

Table 7. Post-Hoc Analysis of Training Scenario Performance Across EPA Treatments

EPA, VoG, or No Feedback	*TC3Sim Scenario %* Mean	*TC3Sim Scenario %* Standard Error	*Significance*
EPA vs.	37.2	.007	$p = .01$
No Feedback	32.4	.014	
VoG vs.	40.6	.014	$p < .001$
No Feedback	32.4	.014	

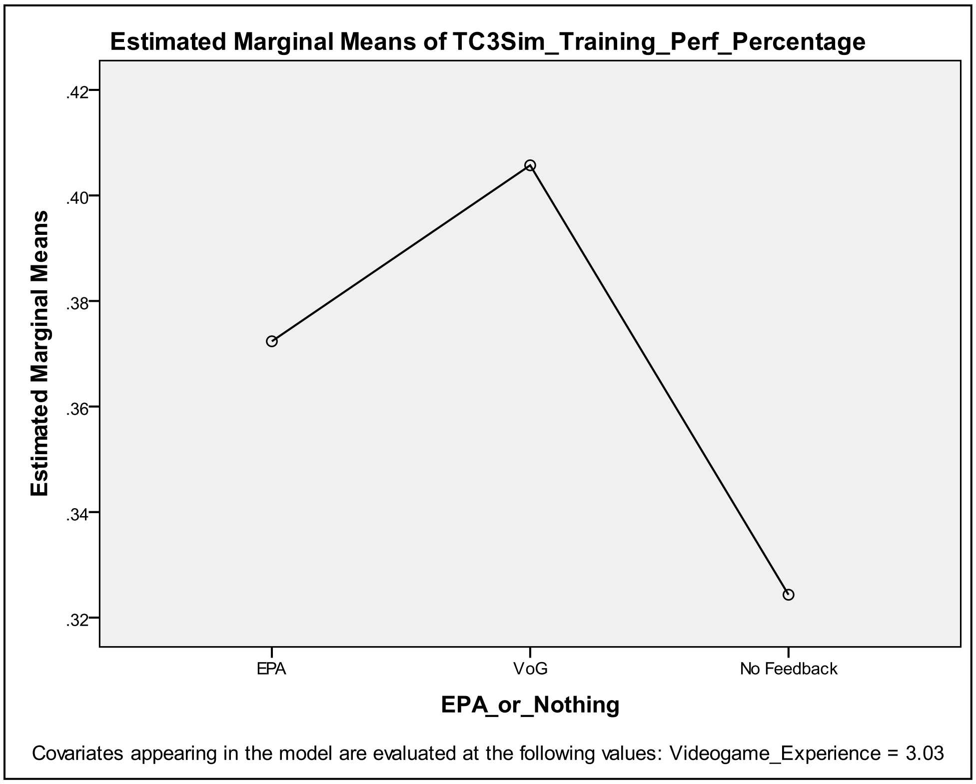


Figure 15. Estimated Marginal Means of TC3Sim Training Scenario Outcomes

Next, analyses were conducted to examine participant's subsequent performance within a capstone scenario directly following training that incorporated no explicit feedback for all conditions. Performance outcomes from this scenario are used to gauge if feedback present in a training scenario will lead to better overall performance on a similar task and for measuring learning gains as they relate to scenario execution. This analysis also assists in determining if the inclusion of an EPA produces larger performance outcomes on transfer assessments. A graphical representation of training performance compared to capstone performance can be seen in Figure 16.

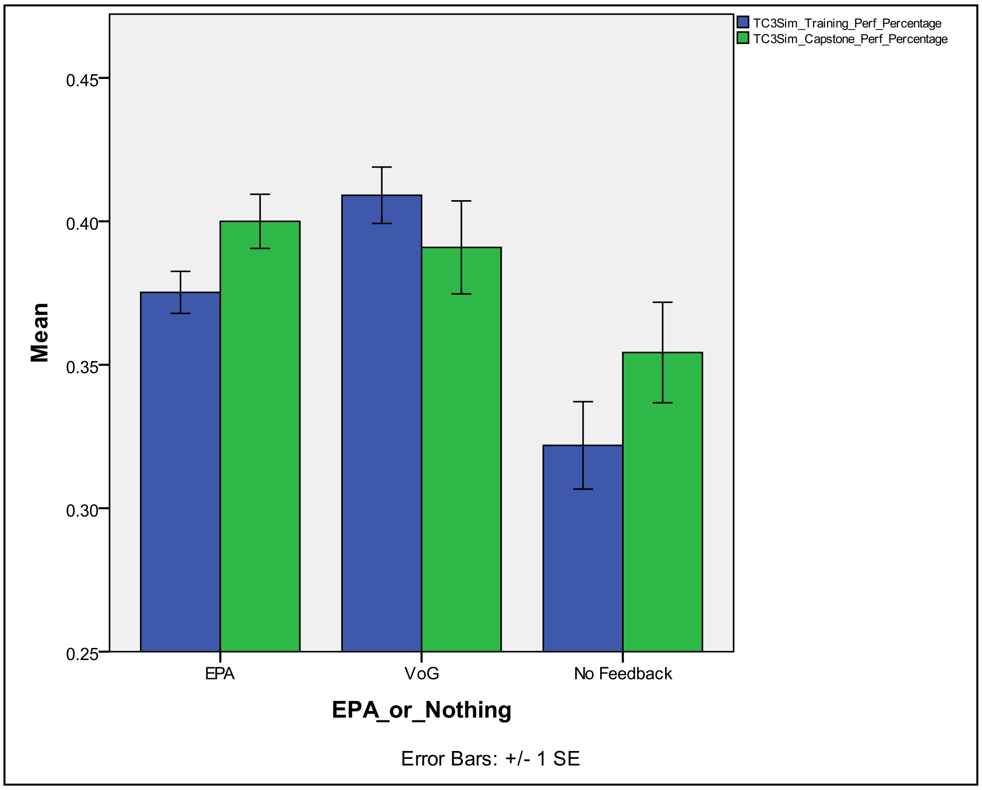


Figure 16. Comparison of TC3Sim Game Performance for EPA, VoG, and No Feedback Groups

A mixed between-within subjects ANOVA was performed examining differences in performance gains between the two game scenarios and to determine if the feedback source had an influence on the associated outcomes. Results show no significant within-subject interaction between scenario and experimental condition (*F* (1, 125) = 2.572, *p* = .080, η_p^2 = .040, power = 0.505). However, the mixed ANOVA revealed a significant between subjects main effect across conditions in terms of TC3Sim performance as deemed by the scores across the two scenarios (F (2, 128) = 4.520, $p < .025$, $\eta_p^2 = .066$, power = 0.762), which shows that regardless of the assigned conditions participants reliably produced different performance scores across the two scenarios. Interestingly, when examining the visual representation of performance across the two scenarios, it was recognized that the VoG condition was the only treatment to produce lower performance scores on the capstone when compared to the training scenario.

Next, a Univariate ANCOVA was conducted to test the finding found above and to identify if associated EPA capstone performance was significantly different when compared against outcomes from the VoG and No Feedback conditions, with a participants training scenario score being defined as the CV. Results show the source treatment to have no significant main effect on game performance within the capstone scenario (F (2, 123) = 1.232, $p = .295$, $\eta_p^2 = .020$, power = 0.264), with a participants performance on the training scenario being a significant CV, (F (1,123) = 19.571, $p < .001$, $\eta_p^2 = .137$, power = 0.992). Regardless of the condition, an individual's score on the TC3Sim training scenario was found to strongly predict their performance on the subsequent assessment scenario (Pearson's $r = .393$, $p < .001$). A visual representation of the resulting estimated marginal means of capstone performance as a result of the ANCOVA can be seen in Figure 17.

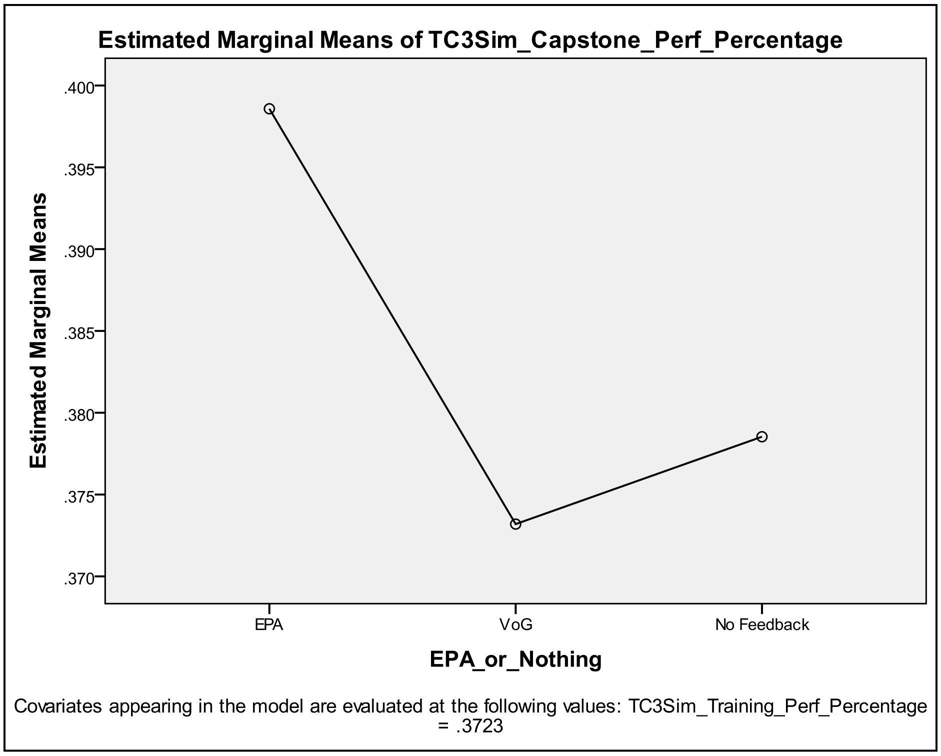


Figure 17. Estimated Marignal Means of the Game Capstone Scneario

It is also interesting to note that in the No Feedback condition participants improved their performance in the capstone scenario despite not having explicit feedback provided to them during training. Yet, their performance in the capstone was also still lower than all other conditions. Thoughts for why the VoG condition produced a negative learning gain will be addressed in the discussion.

Following examination of game-based performance metrics, analyses were performed on outcomes from the two knowledge tests administered at the beginning and end of the experimental session. It is hypothesized that individuals who interacted with game conditions involving explicit feedback from an EPA would gain a better conceptual understanding of the tasks, resulting in larger test gains on associated knowledge tests. A mixed between/within subjects ANOVA was run looking at the differences in performance across the pre- and post-test knowledge scores to identify learning gains and determine if explicit feedback delivered by an EPA impacted overall outcomes. A visual graphic of these performance metrics can be seen in Figure 18. In examining the statistical outputs, results show no significant within subject interaction between Pre-/Post-Test Administration and the source conditions, (F (2, 128) = 2.413, $p < .094$, $\eta_p^2 = .036$, power = 0.479). However, a significant between subject main effect for Experimental Condition was identified (F (2, 128) = 4.520, $p < .025$, $\eta_p^2 = .066$, power = 0.7626) based on a transformed variable computed by averaging an individual's two test scores.

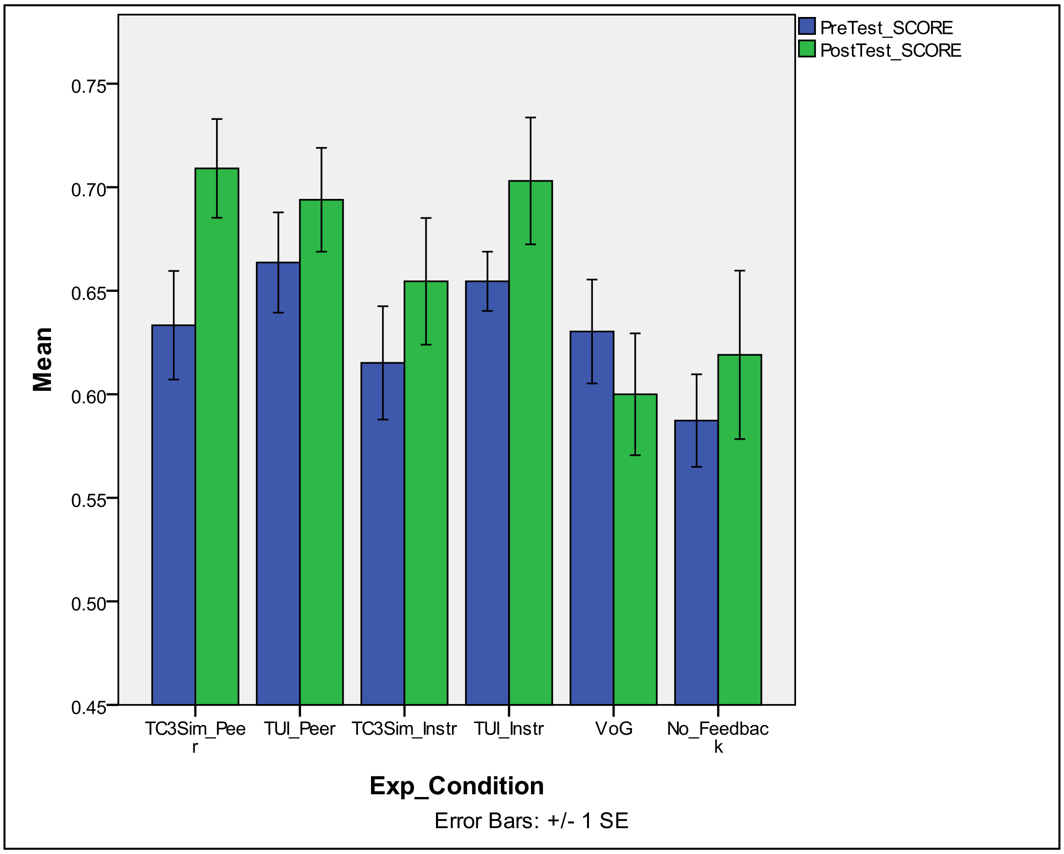


Figure 18. Pre-/Post-Test Performance Outcomes Across Conditions

Because of the identified significant between subjects main effect, post hoc analysis was conducted to identify the conditions to produce reliable differences for knowledge learning gains. To account for performance scored on the administered pre-test, a Univariate ANCOVA was performed to look at the effect source conditions have on post-test outcomes, with the pre-test score being defined as a CV. Results show the source condition to have a significant main effect on the knowledge post-test scores (F (2, 127) = 4.028 , $p < .025$, $\eta_p^2 = .060$, power = 0.710), with an individual's pre-test score showing as a significant CV (F (1, 127) = 12.975, $p < .001$, $\eta_p^2 = .093$, power = 0.947). As found above in game performance, an individual's score on the knowledge pre-test was found to strongly predict their performance on the subsequent post-test, regardless of the condition (Pearson's $r = .321$, $p < .001$). A visual representation of the resulting estimated marginal means of post-test performance as a result of the ANCOVA can be seen in Figure 19.

To examine further, post-hoc analysis was performed with the Bonferroni test, resulting in an identified significant difference on post-test performance between those interacting with an EPA ($M = 68.86$, $SE = .014$) and those in the VoG condition ($M = 60.00$, $SE = .029$; $p = .026$). While the EPA conditions outperformed the No Feedback by more than five percentage points, there was no significant difference found as a result of the ANCOVA.

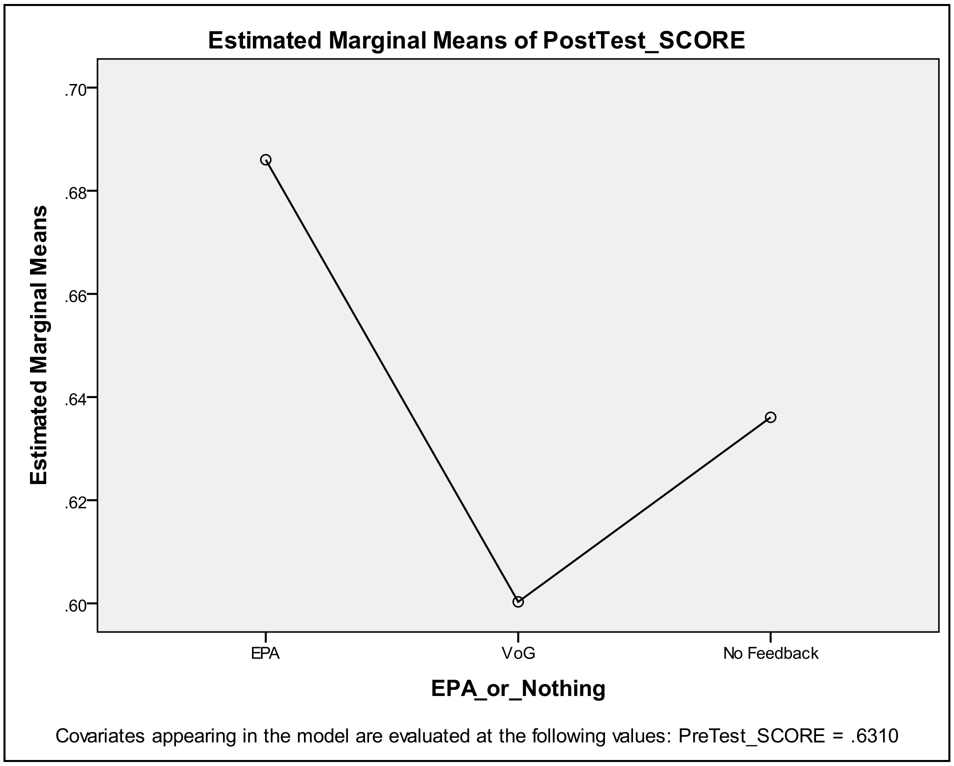


Figure 19. Estimated Marignal Means of the Knowledge Post-Test

Interestingly, in reviewing the visual depiction of the data from Figure 18, all conditions are shown to produce increases in knowledge as deemed by the pre-/post-test comparisons, except for the VoG condition, which is the only treatment to show a decrease in performance. However, it is important to note that participants in the No Feedback condition performed significantly lower on the post-test than all other conditions except for VoG, which shows the mean for this treatment to be lowest across all groups.

It is important to remember that this analysis takes into consideration all participants interacting with the EPA as a single group. In examining the breakdown of specific EPA conditions to produce reliable differences, the EPA vs. No Feedback comparison showed no significant differences as highlighted in the Prediction 1 results, yet reliable differences in post-test performance were identified between three of the four EPA conditions when compared against VoG. This shows that although the VoG condition produced the highest performance marks during the game-based training scenario, all subsequent performance metrics collected, including both the capstone game scenario and knowledge post-test, were significantly lower than conditions where an EPA was present. This relationship will be dissected further in the next chapter.

Prediction 3

Prediction 3 is associated with the location of the EPA during gameplay (TUI vs. Game-Embedded) and if there was an effect on resulting performance outcomes. Because all of the EPA conditions incorporate explicit feedback, it is predicted that there will be no significant differences in outcomes as a result of where the EPA was positioned. As can be seen in Table 4 on page 116, the descriptive statistics across each TC3Sim-tutor and TUI-tutor condition show little variance in performance for the TC3Sim training scenario.

As this is the only aspect of the experimental procedure where a tutor was present, this analysis focused solely on training scenario outcomes to determine if performance was affected by a tutor being located in the TUI while the game was displayed in a windowed mode. A Univariate ANCOVA was performed based around the TUI-embedded and TC3Sim-embedded EPA groupings, with VGE defined as the CV. As predicted the results show no significant differences in training performance when comparing a tutor in the TUI (*M* = 37.3. *SE* = .011) versus being embedded in the game environment (*M* = 37.1, *SE* = .011; *F* (1, 86) = .023, *p* = .879, η_p^2 = .000, power = 0.053). As seen in the groups associated means, there is minimal variance in performance outcomes as a result of where the EPA was located during game interaction. Next, analyses will be presented that investigate the effect of the IVs on an individual's reported level of workload and cognitive demand.

Hypothesis 2

Due to experimental conditions involving variations in the game-tutor interface design, Hypothesis 2 focuses on analysis linked to an individual's MD and associated WL during game interaction, and is based on research surrounding MRT and CLT (Oviatt, 2006). Analysis linked to Hypothesis 2 is based on self-reported WL and MD metrics collected from the NASA-TLX directly following the TC3Sim training scenario. Due to time limitations with the subject pool, we were unable to re-administer the NASA-TLX following the capstone scenario to determine if further exposure to the game reduces the perceived amount of effort to perform effectively. As a result, statistical tests were applied to examine the relationships between IVs and their impact on WL and MD within only one of the two scenarios. For a list of descriptive statistics on associated WL and MD metrics across each individual condition, see Table 8.

Table 8. Experimental Workload and Mental Demand Metrics Across Conditions

Feedback Modality		*NASA-TLX Results*	
Condition		*Workload*	*Mental Demand*
TC3Sim-Peer	*M*	56.53	79.00
(N = 22)	*SD*	6.41	9.52
TC3Sim-Instr	*M*	57.02	82.64
(N = 22)	*SD*	10.71	15.70
TUI-Peer	*M*	52.78	86.68
(N = 22)	*SD*	13.49	9.27
TUI_Instr	*M*	57.74	82.41
(N = 22)	*SD*	11.08	22.18
VoG	*M*	55.65	73.86
(N = 22)	*SD*	9.53	13.89

Feedback Modality		*NASA-TLX Results*	
Condition		*Workload*	*Mental Demand*
No Feedback	*M*	52.37	85.33
(N = 21)	*SD*	13.24	15.89

Prediction 1 and 2

Results for Prediction 1 and Prediction 2 are presented together because they are relatively defined as being inverse of each other. Based on components found within Wicken's (2002) MRT, two separate predictions were created that account for different applications of the theory. Prediction 1 is based around the implementation of two separate interfaces to enable GIFT's TUI to house an EPA for explicit feedback delivery while also displaying the game in a windowed-mode. This approach is being compared against conditions with the EPA embedded in the game environment as a NPC, which takes significantly more time to implement.

Because one of the conditions has the EPA situated in a separate interface, it is predicted that WL and MD will be reported as significantly higher in the TUI conditions when compared to individuals interacting with the tutor embedded within TC3Sim. This is due to the individual having to maintain attention on two separate visual fields, requiring more visual resources to maintain orientation of what is happening. This is believed to make the perceived difficulty of the task higher, resulting in higher WL and MD scores. For descriptive statistics and a visual representation of the data, see Table 9 and Figure 20. Based on this figure, it is interesting to note the vast difference in reported MD when compared to the overall calculated WL score.

Table 9. Workload and Mental Demand Metrics Comparing Feedback Source Modalities

Feedback Modality		*NASA-TLX Results*	
Condition		*Workload*	*Mental Demand*
TC3Sim_Embedded	*M*	56.77	80.82
(N = 44)	*SD*	8.72	12.97
TUI_Embedded	*M*	55.27	84.54
(N = 44)	*SD*	12.45	18.85
VoG	*M*	55.65	73.86
(N = 22)	*SD*	9.53	13.89

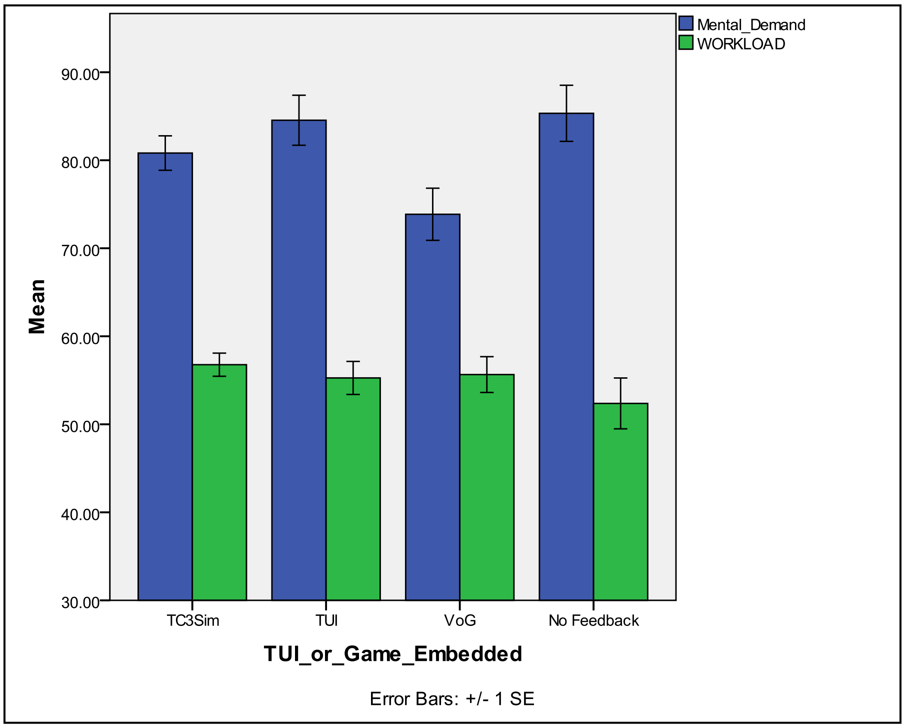


Figure 20. Workload and Mental Demand Metrics Across Source Modalities

In comparison to Prediction 1, Prediction 2 is based on Wickens (2002) description of ambient vision. From this perspective information can be perceived through an individual’s peripheral vision field, allowing that user to maintain a sense of orientation on stimuli in their peripheries while maintaining majority of attentional resources on a primary task. This resource enables an individual to monitor dueling tasks efficiently if one of the tasks only requires ambient vision to process necessary information. While the EPA in the TUI displays movement at times, its location in the browser is relatively static. In addition, the inclusion of an EPA character in an already busy game environment may force a learner to apply more visual resources to maintain location and presence, resulting in higher cognitive load. Because of this, it is predicted that WL and MD will report significantly higher in the TC3Sim Embedded tutor condition when compared to the TUI Embedded tutor due to a user being able to apply ambient vision to reduce the load on visual resources required to effectively perform in the training scenario. These predictions are based around the defined EPA and VoG conditions only, and will be used to determine interface design approaches as they relate to source modalities for explicit feedback delivery.

To establish if there were reliable differences in reported WL and MD scores across treatments, two separate Univariate ANOVAs were performed on each of the cognitive load metrics. Results show the overall WL metric (i.e., metric computed from all six dimensions of NASA-TLX) to reveal no significant differences between conditions (F (2, 107) = .235, p = .791, η_p^2 = .004, power = 0.086), while the MD metric showed reliable differences as a result of whether a participant interacted with the TUI-Embedded tutor, the TC3Sim-Embedded tutor, or the VoG condition with no defined EPA (F (2, 107) = 3.373, p < .05, η_p^2 = .059, power = 0.625).

To examine further, planned comparisons were performed to determine the specific treatments contributing to this statistical finding. Outcomes from these tests showed both the TUI-Embedded tutor (M = 84.54, SD = 18.85) conditions and TC3Sim-Embedded tutor (M = 80.82, SD = 12.97) conditions to report significantly higher MD scores when compared to the

VoG condition (*M* = 73.86, *SD* = 13.89; see Table 10), while no reliable differences were found between the varying EPA source modalities. Outcomes from this analysis signify that the inclusion of an additional interface during gameplay did not result in higher MD scores when compared against those interacting with a tutor in the environment, rejecting both predictions, and supporting the TUI as a viable tool for relaying information in game-based learning events. Although the analysis did not support the associated predictions, the results show the VoG condition to report significantly lower scores on MD when compared to all EPA related treatments. Discussions on this relationship will be explored in the next chapter.

Table 10. Planned Comparisons Results for Mental Demand Scores Across EPA Condtions

Condition	*t*	*p*
TC3-Embedded vs. VoG	(64) = 2.006	<.05
TUI-Embedded vs. VoG	(64) = 2.354	<.25

Prediction 3

With a baseline condition not providing explicit feedback during the TC3Sim training scenario, Prediction 3 is focused on examining if those relying solely on implicit information from the game to gauge performance would report significantly higher WL and MD scores when compared to those receiving feedback based on actions taken. Two analyses were conducted to test this hypothesis. The first was a run of two Univariate ANOVAs looking at both WL and MD against two defined groups of Feedback and No Feedback (See Figure 21 for a visual representation of the two groups). Results for both ANOVAs show no significant differences between the two groups for both metrics (MD: F (1, 129) = 1.364, p = .245, η_p^2 = .010, power = 0.213; WL: F (1, 129) = 1.886, p = .172, η_p^2 = .014, power = 0.276).

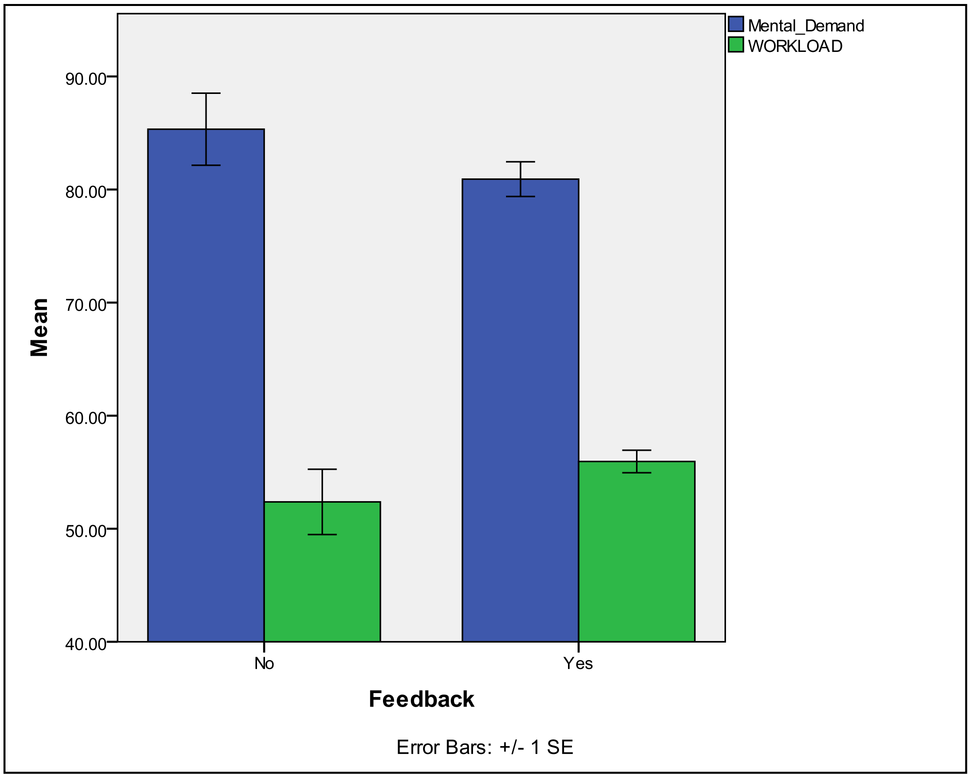


Figure 21. Workload and Mental Demand Scores Based on Presence of Feedback

The next set of analyses looked at each individual condition against those receiving no feedback through defined simple contrasts within a Univariate ANOVA for both WL and MD metrics. This allows a simple comparison of each condition against the control in a single run. As seen in all results for Hypothesis 2, the metric of WL showed no significant differences between the individual conditions (*F* (5, 125) = .910, *p* = .477, η_p^2 = .035, power = 0.317). In addition, results from the ANOVA on MD show no significant differences between conditions (*F* (5, 125) = 1.939, *p* = .092, η_p^2 = .072, power = 0.639), yet the contrasts showed the VoG (*M* = 73.86, *SE* = 3.33) and No Feedback (*M* = 85.33, *SE* = 3.41) conditions to report as significantly different. A planned comparison was run to examine this relationship, with results showing the No Feedback condition to score significantly higher on MD when compared to VoG (*t* (41) = -2.639, *p* <.025). This outcome will be examined further in the discussion.

Prediction 4

A second IV incorporated in this study was a defined EPA profile that was presented to each participant prior to interaction with the TC3Sim training scenario (see APPENDIX L: EPA PROFILE BACKGROUNDS/BIOS). In terms of Prediction 4, WL and MD scores are not expected to be affected by the EPA profile, as it does not affect the interfacing components that present feedback information. To test this prediction, a One-Way ANOVA was performed to determine if the EPA Profile impacted an individual's reported WL and MD. Results from this analysis show no significant differences between the two groups, signifying that the profile treatments had no resulting effect on how someone perceived the demand and cognitive load required to perform. To see a visual of WL and MD across the EPA Profile groups, see Figure 22. The EPA Profile IV will be further explored in Hypothesis 3.

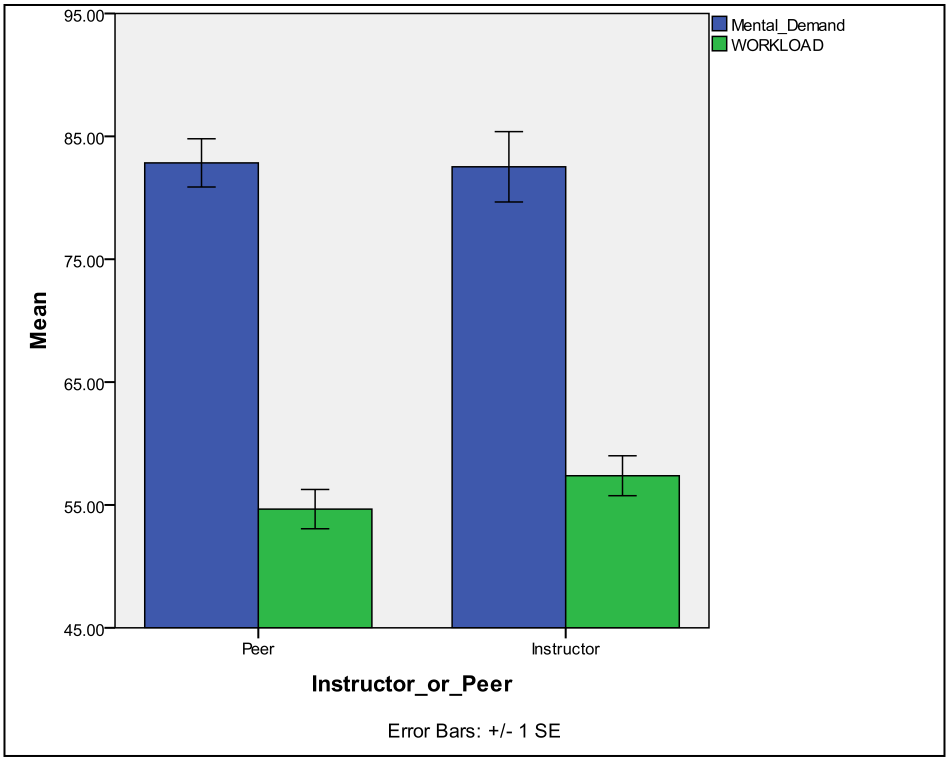


Figure 22. Workload and Mental Demand Across EPA Profile Conditions

Hypothesis 3

As deemed from analyses linked to Hypothesis 1, the inclusion of explicit feedback is shown to significantly impact an individual's performance both within a game-based training event and during subsequent domain knowledge tests. With feedback reliably shown to affect performance outcomes, Hypothesis 3 is interested in examining the IV 'EPA Profile' and the associated subjects' perceptions of the EPA during the TC3Sim training scenario, as collected from the API (Ryu & Baylor, 2005). Though performance is shown to increase, it is important to understand how individuals interact with these types of agents, and if they perceive them to add value to the experience.

Based on research influenced by SCT and the agent persona effect (Veletsianos, 2010), two EPA profiles were created to test if how an agent is defined will impact how a user perceives its usefulness. The background profiles were constructed for an EPA to act as a ‘Peer’ team member or as an ‘Instructor’ with an accomplished career (see APPENDIX L: EPA PROFILE BACKGROUNDS/BIOS). Depending on the assigned treatment, the EPA profiles were displayed to each participant just prior to entering the training scenario. This introduces the tutor to the learner, and provides a context for the tutor’s intended role. The VoG condition did not receive any EPA introduction, as the feedback modality was designed so that information was not grounded to any type of source. Because each condition with explicit feedback received the same reflective prompts during interaction, the API will determine the effect character profiles have on the stereotyping of source credibility, as governed by interaction between the two profiles. It is hypothesized that the character profile linked to an EPA will significantly affect scores across the dimensions of the API (see Table 11 for a list of the descriptive statistics associated with the API across each experimental condition). In terms of associated scores and their interpreted meaning, the API is scored on a 5-point Likert scale (i.e., 1 = Strongly Disagree to 5 = Strongly Agree, with 3 = Neither Agree or Disagree). It is important to note that because the No Feedback condition had no designated feedback agent, the API was not administered to these participants.

Table 11. Agent Persona Instrument (API) Descriptive Statistics Across Conditions

Feedback Modality Condition		*Agent Persona Instrument (API) Dimensions* *Facilitating Learning*	*Credibility*	*Human-Likeness*	*Engaging*
TC3Sim-Peer	*M*	3.36	3.86	3.28	3.24
(N = 21)	*SD*	.544	.666	.666	.404
TUI-Peer	*M*	3.44	3.45	3.44	3.45
(N = 20)	*SD*	.522	.605	.613	.636
TC3Sim-Instr	*M*	3.46	3.36	3.65	3.42
(N = 22)	*SD*	.522	.551	.427	.470
TUI_Instr	*M*	3.65	3.97	3.55	3.41
(N = 21)	*SD*	.495	.587	.565	.522
VoG	*M*	3.51	3.51	3.62	3.55
(N = 22)	*SD*	.486	.661	.559	.494

Prediction 1

The API is composed of 25-items that rate components of interaction with an EPA across four dimensions: Facilitation to Learning, Credibility, Human-Likeness, and Engagement. Prediction 1 is focused on the dimensions of Human-Likeness and Engagement to examine Baylor & Kim's (2005) assertion that EPAs can be designed to facilitate different instructional roles. With findings from their research showing individuals to perceive 'mentor' based agents as more human-like, while agents with 'expertise' were more facilitative to learning, it is predicted that individuals interacting with the defined 'Instructor' EPA will produce significantly lower scores on the two dimensions of 'Engagement' and 'Facilitation to Learning'. This is also influenced by agent stereotype research that shows individuals to automatically create impressions of an EPA based on their assigned role and appearance (Veletsianos, 2010). For descriptive statistics of the API dimensions across the three groups of Peer, Instructor and VoG, see Table 12.

Table 12. Agent Persona Instrument (API) Descriptive Statistics Across EPA Profile Groups

Peer, Instructor, or VoG Treatments		*Agent Persona Instrument (API) Dimensions*			
		Facilitating Learning	*Credibility*	*Human-Likeness*	*Engaging*
Peer EPA	*M*	3.40	3.66	3.35	3.35
(N = 44)	*SD*	.529	.662	.638	.537
Instructor EPA	*M*	3.56	3.67	3.60	3.41
(N = 44)	*SD*	.510	.641	.497	.491
VoG	*M*	3.51	3.51	3.62	3.55
(N = 22)	*SD*	.486	.661	.559	.494

As can be seen from the visual representation in Figure 23, there is minimal variation between groups on both of the dimensions of interest. While the 'Peer' EPA scores are lower on both scales, the variance is not enough to show reliable differences when compared to the other groups. This is backed up by results from a Multivariate ANOVA (MANOVA), with outputs from Pillai's trace test showing a non-significant outcome ($V = .064$, $F\ (2, 214) = 1.771$, $p = .136$, $\eta_p^2 = .032$, power = 0.535). Univariate ANOVAs were also run for each dimension, with results again showing no reliable differences (Human-Likeness: $F\ (2, 107) = 2.492$, $p = .088$, $\eta_p^2 = .044$, power = 0.491; Engagement: $F\ (2, 107) = 1.179$, $p = .312$, $\eta_p^2 = .022$, power = 0.254). As deemed from these statistical analyses, the EPA profile of 'Peer' and 'Instructor' produced no differences in individuals' perceptions of human-likeness and engagement when judging the tutor agent. It is also interesting to note that although participants in the VoG condition did not interact directly with an EPA, they still gave relatively high marks on both of these scales. The next prediction analyzed is focused on the dimensions of 'Facilitating to Learning' and 'Credibility'.

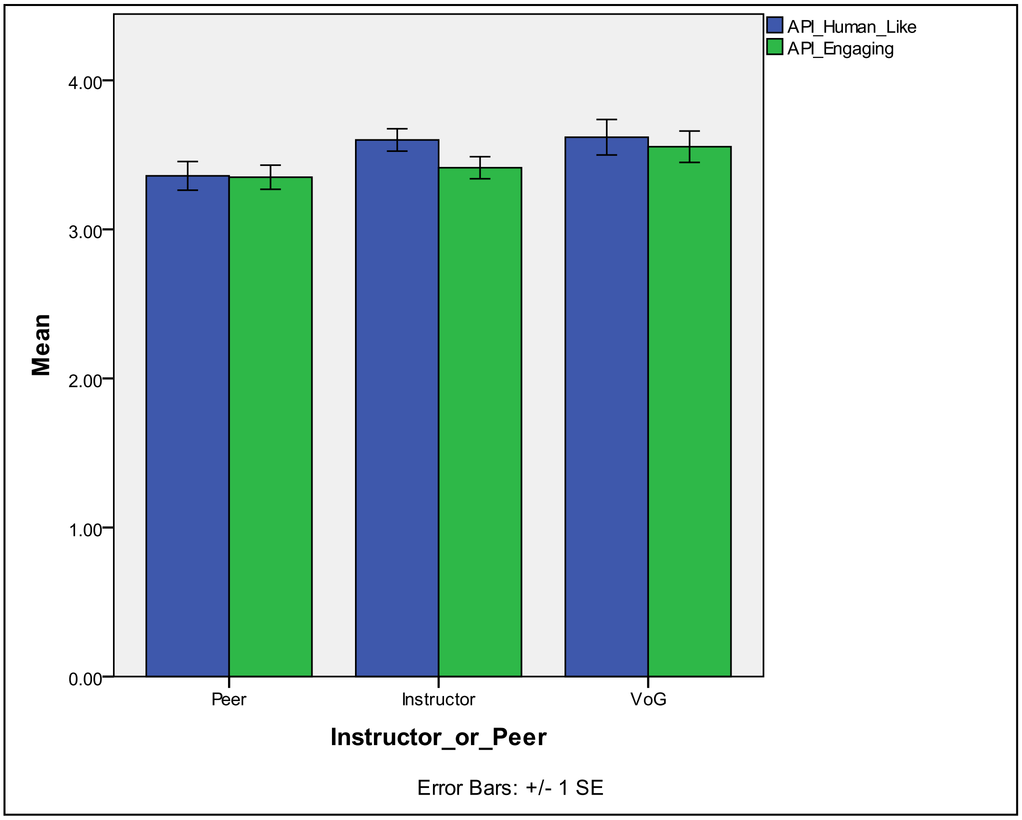


Figure 23. API Scores for Human-Likeness and Engagment across EPA Profile Groups

Prediction 2

With the previous analysis examining differences across the API dimensions of 'Human-Likeness' and 'Engagement', Prediction 2 is interested in how the EPA Profile IV affects a subject's perceived rating across 'Facilitation to Learning' and 'Credibility' (see Figure 24 for graphical representation). In contrast to Prediction 1, it is believed that those interacting with the 'Instructor' EPA would report significantly higher marks on these dimensions, as the defined instructor is credited with having expertise in the TC3 domain. This prediction is supported by stereotype research conducted by Veletsianos (2010) and outcomes from the Baylor & Kim

(2005) study that showed a defined agent's role (e.g., expert and mentor) to impact a learner's perception of their usefulness in a training environment.

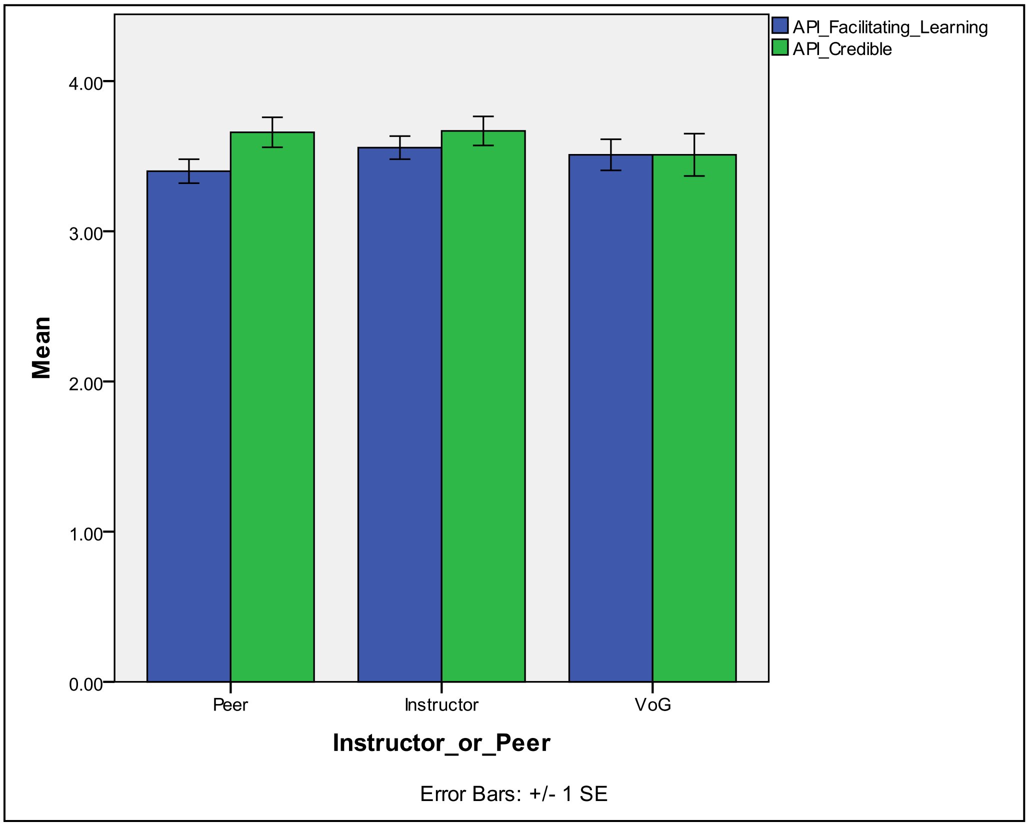


Figure 24. API Scores for Facilitation to Learning and Credibility across EPA Profile Groups

To test this prediction, a MANOVA was run looking at the API dimensions of 'Facilitation to Learning' and 'Credibility' across the three EPA Profile groups. Similar to prediction 1, the visual representation of the API data across the three groups for Prediction 2 shows minimal variation, signifying relatively equal scores on the two dimensions across the three groups. Results from the MANOVA support this claim, with outputs from Pillai's trace test showing a non-significant effect of EPA profile on the recorded scores of 'Facilitation to Learning' and 'Credibility' ($V = .083$, $F\ (2, 214) = 2.311$, $p = .059$, $\eta_p^2 = .041$, power = 0.666).

With the *p*-value of .059 approaching significance, Univariate ANOVAs were conducted on the two dimensions by themselves to observe if there are any reliable difference between groups. Results from these tests show the dimensions ‘Facilitation to Learning’ and ‘Credibility’ to not be significantly different between groups (Facilitation to Learning: F (2, 107) = 1.059, p = .350, η_p^2 = .019, power = 0.231; Credibility: F (2, 107) = .494, p = .611, η_p^2 = .009, power = 0.129). While the EPA Profile IV has been shown to produce no significant differences across any of the API dimensions, the next two predictions associated with Hypothesis 3 are focused on the perceived effect of EPA location (i.e., source modality) on the four API dimensions.

Prediction 3

The next analyses examine to what effect the location of the EPA (e.g., Internal TC3Sim-Embedded EPA or External TUI-Embedded EPA) has on reported scores across the API dimensions (See Table 13). In terms of Prediction 3, it is expected that the TC3Sim Embedded EPA will produce significantly higher scores on the dimensions of Human-Likeness and Engagement. Because the EPA is directly interacting with the virtual environment the scenario is taking place within, the agent is predicted to be perceived as more engaging and life-like as a result of seeing it move naturally with other entities in the game. This is in comparison to the TUI-Embedded condition, where the EPA is present in a separate internet-browser window. Engaging with this character requires attention to be taken from the game, lending to the prediction that the Engagement dimension will report significantly lower in the TUI treatment. In addition, though the appearance of the TUI-Embedded EPA is visually realistic, the agent’s movements are relatively static, which is the basis for predicting Human-Likeness will score significantly higher in the TC3Sim-Embedded conditions.

Table 13. Agent Persona Instrument (API) Descriptive Statistics Across EPA Source Modalities

EPA Source Modality Conditions		*Agent Persona Instrument (API) Dimensions*			
		Facilitating Learning	*Credibility*	*Human-Likeness*	*Engaging*
TC3Sim-Embedded	*M*	3.41	3.61	3.46	3.33
(N = 44)	*SD*	.527	.655	.583	.442
TUI-Embedded	*M*	3.55	3.71	3.50	3.43
(N = 44)	*SD*	.515	.645	.586	.575
VoG	*M*	3.51	3.51	3.62	3.55
(N = 22)	*SD*	.486	.661	.559	.494

To test Prediction 3, a MANOVA was performed examining both dimensions of Human-Likeness and Engagement together against the Feedback Source Modality treatment groups. Results show when using Pillai's trace that there was no significant effect of EPA Source Modality on the reported scores of the two API dimensions ($V = .029$, $F\ (2, 214) = .778$, $p = .541$, $\eta_p^2 = .014$, power = 0.247). To assess further, Univariate ANOVAs were conducted with results showing the EPA conditions to not have a reliable effect on the Human-Likeness and Engagement scores when analyzed by themselves (Human-Likeness: $F\ (2, 107) = .537$, $p = .586$, $\eta_p^2 = .010$, power = 0.137; Engagement: $F\ (2, 107) = 1.437$, $p = .242$, $\eta_p^2 = .026$, power = 0.302). Outcomes from these statistical tests show no support for Prediction 3, with results showing the EPA Source Modality to reliably produce similar responses on the Human-Likeness and Engagement dimensions of the API (see Figure 25 for a graphical representation of the data). The next analyses looked at EPA Source Modality and the dimensions of Facilitation to Learning and Credibility.

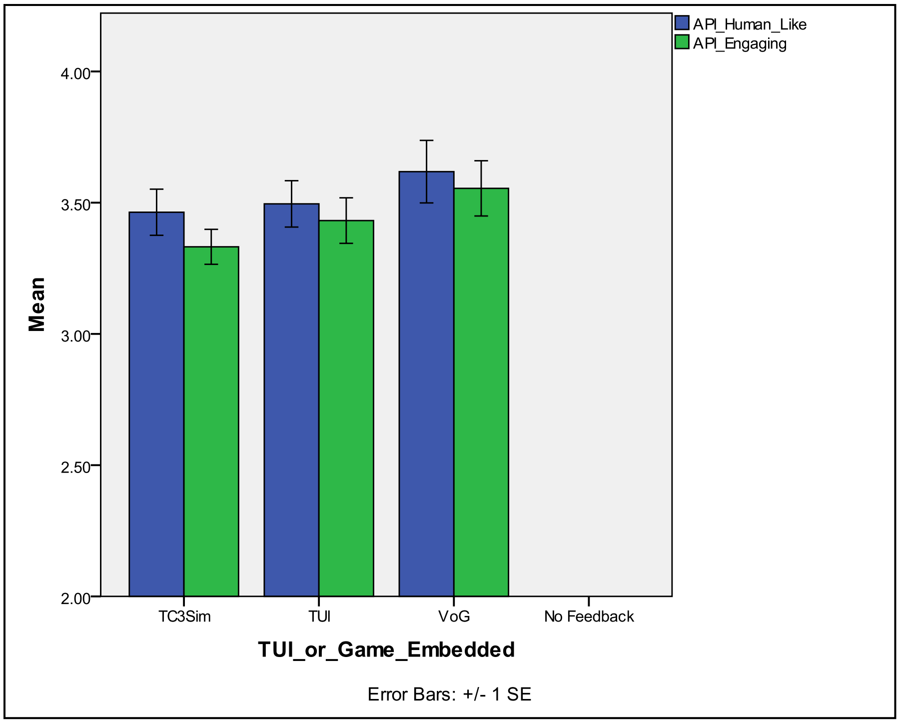


Figure 25. API Scores for Human-Likeness and Engagement across EPA Source Modality Groups

Prediction 4

While the previous prediction is based on the API dimensions of ‘Human-Likeness’ and ‘Engagement’, Prediction 4 is interested with how EPA Source Modality affects a subject’s response across the dimensions of ‘Facilitation to Learning’ and ‘Credibility’ (for a review of the descriptive statistics, see Table 13). In considering the variable of EPA Source Modality, it is important to note that one agent is always visible to the user in a separate browser, while one agent is embedded in the game and is only visible when the character is in the player’s line of sight. Because of this distinction, it is predicted that the TUI-Embedded condition will rate responses on the dimensions of Facilitation to Learning and Credibility significantly higher when compared to TC3Sim-Embedded and VoG conditions. This is a result of the learner having constant visibility of the EPA in the TUI-browser, thus creating a perception that the agent facilitates the delivery of feedback in a more credible manner. This prediction is based on the assumption that the TC3Sim-Embbed EPA is rarely viewed by the learner due to the dynamic nature of the task, while the TUI has a social character in constant view which provides additional grounding of the explicit feedback delivered, making it perceived as more credible than just hearing the words spoken. There is no previous empirical research found investigating this relationship. See Figure 26 for a graphical representation of the data across the EPA Source Modality groups.

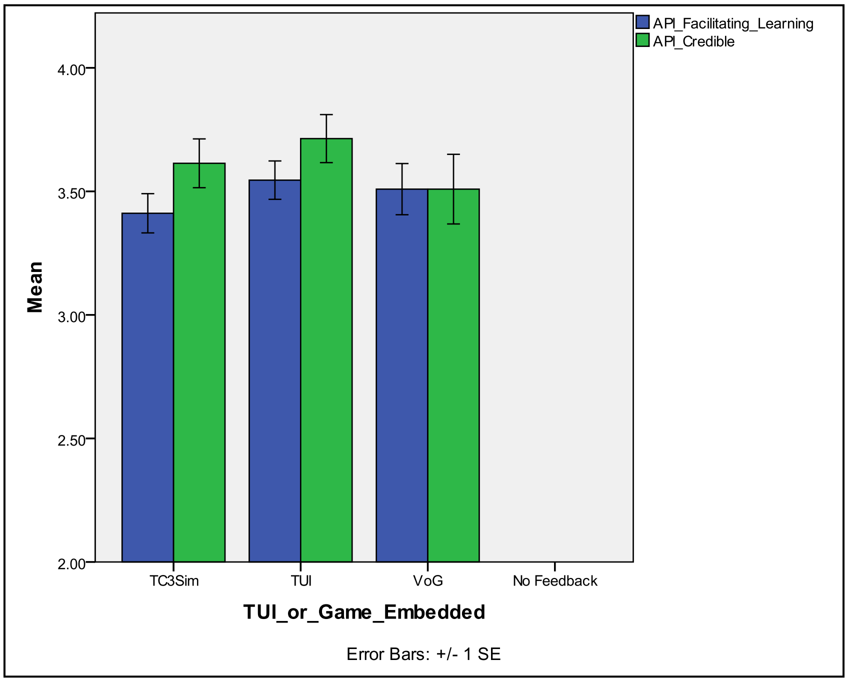


Figure 26. API Scores for Human-Likeness and Engagement across EPA Source Modalities

To determine the efficacy of Prediction 4, a MANOVA was performed examining the EPA Source Modality against the two API dimensions of Facilitation to Learning and Credibility. In examining Source Modality by itself, the MANOVA shows no significant effect of the IV on the two API dimensions of interest, as reported by Pillai's trace ($V = .051$, $F(2, 214) = 1.402$, $p = .234$, $\eta_p^2 = .026$, power = 0.433). For further analysis, Univariate ANOVAs were conducted with results showing the EPA conditions to have no reliable effect on the 'Facilitation to Learning' and 'Credibility' scores when analyzed by themselves (Facilitation to Learning: $F(2, 107) = .779$, $p = .462$, $\eta_p^2 = .014$, power = 0.180; Credibility: $F(2, 107) = .753$, $p = .473$, $\eta_p^2 = .014$, power = 0.175). As determined by the four predictions associated with Hypothesis 3, both the EPA Source Modality and EPA profiles were found to have no statistical effect on how participants responded to the items within the API. For the next set of analyses, statistical tests are conducted looking at subject responses to the RETRO Flow Scale, and how EPA Source Modality and EPA Profile impacted observations.

Hypothesis 4

Following completion of the TC3Sim training scenario, participants completed the RETRO Flow Scale, a 35-item instrument used to gauge an individual's self-reported flow state across eight dimensions. This questionnaire was administered to assess if the feedback source modality manipulations had an effect on how someone rated the level of flow they experienced. Hypothesis 4 predicts that the source modality of feedback will significantly influence an individual's sense of flow within the TC3Sim game environment. The basis of this hypothesis is centered on the incorporation of GIFT's TUI, and the tradeoffs required to implement its function. The game is displayed in a windowed mode for visual access to the EPA situated in the TUI, which is predicted to impact a subject's level of immersion and flow. Predictions associated with Hypothesis 4 examine the Feedback Source Modality and EPA Profile variables to determine the specific effect they have on scores linked to the RETRO-Flow Scale. As the flow scale was administered only following the TC3Sim training scenario, the analysis is limited to examining between-subject effects through the application of ANCOVAs. For a list of descriptive statistics associated with the RETRO-Flow Scale, see Table 14.

Table 14. RETRO-Flow Scale Descriptive Statistics Across Dimensions

Feedback Modality Condition		*Antecedents of Flow*	*Flow Experience*	*Overall Flow*
		RETRO-Flow Scale		
TC3Sim-Peer	*M*	3.36	2.97	63.00
(N = 22)	*SD*	.600	.660	.11.87
TUI-Peer	*M*	3.24	3.06	62.63
(N = 21)	*SD*	.555	.558	.10.42
TC3Sim-Instr	*M*	3.43	3.40	68.71
(N = 22)	*SD*	.506	.360	6.65
TUI-Instr	*M*	3.34	3.16	64.99
(N = 22)	*SD*	.420	.523	8.12
VoG	*M*	3.42	3.10	64.27
(N = 22)	*SD*	.486	.564	9.45
No Feedback	*M*	3.12	3.20	65.48
(N = 21)	*SD*	.446	.458	6.40

Prediction 1

Prediction 1 assesses the variable of Feedback Source Modality and its impact on an individual’s self-reported level of Flow. It is predicted that those interacting with the TC3Sim-Embedded EPA conditions will report significantly higher scores on the dimensions of Flow Experience (i.e., average of inputs across the dimensions of: Concentration, Temporal Dissociation, Loss of Self-Consciousness, Autotelic Experience, and Merging of Action and Awareness) and Overall Flow (i.e., sum of all items divided by maximum total possible, then multiplied by 100; does not include questions on Visual Quality dimension) when compared against the TUI-Embedded, VoG, and No Feedback conditions. This is because the embedded

EPA tutor allows for the game to be played in a full-screen mode, and does not include elements that can lead to distraction, such as GIFT's TUI browser. In comparison, it was predicted Feedback Source Modality would have no effect on a subject's reported score across the dimensions of Antecedents of Flow (i.e., average of inputs across the dimensions of: Mastery of Gameplay and Feedback). This was due to all subjects receiving the same performance-based explicit feedback, thus providing the required resources for an individual to enter and maintain a state of flow. For a list of descriptive statistics on the RETRO-Flow Scale across the Feedback Source Modality groupings, see Table 15.

Table 15. RETRO-Flow Scale Descriptive Statistics Across Feedback Source Modality Groups

Feedback Source Modality Groups		*RETRO-Flow Scale* *Antecedents of Flow*	 *Flow Experience*	 *Overall Flow*
TC3Sim-Embedded	*M*	3.40	3.19	65.85
(N = 44)	*SD*	.550	.568	9.93
TUI-Embedded	*M*	3.29	3.11	63.83
(N = 44)	*SD*	.487	.536	9.27
VoG	*M*	3.42	3.10	64.27
(N = 22)	*SD*	.486	.564	9.45
No Feedback	*M*	3.12	3.20	65.48
(N = 21)	*SD*	.446	.458	6.40

To examine prediction 1, separate Univariate ANCOVAs were performed on the DVs of Flow Experience and Overall Flow, as described above. For this analysis, an individual's reported VGE was applied as a CV, to determine if how often someone plays videogames influences the level of flow they perceive to experience. Results for both tests show the IV of Feedback Source Modality to have no significant main effect for the two metrics of Flow Experience (F (3, 123) = .466, p = .707, η_p^2 = .011, power = 0.142) and Overall Flow (F (3, 123) = .674, p = .569, η_p^2 = .016, power = 0.189). In terms of VGE being a strong predictor of perceived flow, the metric was found to be a significant CV for both variables assessed in this analysis (Flow Experience: F (1, 123) = 4.321, $p < .05$, $\eta_p^2 = .034$, power = 0.541; Overall Flow: F (1, 123) = 6.359, $p < .025$, $\eta_p^2 = .049$, power = 0.706). For graphical representations of the data for each Feedback Source Modality Group, see Figure 27.

Results from this prediction analysis show the modality of feedback to have no significant effect on the level of flow an individual experiences, while also showing a direct correlation between how often individuals play videogames and the flow state they perceive to experience. This finding is important in terms of utilizing GIFT's TUI for feedback delivery, as the resulting windowed display of the game is not enough to remove the immersive element associated with flow and game-based training.

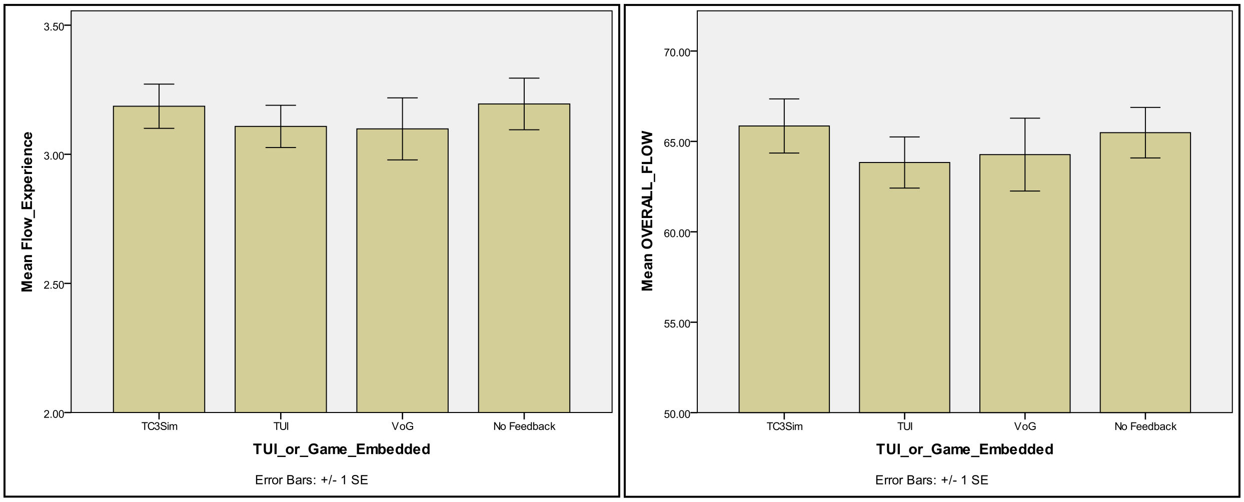


Figure 27. Flow Experience and Overall Flow Scores Across Feedback Source Modality Groups

Prediction 2

Prediction 2 assesses the variable of EPA Profile and its impact on enabling individuals to enter a state of flow. It was predicted that those interacting with the EPA-Instructor conditions would report significantly higher scores on the dimensions of Antecedents of Flow (i.e., average of inputs across the dimensions of: Mastery of Gameplay and Feedback). This was centered on stereotype effects and the persona effect research found in SCT. It was believed that feedback delivered by the defined Instructor EPA would be perceived as more useful, resulting in better gameplay experiences and higher reported flow scores. As a result, Antecedents of Flow was predicted to score higher for subjects interacting with the EPA Instructor Profile when compared against the EPA Peer and VoG conditions.

In addition, to further explore prediction 2 the single dimension of Feedback will be examined to determine if the feedback provided by GIFT was effective enough to be an antecedent of flow when compared against the No Feedback condition. Three questions in the RETRO-Flow Scale were administered to determine if a game provides enough information for a player to gauge performance for achieving objectives (e.g., I received feedback on my progress in the game; I received information on my success (or failure) of intermediate goals immediately; and I knew how well I was playing the game). Responses to these items will be assessed across the EPA Profile conditions to determine if subjects viewed feedback from the game as helpful for performing task elements to reach scenario objectives. For all descriptive statistics of the RETRO-Flow Scale inputs across the EPA Profile conditions, see Table 16.

Table 16. RETRO-Flow Scale Descriptive Statistics Across EPA Profile Groups

EPA Profile Treatment Groups		*RETRO-Flow Scale*			
		Antecedents of Flow	*Feedback*	*Flow Experience*	*Overall Flow*
EPA-Instr	*M*	3.39	3.35	3.28	66.85
(N = 44)	*SD*	.462	.655	.460	7.57
EPA-Peer	*M*	3.30	3.29	3.01	62.82
(N = 44)	*SD*	.575	.735	.607	11.05
VoG	*M*	3.42	3.53	3.10	64.27
(N = 22)	*SD*	.486	.640	.564	9.45
No Feedback	*M*	3.11	2.64	3.20	65.48
(N = 21)	*SD*	.446	.666	.458	6.40

The first test run for Prediction 2 was a Univariate ANCOVA to determine the effect the assigned EPA Profile condition had on reported scores for items related to Antecedents of Flow. To remove any relationship a subject's VGE has on these inputs, VGE was assigned as the CV for this analysis. The ANCOVA results show EPA Profile to have no significant main effect on the Antecedents of Flow metric outcomes (F (3, 123) = 1.932, p = .128, η_p^2 = .045, power = 0.489), along with VGE not being recognized as a significant CV (F (1, 123) = 1.804, p = .182, η_p^2 = .014, power = 0.266). A graphical representation of these relationships can be seen in Figure 28. Though the findings from this analysis were not significant, it is worth noting that individuals in the No Feedback condition reported the lowest marks for the Antecedent of Flow metric. To examine further, the single dimension of Feedback within the RETRO-Flow Scale was analyzed to determine if GIFT feedback produces higher scores on the three items when compared against those who relied specifically on implicit information within the game.

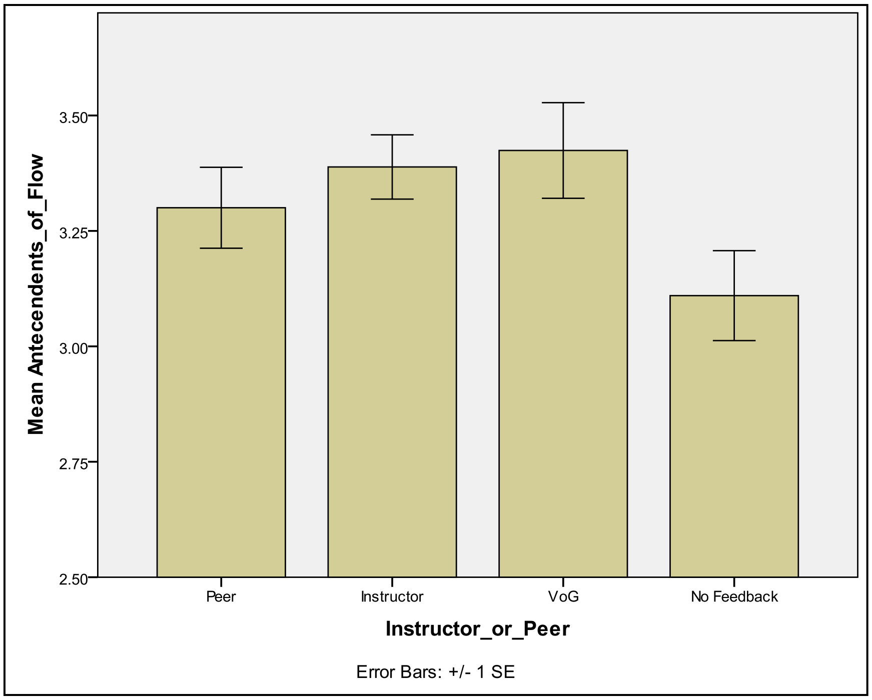


Figure 28. Antecedents of Flow Scores Across EPA Profile Groups

The Feedback flow dimension is calculated by averaging responses across the three questions identified above. With this associated metric, analysis can be conducted to identify significant differences in scores as a result of interaction with a particular treatment. In this instance, a Univariate ANCOVA with VGE being defined as the CV was performed looking at differences in Feedback scores across groups related to EPA Profile. Recorded scores for this variable can be seen in Table 16. Outputs from this test show EPA Profile treatments to have a significant main effect on how individuals scored items within this particular flow dimension (F (1, 123) = 7.609, $p < .001$, $\eta_p^2 = .157$, power = 0.985). The defined CV was not found to be a strong predictor of how participants responded across the associated questions.

To further assess the identified main effect, a post-hoc analysis was performed using the Bonferroni test to identify the specific conditions that produced reliable differences. Results show all treatments with feedback (e.g., EPA-Instructor, EPA-Peer, and VoG) to produce significantly higher scores on the Feedback dimension metric when compared to the baseline scenario. See Table 17 for results linked to the post-hoc analysis and Figure 29 for a visual display of the Feedback metric data. Essentially, results from this analysis show participants receiving explicit feedback during the training scenario, regardless of the condition, enables individuals to track progress towards objectives better than when relying specifically on implicit information to determine how one is performing. This relationship will be examined further in the next chapter.

Table 17. Post-Hoc Analysis of RETRO-Flow Feedback Metric Across EPA Profile Treatments

EPA Profile Groupings	*RETRO-Flow Feedback Dimension*		*Significance*
	Mean	*Standard Error*	
EPA-Peer vs.	3.27	.104	*p* = .003
No Feedback	2.63	.147	
EPA-Instr vs.	3.38	.103	*p* < .001
No Feedback	2.63	.147	
VoG vs.	3.53	.144	*p* < .001
No Feedback	2.63	.147	

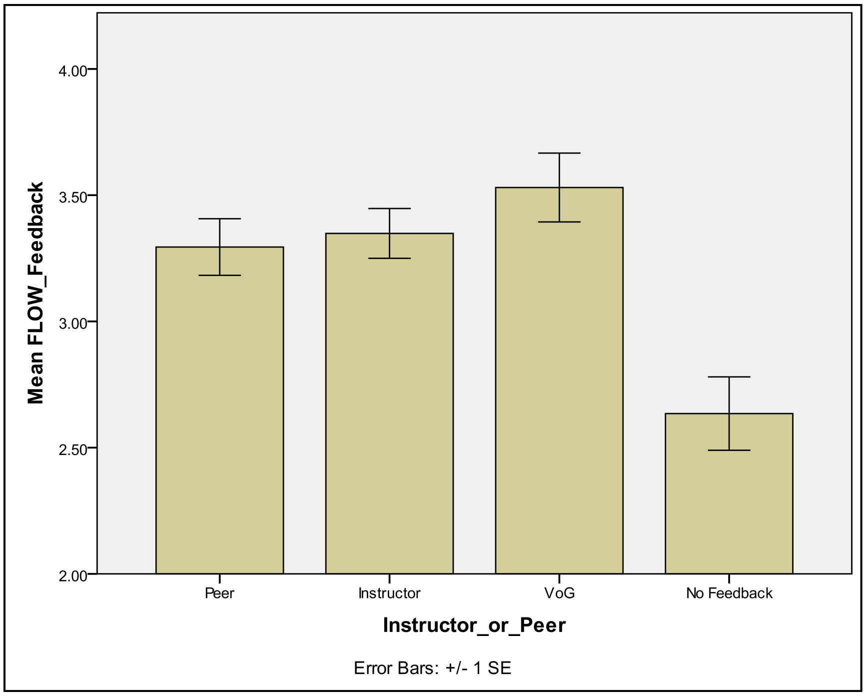


Figure 29. RETRO-Flow Scale Feedback Dimension Scores Across EPA Profile Groups

Prediction 3

The last prediction associated with Hypothesis 4 is concerned with an individual's perceived level of presence or immersion within the game environment, and how the associated treatments impacted a subject's self-reported score. Due to the absence of explicit feedback channels removing the user from the game experience, an initial prediction is that the control condition with no feedback will score the highest on presence dimensions related to the RETRO-Flow Scale (i.e., average of scores across Concentration, Temporal Dissociation, and Loss of Self-Consciousness dimensions; see Table 18). This is based on participants having to rely on implicit information within the environment to gauge progress and next actions, resulting in increases of perceived presence. In addition, it was also expected that participants in the TC3Sim-Embedded EPA conditions would score significantly higher on the dimensions linked to presence when compared against the TUI-Embedded grouping. To examine these hypotheses, an ANCOVA was performed with VGE defined as the CV.

Table 18. RETRO-Flow Scale Descriptive Statistics for Presence Across Feedback Source Modality Groups

Feedback Source Modality Groups		*RETRO-Flow Scale Presence/Immersion*
TC3Sim-Embedded	*M*	3.15
(N = 44)	*SD*	.569
TUI-Embedded	*M*	3.06
(N = 44)	*SD*	.533
VoG	*M*	3.05
(N = 22)	*SD*	.646
No Feedback	*M*	3.10
(N = 21)	*SD*	.561

Results from the performed ANCOVA show no significant main effect of Feedback Source Modalities on the level of immersion/presence reported by subjects following completion of the game event (F (3, 123) = .283, p = .837, η_p^2 = .007, power = 0.103). This signifies that the inclusion of explicit feedback in the game environment does not impact an individual's perceived level of immersion when compared against those who rely solely on implicit information channels to determine what action to perform next. In addition, this result conveys that the incorporation of GIFT's TUI next to the game display does not significantly impact the level of presence a subject reports as experiencing. For a graphical representation of Flow Presence scores across the groups of interest, see Figure 30. With produced results informing each hypothesis, the next chapter will review the experimental outcomes with a discussion centered around the implications and tradeoffs associated with the statistical findings.

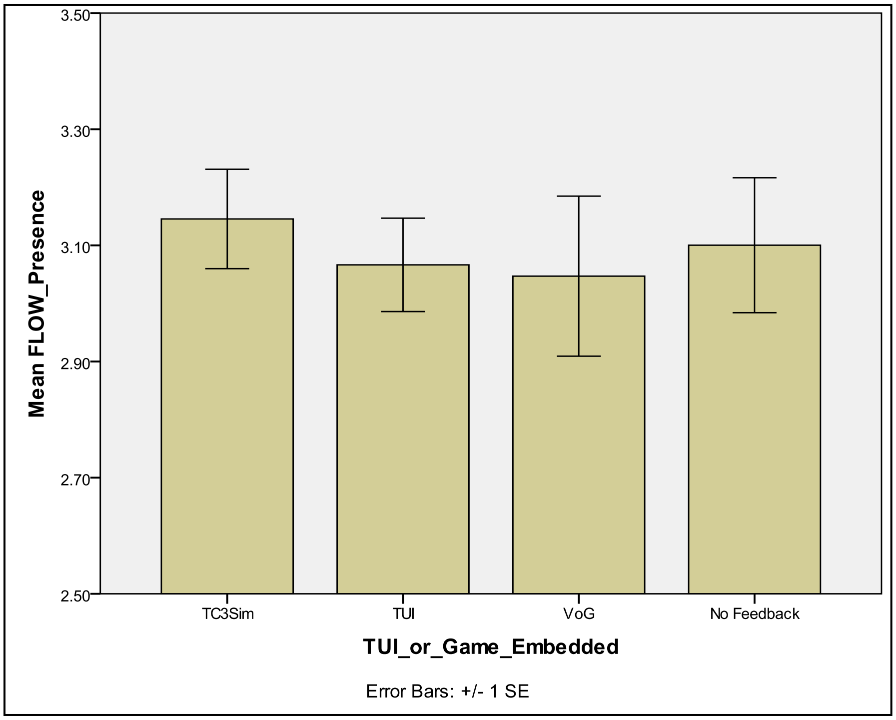


Figure 30. RETRO-Flow Scale Presence Metric Scores Across Feedback Source Modality Groups

CHAPTER 6: DISCUSSION

The use of serious games within education and training communities are on the rise. They provide innovative opportunities for instructors to enable their students to apply newly acquired knowledge and skills in unique environments and under novel situations. While many of the produced games provide these characteristics, where they lack is in the ability to contextualize interaction within a scenario to overarching learning objectives the game was designed to train. Due to this constraint, many of the serious games utilized for education and training require monitoring from an instructor for linking game actions to intended learning events. To combat this limitation, research is being conducted to examine innovative opportunities to embed intelligent tutoring functions within serious game environments that provide the explicit feedback element necessary for effective instruction.

A tool developed to meet this need is GIFT, which is a modular framework that incorporates standardized processes for authoring and managing adaptive functions across linked training applications. The research presented here focuses around the application of GIFT within a serious game environment used by the U.S. Army to train KSAs associated with being a CLS and combat medic. The study focused on two primary outcomes. First, the research was intended to provide empirical evidence supporting the inclusion of explicit feedback in serious game environments by examining performance outcomes across a game integrated with GIFT versus a baseline version. This analysis was designed to determine whether a game embedded with functions provided by GIFT produced benefits that justify its application.

Second, multiple experimental conditions were designed to examine feedback delivery modalities within serious game environments. Research questions were designed around two IVs (e.g., Feedback Source Modality and EPA Profile) that focused on two fundamental theories relevant to available tools within the GIFT architecture: CLT and SCT. Specifically, the study examined if there are significant benefits to incorporating EPAs as feedback delivery mechanisms in game-based environments, and to what effect different interfacing modalities have on dependent variable outcomes (e.g., performance, agent perception, cognitive load, and flow). The source of feedback was manipulated across six conditions, with participants being assigned to one of four primary setups. These involved an EPA located in GIFT's TUI, an EPA located directly in the game environment, feedback delivered from no EPA source (i.e., voice of God), and a baseline condition with no explicit feedback at all. For the EPA related conditions, an additional IV was incorporated (e.g. EPA Profile) to determine if how an agent's profile is presented to a learner affects their perception of the entity's usefulness. Analyses linked to this experimental approach are intended to provide empirical evidence for the efficacy of including virtual human agents as defined EPAs in game-based training environments, with hypotheses developed to recognize tradeoffs between the manipulated IVs of interest.

Outputs from these analyses provide design recommendations for the GIFT user community with specific suggestions for integrating explicit feedback functions in simulation-based training events. The following section reviews the associated results from this experiment and what they mean in the context of game-based training delivery. Research questions are presented as they relate to the experimental design and tradeoffs are identified across the varying feedback source modalities.

Summary of Results

Analyses for this experiment were based on four primary hypotheses. Results are intended to inform research questions associated with explicit feedback delivery in serious game environments and the effect variations in source modalities play on a number of dependent variables. Table 19 was generated to provide a summary of the results and how they relate to the overarching research questions posed in Chapter 1. The following discussion focuses around implications of what the data tells us with respect to the variables of interest and what tradeoffs are identifiable in terms of the varying condition manipulations. Each research question is addressed to highlight specific findings as they relate to statistical outcomes associated with hypothesis testing. Following, tradeoffs between the varying conditions are presented based on the dependent measures collected resulting in a list of recommendations for implementing explicit feedback in game-based environments.

Table 19. Summary of Research Questions, Associated Hypotheses, and Analyses Outcomes

Question	Associated Hypothesis	What the Results Tell Us
1) Does the inclusion of explicit feedback in TC3Sim significantly impact performance?	*Hypothesis1 (Prediction1)*	• Inclusion of feedback is found to have a significant main effect on game performance within the training scenario
2) Does explicit feedback delivered by an EPA provide a distinct benefit when compared against feedback delivered as audio alone (i.e., VoG)?	*Hypothesis1 (Prediction2)*: Game Performance	• Feedback source modality was found to have a significant main effect on performance outcomes within the TC3Sim training scenario • No significant difference was identified between the EPA and VoG conditions • VoG condition was found to produce the highest overall performance scores for the training scenario

Question	Associated Hypothesis	What the Results Tell Us
	Hypothesis1 (Prediction2): Learning Gains/Transfer	• Analysis looking at learning gains associated knowledge assessments show feedback source modality to have a significant main effect • Participants receiving feedback from EPAs performed significantly better on the post-test when compared against the VoG Condition
3) Does embedding the EPA directly in the game world provide a distinct benefit on Cognitive Load, Agent Perception, and Flow when compared to more simplistic interfacing approaches (i.e., TUI)?	*Hypothesis2 (Prediction1 & 2)*	• Feedback source modality is found to produce significant differences on the MD dimension of the NASA-TLX • Both the TUI-Embedded and Game-Embedded tutor groups scored significantly higher on MD when compared against the VoG condition
	Hypothesis2 (Prediction3)	• Presence of feedback did not significantly affect responses on WL and MD metrics • Individuals in the no feedback condition rated MD significantly higher when compared against the VoG condition • No significant difference in WL and MD between No Feedback and all EPA related conditions
	Hypothesis3 (Prediction3 & 4)	• The location of the EPA was found to have no significant effect on how subjects scored responses across all dimensions of the API
	Hypothesis4 (Prediction1 & 3)	• Feedback Source Modality conditions were found to have no significant effect on the reported state of flow participants experienced while interacting with the TC3Sim training scenario • In examining the specific dimensions of the RETRO-Flow Scale related to presence and immersion, the location of the EPA was found to have no effect on reported scores

Question	Associated Hypothesis	What the Results Tell Us
4) Does an EPA's defined profile background impact an individual's perceived level of experienced cognitive load and flow during gameplay?	*Hypothesis2 (Prediction4)*	• Analysis shows minimal variance in reported scores of WL and MD when comparing Instructor vs. Peer affiliations as they relate to the EPA Profile Groups
	Hypothesis4 (Prediction2)	• The Defined EPA Profile groups did not produce significant differences on the Antecedents of Flow scores (i.e., average of Feedback and Mastery of Gameplay dimensions) • In examining Feedback dimension alone, EPA Profile was found to have a significant main effect • All associated tutor groups scored significantly higher on Feedback Usefulness than the No Feedback dimension as reported from the Flow Scale • No identified differences between EPA Profile groups and the VoG condition
5) Does an EPA's defined profile and background influence their perceived competency and usefulness across learners when there are no differences in interaction?	*Hypothesis3 (Prediction1)*	• In examining the Persona Effect highlighted in SCT, the defined EPA Profile groups were found to produce no significant differences on the dimensions of Facilitation to Learning and Credibility
	Hypothesis3 (Prediction2)	• Similarly to Prediction1, analysis shows the EPA Profile groups to produce minimal variance in subject response for the API dimensions of Human-Likeness and Engagement

The first question addressed in Table 19 focuses on the application of explicit feedback in a serious game environment to determine if this added functionality significantly impacted performance outcomes. With much of the previous literature on this topic emphasizing the benefit of providing explicit feedback information in challenging learning contexts (Astwood et al., 2008; Mory, 2004; Shute, 2007), statistical analyses were run comparing performance between individuals receiving feedback and those in the control No Feedback treatment. Hypothesis1 predicted that individuals receiving the explicit feedback information would outperform those individuals in the baseline condition where they relied on implicit feedback to monitor performance. Results show TC3Sim embedded with GIFT's explicit feedback functions produced significantly better scores on game performance when compared against the current baseline version, yielding an effect size of .133 sigma. This outcome shows the mere presence of reflective prompts within the training scenario to influence next actions taken, resulting in better performance marks for the player. Though the effect size reported as rather small, it is important to remember that this shift in performance was the result of a single scenario interaction covering multiple learning objectives. If more exposure to the game was provided where the tutor manipulations were present, it is believed that this disproportion in performance would increase. It is also important to remember that this effect size associated with question one is based on performance from the single training scenario that lasted an average of five minutes.

With data supporting the application of GIFT in the game TC3Sim, question two focused on the inclusion of a social element in the explicit feedback delivery. Based around SCT, research has shown social interaction in a learning setting to increase motivation and comfort with tasks, enhance flow of information, and improve task performance and understanding of material (Bandura, 2011; Vygotsky, 1987). For this reason EPAs were included in the experimental design to determine if this relationship extends into interaction within a game-based learning environment. Research has shown incorporation of EPAs in intelligent tutoring and computer-based instruction to make a difference (Graesser & McNamara, 2010; Kim & Baylor, 2006b; Moreno et al., 2001), yet much of this research was conducted within rather static learning environments that do not dynamically change throughout the experience. The question this research sought to answer is if it is worth the effort to include social elements for feedback delivered by an external ITS embedded in a highly interactive gaming environment, with the presumption being that this would assist in grounding the feedback to a source so learners had a better chance of interpreting the information efficiently to assist in task execution and retention.

Hypothesis1 further predicted that individuals receiving feedback from an EPA would demonstrate significantly better performance outcomes when compared against the two control conditions. In examining the effect an EPA had on performance within this study, it was found that individuals within the VoG condition scored highest in the TC3Sim training scenario when compared against all EPA related conditions and the baseline with no feedback. From this perspective, the inclusion of an EPA shows no true benefit. Individuals who received feedback prompts as audio alone performed the best, but results were not significantly better than those with EPA treatments.

The real insight on an EPA’s effect on performance is seen in examining performance on subsequent assessments (i.e., capstone scenario and post-test). According to Schmidt and Bjork (1992) it is critical to add transfer and retention phases when comparing treatment conditions on learning effect, as these subsequent measures are often better indicators of the IVs influence on performance differences across groups. In these analyses the EPA conditions were found to perform significantly better than the VoG. The results from this analysis indicate that the presence of an EPA during game interaction led to better outcomes on subsequent interaction within similar problem spaces, leaving the VoG condition as the only treatment to produce negative learning gains and transfer across both the game and knowledge-test metrics. Hence, while VoG was shown to result in the highest performance outcomes in the TC3Sim training scenario, this treatment was shown to have the weakest transfer to alternate problems and retention of domain related facts. This finding supports SCT in that grounding information through a social source aids in perception of information and management of short- and long-term memory, resulting in better conceptual understanding of the material (Gulz, 2004).

With evidence supporting the incorporation of EPAs as feedback delivery mechanisms in serious game environments, the remaining research questions were based around the two defined IVs of interest: Feedback Source Modality and EPA Profile. With the intelligent tutoring architecture GIFT playing a key role in the experimental design, a major thrust of this research was to examine interfacing options offered by the framework to provide empirical evidence to support the efficacy of their use. The component of interest for this study was GIFT's TUI and how it can be used to interface feedback information with a learner during a game-based learning event.

The research question generated around the TUI feature was based on identifying distinct advantages/disadvantages associated with the different modalities used in experimentation. Question three in Table 19 covers work surrounding CLT and how individuals interface with technology, SCT and how learners perceive agents based on appearance and application, and elements of perceived Flow and immersion during game interaction. Hypotheses were defined for each research avenue mentioned above, with Hypothesis2 focusing on the effect Feedback

Source Modality has on individuals' self-reported WL and MD rankings. As highlighted by Sweller, Van Merrienboer, and Paas (1998), CLT within an instructional design context is concerned with identifying the optimal approaches for delivering information to a learner that avoids overloading their WM capacity. While the goal of CLT based instructional design is to reduce the amount of extraneous load a learner experiences during interaction, it is important to understand how integrating explicit feedback in game-based environments affects the level of mental effort necessary to efficiently interpret this channel of information without taking cognitive resources away from the task being performed in the virtual environment.

For this study two predictions were posed as they relate to where the EPA was situated during the TC3Sim training scenario and its effect on perceived cognitive demand. Each prediction was based around different perspectives of Wicken's (2002) MRT, with dual task and ambient vision theories providing the basis for the design. Interestingly, the data revealed no differences in self-reported WL and MD as collected from the NASA-TLX across all four associated EPA conditions, yet both the TUI-Embedded and TC3Sim-Embedded EPA treatments scored significantly higher on the MD metric when compared against the VoG condition. This result conveys that the incorporation of an EPA increased the level of mental effort used by a subject when interacting within the serious game environment. If a learner knows information will be delivered that will assist them in performing their tasks, they will be more prone to apply additional cognitive resources so explicit information is not missed over. In the context of the VoG condition, participants were not notified explicit feedback would be provided, resulting in less effort to monitor information not implicitly provided by the game.

This finding may assist in explaining why individuals in the VoG condition scored the highest during the training scenario, while producing the worst transfer results on the subsequent assessments. In the VoG treatment, participants are reacting to feedback provided by GIFT as if it is part of the game, due to removal of the EPA introduction that notifies the subject explicit information will be provided. Based on this association, it appears to be beneficial to provide upfront information to the learner that feedback will be provided linking game interaction to overall learning objectives the system is designed to train. This may assist the learner in associating formative feedback information with knowledge schemas in memory for correcting or reaffirming knowledge components (Shute, 2007). An additional prediction posed to Hypothesis2 was that subjects in the baseline No Feedback treatment would report the highest WL and MD scores due to relying on implicit information from the game alone to gauge performance towards meeting objectives. Similarly to all EPA conditions, the No Feedback condition reported higher MD scores when compared against the VoG condition, with no significant differences seen between the control and the Feedback Source Modality treatments.

The next analyses run against question number three considered whether the Feedback Source Modality IV influenced how individuals scored on metrics associated with agent perception and flow based on self-response measures collected from the API and RETRO-Flow Scale. In terms of agent perception, results did not support predictions defined within Hypothesis3. While it was believed that conditions with the EPA present in the game environment would produce higher scores on the dimensions of Human-Likeness and Engagement and lower scores on the dimensions of Facilitation to Learning and Credibility, the collected data showed the location of the EPA to have no effect on how subjects responded across all items on the API.

Predictions were also made within Hypothesis4 that were concerned with question three and Feedback Source Modality's effect on an individual's perceived level of flow during game interaction with an EPA. Prediction1 posited that the TC3Sim-Embedded EPA conditions would score significantly higher on items linked to the dimensions of Flow Experience when compared against the TUI-Embedded treatments, while Prediction3 hypothesized that the control with No Feedback would score highest on the specific dimensions linked to presence and immersion. Prediction1 was based on the notion that the incorporation of the TUI requires the game to be displayed in a windowed mode, removing the element of full-screen immersion. Results from the analysis showed the Feedback Source Modality IV to have no significant effect on Flow Experience and Presence dimensions within the RETRO-Flow Scale. This finding supports the application of the TUI as an effective tool to house an EPA for feedback delivery during game-based interaction. Though the visual field of the game environment is reduced, the display was large enough for players to become cognitively immersed in the environment. This is an important finding, as results suggest the inclusion of an EPA to be beneficial, yet their application can often be expensive and labor intensive to implement. With the TUI producing similar cognitive load and flow scores when compared to the TC3Sim-Embedded treatments, the true benefit is in the domain-independency and reusable agent entities the TUI provides in authoring EPA interaction functions.

The next research question posed in Table 19 focuses on the IV of EPA Profile. Specifically, question four seeks to identify if how an agent is presented to a learner prior to game interaction affects the way that individual scored items associated with the dependent measures of cognitive load and flow. It was predicted in Hypothesis2 that the background and biography of the EPA presented to the learner would not produce significant differences in cognitive load based on responses for WL and MD measures from the NASA-TLX. Though the EPAs are presented as being different from one another, the interaction they provide during the scenario is the same for all associated conditions. Because the events within the scenario remained the same, the EPA Profile was not expected to change an individual's perception of how difficult the game was. Results from the data show minimal variance in reported scores of WL and MD when comparing Instructor vs. Peer affiliations as they relate to the EPA Profile groupings.

Part two of question four is concerned with the effect EPA Profile has an individual's reported level of flow experienced during gameplay. Based on the defined role of the EPA, it was predicted in Hypothesis4 that the EPA Profile would produce a significant effect on how subjects scored items on the RETRO-Flow Scale's Antecedents of Flow dimensions (e.g., mastery of gameplay and feedback). The antecedents of flow references the elements that need to be in place for an individual to effectively enter a state of flow as described in Chapter2. In terms of the RETRO-Flow Scale, a game must have elements that enable an individual to feel a sense of control over the game (e.g. mastery) and feedback information to assist that individual in succeeding through scenario interactions. Based on research surrounding the *persona effect* (Baylor & Kim, 2005; Lester et al., 1997b; Veletsianos, 2010) it was predicted that the Instructor profile would be perceived as more useful by the learner when compared to the Peer conditions resulting in higher marks, and that both EPA Profiles would score higher when compared against the VoG treatment. Results from the analysis failed to support this prediction as the defined EPA Profiles did not produce significant differences between any of the conditions.

Following, analysis was performed looking at individuals' responses to the items associated specifically with the Feedback dimension to see if significant differences existed across EPA Profile conditions. The RETRO-Flow Scale includes three questions that gauge how useful someone perceives feedback to be within a game environment, and these items were examined together. By examining the Feedback dimension alone, results showed all conditions that incorporated explicit feedback from GIFT to score significantly higher on these items when compared to the control No Feedback condition. This further supports the application of GIFT in serious games. Not only did the explicit feedback produce better results on performance assessments, subjects interacting with TC3Sim embedded with GIFT reported the game to provide more helpful information to assist in achieving task objectives when compared against the current baseline version being used for training. No differences were found between the EPA Profile and VoG conditions.

The last question presented in Table 19 looks at the EPA Profile IV and the impact it has on self-reported scores across the dimensions of Ryu and Baylor's (2005) API. Based on the different defined profiles, Hypothesis3 presents varying predictions as they relate to the four dimensions that make up the API. This research question is linked to previous work on the *persona effect* and associated stereotyping research involving interaction with virtual agents in synthetic environments (Lester et al., 1997a; Veletsianos, 2010). In terms of this experiment, it was believed that the Instructor and Peer profiles would lead individuals to score the varying dimensions of the API differently. Prediction1 within Hypothesis3 expected the Peer profile to score significantly higher on Human-Likeness and Engagement dimensions of the API, while prediction2 stated that the Instructor profile would produce larger outcomes on the dimensions of Facilitation to Learning and Credibility. Results from the analysis did not support these predictions, as minimal variance was produced between EPA Profile groups across all of the API dimensions. The data show that an introductory bio/profile description of the EPA did not produce perceptions as a result of the *persona effect* as the interaction in the game was the same regardless of the condition assigned. In terms of TC3Sim, the use of varying backgrounds and profiles as a form of instructional strategy is not recommended. However, results from the various analyses show the mere inclusion of an EPA introduction to be beneficial.

With a summary of results linking hypothesis outcomes to defined research questions, the next section focuses on identifying tradeoffs between the experimental conditions applied and what they mean in terms of implementation. The section will conclude with recommendations for authoring EPA functions in game-based environments based on tools and methods provided by the GIFT architecture.

Tradeoff Analysis

Results from this experiment showed variations in a feedback's source modality to have an effect on measures related to performance, cognitive demand, and flow. While the data exhibits differences in outcome values that are attributable to the IVs of interest, it is important to recognize the strengths and weaknesses associated with each approach to identify tradeoffs that require consideration when authoring adaptive functions using the GIFT architecture. The technique being applied for this task is derived from Kalyuga et. al.'s (1999) methodology to produce an instructional effectiveness metric, as described in Chapter 2's Cognitive Load Theory Applied to Instructional Design section. The benefit behind this approach is that it allows you to observe the effect experimental conditions have on outcomes for two defined DVs and their relationship when compared against the designated control. The approach will be administered similarly to Kalyuga et al.'s (1999) implementation in that it will be applied only to the experiment's associated transfer tests.

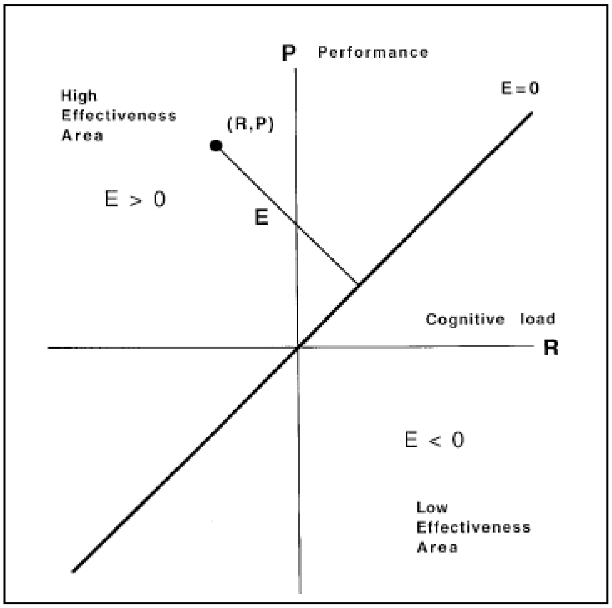


Figure 31. Representation of relative condition effectiveness (Kalyuga et al., 1999)

The effectiveness metric is derived from calculated *Z*-scores across two variables as they relate to the control treatment, which are then represented as a coordinate system to provide a visual representation of the experimental condition's efficiency. For this tradeoff analysis, the variables examined are those that were found to have significant differences across experimental treatments (i.e., Test Performance, Game Performance, Mental Demand and Feedback

Usefulness). With *Z*-score values calculated for each variable, the following formula is applied: $E = (P \ \ +/- R)// SQRT(2)/$ (Kalyuga et al., 1999), where performance *Z*-score (P) and Mental Demand/Feedback Usefulness *Z*-score (R) produce a value to determine training effectiveness (E). In this instance, the MD *Z*-score is subtracted from performance, while the Feedback Usefulness *Z*-score is added. This is based on an assumption that low MD compared to the control is desired, while Feedback Usefulness is desired to be higher. The resulting coordinate point is measured against the line of zero effectiveness ($E = 0$) (see Figure 31).

The analysis was conducted from two perspectives. The first is looking at in-game performance in relation to the subject's perceived level of mental effort exerted and their rating of how useful the feedback provided during interaction was. The self-reported scores of MD and Feedback Usefulness in the TC3Sim training scenario were compared against the subject's performance on the subsequent capstone scenario administered for skill evaluation. First, this shows if the level of perceived cognitive demand linked to the training scenario correlates with performance outcomes on the subsequent delivered assessments. In addition, this technique also shows the relationship between the usefulness of feedback information in a training scenario and its effect on performance in a transfer setting. Each experimental condition will be represented, with an associated effectiveness score provided based on the formula presented above (see Figure 32 and Table 20). The second perspective associated with this analysis is by examining the same DVs of MD and Feedback Usefulness in conjunction with outcomes from the knowledge post-test following completion of the capstone scenario in TC3Sim. This assists in examining if the feedback provided is attributable to increases in knowledge acquisition, as well as providing a way to observe how elements linked to game interaction affect test scores.

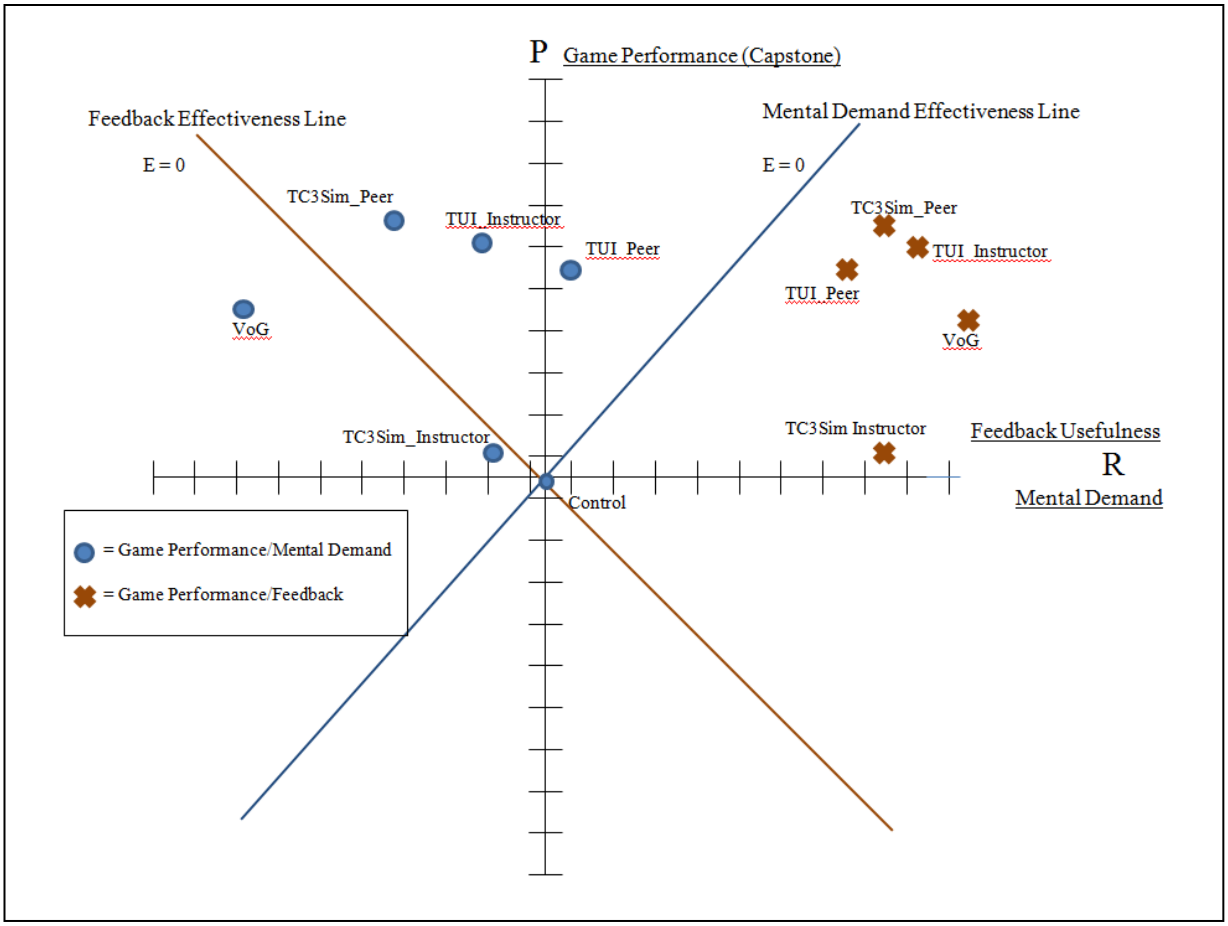


Figure 32. Relative Condition Effectiveness When Comparing Game Performance With Mental Demand and Feedback Usefulness (*X* and *Y* axes represent relative z-score in relation to control condition)

Table 20. Condition Effectiveness Scores fore Game Performance

	Effectiveness Score (TC3Sim Capstone Scenario/Mental Demand)	Effectiveness Score (TC3Sim Capstone Scenario/Feedback Usefulness)
TC3Sim_Peer	0.776449573	1.068820522
TUI_Peer	0.321286185	0.906424501
TC3Sim_Instructor	0.212958657	0.680994443
TUI_Instructor	0.576263348	1.064891379
VoG	0.878041922	1.117114203
No Feedback	0	0

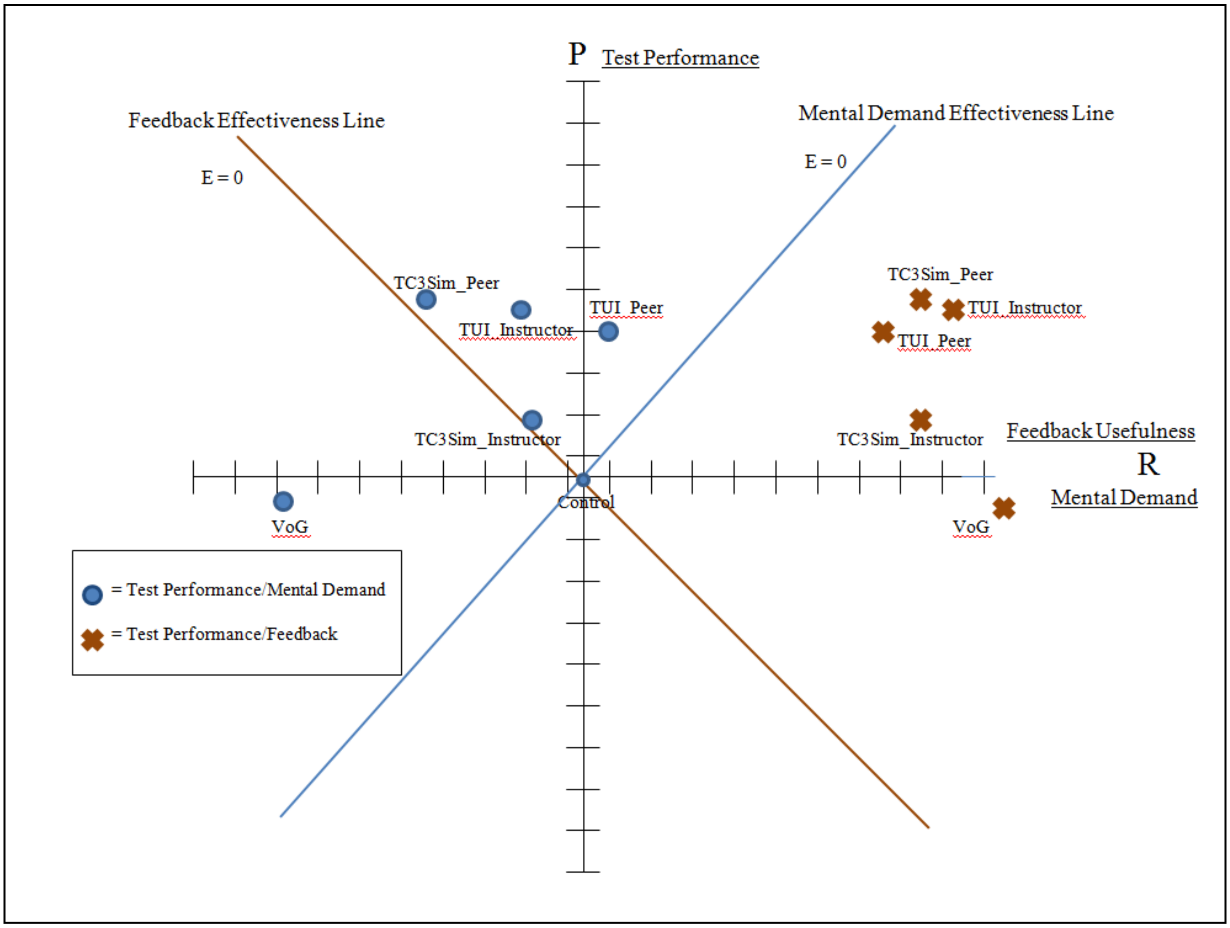


Figure 33. Relative Condition Effectiveness When Comparing Post-Test Performance With Mental Demand and Feedback Usefulness (*X* and *Y* axes represent relative z-score in relation to control condition)

Table 21. Condition Effectiveness Scores fore Post-Test Knowledge Performance

	Effectiveness Score (Knowledge Post Test/Mental Demand)	Effectiveness Score (Knowledge Post Test/Feedback Usefulness)
TC3Sim_Peer	0.648370892	0.940741841
TUI_Peer	0.218835907	0.803974223
TC3Sim_Instructor	0.265292837	0.733328624
TUI_Instructor	0.460266144	0.948894175
VoG	0.483211518	0.722283799
No Feedback	0	0

In examining the condition effectiveness scores for both DVs in relation to game performance, it would appear that the VoG condition rates superior in respects of MD experienced and rating of Feedback Usefulness. However, there is a discrepancy when interpreting the tradeoffs associated with these outcomes. In looking at the visual layout of the data on Figure 32, it shows the VoG condition to score lower on game performance for three of the four treatments involving an EPA character. Due to this, the analysis is utilized as a way to facilitate tradeoff discussions, not as a metric to produce recommendations from. Because game performance in the VoG condition is below EPA treatments, it is important to breakdown the game performance graph above to better understand how the effectiveness scores favoring the VoG condition were produced.

The strength in the VoG's effectiveness is primarily attributable to inputs for both MD and Feedback Usefulness. For MD, scores in VoG were found to be significantly lower than all other experimental conditions. The question is why do subjects in the VoG condition report interaction to be less cognitively demanding? It is the opinion of the author that this is due to a cognitive prompting effect linked with the EPA Profile treatment. Participants in the EPA related conditions are presented a tutor profile introducing the character and notifying them that their performance is being monitored and that feedback would be provided based on real-time assessment, thus prompting the individual that feedback would be linked to objectives the game is intended to instill. This was not the case for the VoG condition, as feedback was delivered in audio format alone with no grounded social source to link the delivery to. As a result, these participants were not expecting feedback and most likely viewed the reflective prompts triggered by GIFT as elements associated with the scenario itself. Due to this, subjects marked MD as low while also scoring Feedback Usefulness higher than all other conditions. This trend results in the VoG to produce the highest effectiveness scores as performance relates to game interaction alone. Another possible explanation is based around the inclusion of a social character element that adds an element to the game requiring additional cognitive resources. However, as an increase in performance is the overarching goal of including explicit feedback in games, the value of this variable must be considered higher when talking about tradeoff considerations.

To follow-up effectiveness interpretations based on game performance, the same effectiveness scores were produced in relation to performance outcomes on the knowledge post-test. In this case, all four EPA conditions produced higher relative effectiveness scores when compared to the VoG in terms of Feedback Usefulness. This is in contrast to game performance. While the VoG reported the lowest cognitive demand and the highest in perceived Feedback Usefulness during training, these subjects showed the poorest transfer of knowledge. This goes back to the argument posed by Schmidt and Bjork (1992) in that measuring effectiveness of an experimental treatment often requires analysis of performance within subsequent assessments, rather than interpreting outcomes from the interaction where the manipulation was present. Specifically, participants in the VoG condition would appear to disregard the explicit feedback provided as it is not grounded to a pedagogical function, resulting in both lower MD during the scenario and lower retention of knowledge as measured in the subsequent post-tests.

In terms of identifying tradeoffs between EPAs and the VoG approach, the data support the inclusion of an EPA in the context of this training application. Though the VoG was effective in aiding a learner to perform, the reduction in performance on transfer tests creates concerns on its effect in aiding a learner to commit feedback content to memory for future application. When comparing the effectiveness scores for EPA conditions alone, the outcomes indicate that one experimental condition is at the top of both performance categories. If one were to take these values as whole truth, the recommendation of feedback source modality including an EPA would be TC3Sim_Peer, suggesting the most effective implementation of explicit feedback in TC3Sim would be by embedding a GIFT agent directly in the scenario environment and defining that agent as a peer or fellow teammate. However, based on the extensive analysis looking at all associated treatments, it would be difficult to pick one condition that is hands down better than the rest. Based on outcomes from the analysis as a whole, embedding a tutor in the game world rather than using GIFT's TUI shows no distinct benefit on performance or across any of the collected DVs linked to cognitive load, agent perception, and flow. Because of this, using GIFT's TUI can provide a large advantage for incorporating an EPA element in a game-based training application because it drastically reduces the amount of time, effort, and money to modify a game to support character interaction requests from the tutor's pedagogical model.

In addition, the EPA profile was not found to affect performance or responses on the self-report DV measures, so the outcomes do not support the use of one approach over the other. One finding supported by the data is the use of a profile description notifying the participant that an agent will be present and that feedback will be provided based on performance. This simple narrative notifies the learner that information will be provided that they may want attend to and remember. Another interesting thought pertains to the effect the inclusion the EPA had on MD measures. Although EPAs were found to produce higher MD scores in comparison to the VoG treatment, this increase in perceived MD may signify a causal factor for why these subjects performed better on subsequent assessments. The more mental effort devoted to a training task, the more structured the knowledge schemas associated with that interaction become, as long as enough resources are available to promote positive germane load. This may be caused by incorporating an introduction informing the learner of social elements that will deliver explicit feedback, prepping the individual to use additional cognitive resources to efficiently perceive those channels of information. While the data enables in depth discussion on tradeoffs for implementing explicit feedback in game environments, it is necessary to cover the limitation of the study before conclusions can be drawn.

Study Limitations

In the execution of this experiment there were a number of limitations encountered that should be brought to light. First, this was the initial use case of TC3Sim paired with GIFT, which limited the pedagogical functions available for providing feedback. As this study fed development requirements for enabling GIFT to monitor interaction in real-time within TC3Sim, a challenge that required special attention was how to associate assessments being performed by SIMILE with concepts and objectives being tracked by GIFT. As SIMILE enables a game developer to build rule-based models of performance around game-state messages, linking these rules to concepts and what that means in real-time was what made this process difficult. As a result, the first implementation of linking performance to objectives was by monitoring events in the game as they relate to time and entity location. It was recognized that time and entity locations are major performance variables in such dynamic operational environments. Outcomes in hostile environments are context specific, and time to act and location of entities are critical metrics that require monitoring. From there, if a participant had not performed an action in the game or violated a rule that maps to an associated concept, GIFT could provide reflective prompts to assist the individual on what action to perform next.

Essentially, certain events in the scenario defined different windows of assessment that were associated with different grading parameters. For example, when the explosive device is detonated in each scenario, that triggers a timer associated with concepts linked to Care Under Fire. It is assumed that following the trigger of an event an expert would perform certain actions within a certain timeframe, and that was the basis for development of the assessment rules in SIMILE. An example during Care Under Fire is that a Soldier is tasked with suppressing enemy advances by engaging them. If the game does not report a state message communicating the player has fired his weapon within the first 10-seconds of that phase, a rule would be violated which is then communicated to GIFT for triggering a feedback intervention to inform the subject that they should be returning fire. As the experiment was primarily focused on the effect of different variations in feedback modalities, the type of feedback is not a pressing issue, but many of the findings and discussion is based around the feedback source modality's effect on performance. That is why it is necessary to discuss the limitations and assumptions associated with the feedback used, as it may not be the optimal approach in the context of the learning environment with novice users.

The second limitation of this experiment was the amount of time allotted for data collection. With the cadets of West Point being the population of interest, each experimental session was limited to a maximum of 1-hour due to their associated heavy workload. Being the case, the experimental procedure had to be designed with this is mind. In an ideal situation, more time would have been allotted for a number of the phases in the procedure. Particularly, each participant would have completed all the courseware on TC3 linked to the learning objectives covered in the game and on the tests. However, this was not realistic. To accommodate the time limitation, a custom set of slides was created from the available materials that covered the most relevant information as it pertained to the context of the experiment. Though it failed to cover every aspect of knowledge associated with the domain, the resulting slide deck was a solid representation of all the procedures required to complete the defined task effectively. If more material was covered in this phase of the procedure, performance scores may have altered as a result. In addition, the length of the experiment may have impacted the level of effort an individual subject put forth towards performance. Because of its short runtime and the outcome having no consequence on a subject, a longer experimental session would have been ideal (Van Merriënboer & Sweller, 2005).

In addition, time limited the amount of exposure a participant had within the game environment. If time were not an issue, each subject would have had more time to learn the controls of the game and practice treating casualties before being exposed to the first of two assessment based scenarios. For the experiment participants were given time within a tutorial to go over all the interfaces but were thrown directly into a difficult scenario involving enemy forces and multiple casualties. This may have been a bit too much for some of the less experienced gamers in the sample, as a short window in a tutorial is not enough to learn the controls proficiently enough to perform at standard in a challenging scenario. In this instance, some subjects knew what steps needed to be performed, but they struggled with the controls to find the proper input. In the event where a learner does not proficiently know the controls of a game, the external cognitive load may be so high the performance is compromised as a result of not knowing how to interface with elements in the game environment. Another possibility is that the scenarios in TC3Sim were just too difficult. That is why the majority of participants only performed half of the actions linked to expert performance as defined by the SIMILE assessment models.

A third limitation of the study was the available technologies for incorporating an EPA within GIFT's TUI. The application used for the experiment was MediaSemantics virtual human software, which is a simple low-cost plug-and-play animation package. The characters are not the most life-like and their movements are quite limited, but the program met the requirements laid out for this study. In future research it may be beneficial to test the research questions addressed in this experiment against virtual human software that produces a much more interactive and visually rich EPA. Available technologies include the Institute for Creative Technology's (ICT) Virtual Human Toolkit (ICT, 2013) and VCom3D's VCommunicator Studio and Gesture Builder (VCom3D, 2013). The distinction in the character's appearance and movement may have been enough to cause participants to reduce their scores on items linked to the agent persona.

A final limitation worth mentioning was the selected approach for collecting flow-based metrics. To ensure collection of information centered on flow while maintaining a 60-minute experimental procedure, a self-report method was selected for ease of administration. The RETRO Flow Scale was selected because of its multidimensional design and its inclusion of items centered on the required conditions of flow. Another avenue receiving a lot of recent attention is the use of sensor-based technologies to collect physiological and behavioral information correlated with affective and cognitive states. Previous research has examined sensor-based modeling approaches across a number of psychological constructs, such as engagement, attention, anxiety, fear, and frustration. In terms of ITS research, this approach can potentially enable a system to track a learner's reactive states and adapt instruction when a negative state to learning is being experienced (e.g., boredom, frustration, etc.). Producing models that monitor markers of flow in real-time (e.g., engagement, eye tracking, posture, etc.) could advance the assessment capabilities of game-based learning environments.

The challenge with this approach, and why it hasn't seen wide application yet, is the difficulty in accurately assessing the state being experienced across a large population and the costs associated with quality equipment required to obtain quality data. In addition, sensor technologies often require calibration and baseline procedures that are often difficult to conduct and time consuming to run. In terms of this experiment, five Affectiva Q-Sensors were used over the course of the five day data collection. The Q-Sensor is a wireless Bluetooth device that collects Electradermal Activity on the surface of the skin, and has been found to correlate with variables linked to arousal. The data was not considered in this analysis, as the inclusion of the sensor data falls outside the scope of the addressed research questions and the logs showed a lack of useable data for a large portion of the subjects. The data will be explored for subsequent publications.

Future Work

The outcomes resulting from this study will inform future research efforts associated with the GIFT architecture and instructional strategy implementation for individualized tailored learning. While GIFT provides the tools necessary to author and deliver adaptive learning applications, an additional function of the framework is to operate as a testbed for the purpose of running empirical evaluations on research questions that will influence future developmental efforts. Empirically evaluating developed models and techniques is essential to ensuring the efficacy of GIFT as a sound instructional tool. To accommodate this requirement, while maintaining domain-independency, GIFT's design is completely modular. This allows for the swapping of specific parts within the framework without affecting other components or models. Modularity enables easy authoring of comparative systems designed to inform research questions driving future development. The framework is structured to support a variety of experimental design approaches, including ablative tutor studies, tutor vs. traditional classroom training comparisons, intervention vs. non-intervention comparisons, and affect modeling and diagnosis research (Sottilare, Goldberg, Brawner, & Holden, 2012). As GIFT is scheduled to deliver a new version of the software to the public every six months, this iterative development allows for an easy transition of experimental outcomes into a baseline version the user community can access.

Yet, for GIFT to be effective across all facets of learning, there are a number of research questions that need to be addressed. These include, but are not limited to: (1) How can GIFT be used to manage the sequence, pace, and difficulty of instructional content before a learning session begins, as well as how to adapt instruction in real-time based on learner model metrics?; (2) What information is required in the learner model to make informed decisions on instructional strategy selection?; (3) How can GIFT best manage guidance and feedback during a learning session based on competency and individual differences?; and (4) What is the optimal approach for delivering GIFT communications to a learner during system interaction? These questions vary from those previously explored in the field due to GIFT's domain independency, requiring generalized methodologies that can be applied across multiple systems and course topics. While GIFT is not directly considered an ITS on its own, it provides all the tools and applications necessary to author stand-alone applications that can be delivered to a group of learners. With that said, much of the research focused around GIFT at the current moment is developing tools to aid in the authoring process and to assist in instructional design by recommending pedagogical strategies on a general level that have been empirically found to impact learning outcomes.

In terms of the feedback research addressed in this work, the experiment was intended to examine GIFT's utility within a dynamic serious game and to evaluate approaches for delivering external communication without negatively affecting performance outcomes. The results conveyed interesting findings that support further application of GIFT's TUI to interface real-time explicit feedback information with a learner. More research is needed to explore the varying options the TUI provides for delivering information, and to determine what applications the various approaches work best within. A specific fallout study resulting from this research is investigating the effect the inclusion of text in the TUI has when an EPA is also present during game interaction. This is contrary to findings from research surrounding the modality principle and redundancy effect (Mayer & Moreno, 2002; Shute, 2007). However, it is believed that with some of these applications being highly dynamic, especially TC3Sim, having text present in the TUI as a form of feedback history may be beneficial for the learner as events in the environment may hinder cognitive resources required to effectively interpret the information provided to assist performance.

In addition, it is necessary to examine the methods applied in this experiment across different game genres, as the results from this study are most likely not generalizable outside of first-person shooter (FPS) type applications. As such, the likeability of virtual entities may be of more importance in games where interaction is more static and character inputs are vital to game progression. This is evident in Role-Playing Games (RPGs) where specific narrative and discourse is performed between avatars and NPCs, which facilitates the core game interactions within a scenario. This is drastically different from the interaction experienced in this experiment, where the EPA was an added element that did not impact scenario progression. In terms of serious games that utilize RPG type formats, intelligent tutoring approaches will vary as the targeted learning objectives will be modeled around the realistic actions undertaken in gameplay. In addition, avenues to communicate explicit feedback information will vary, as embedding agents directly in the environment to facilitate this function may be difficult or not feasible. In this case, the use of GIFT's TUI may provide a distinct benefit to incorporating additional social entities that were not originally included in the game development.

It will also be beneficial to look at available software applications for authoring EPAs that can live within the TUI. As MediaSemantics provided a nice base for this research, the visual and immersive characteristics of those agents left a lot to be desired, as well as the fact that a license is required defeats the open source intent of GIFT. In terms of visual appeal, the MediaSemantic characters are static in movement and lack many of the gestures and expressions that make a well designed virtual human realistic. There are other options that can be explored that were mentioned above, such as ICT's Virtual Human Toolkit (ICT, 2013). As more tools are made available within GIFT that provide this function, further empirical evaluations can be conducted to identify the ideal approach for integrating EPA

In addition to formative feedback research, a recent function added to the GIFT architecture that will inform future studies is the University of Memphis' AutoTutor, a natural language dialog-based ITS used to support conversational learning activities through Q&A that promotes reflection and deep understanding of domain material (Graesser & McNamara, 2010; Graesser, Person, Harter, & Group, 2001). In terms of GIFT, this provides a new set of pedagogical options when authoring a new adaptive capability in a training system. The question is when and how best to use this type of technology. Research is required to examine AutoTutor applications outside of its original intended use, which involved presentation of material followed by AutoTutor managed dialog covering key facts and objectives linked to the content. This same capability can be managed by GIFT, along with new mechanisms that have yet to be explored such as using AutoTutor as an intervention in a game when a learning event presents itself. Rather than give simple feedback when an error in performance is recognized, GIFT can pause the game and initiate a dialog so a learner can instantly reflect on the actions just experienced. Another avenue of research would be utilizing functions of AutoTutor to facilitate a comprehensive After Action Review based on performance within a training system.

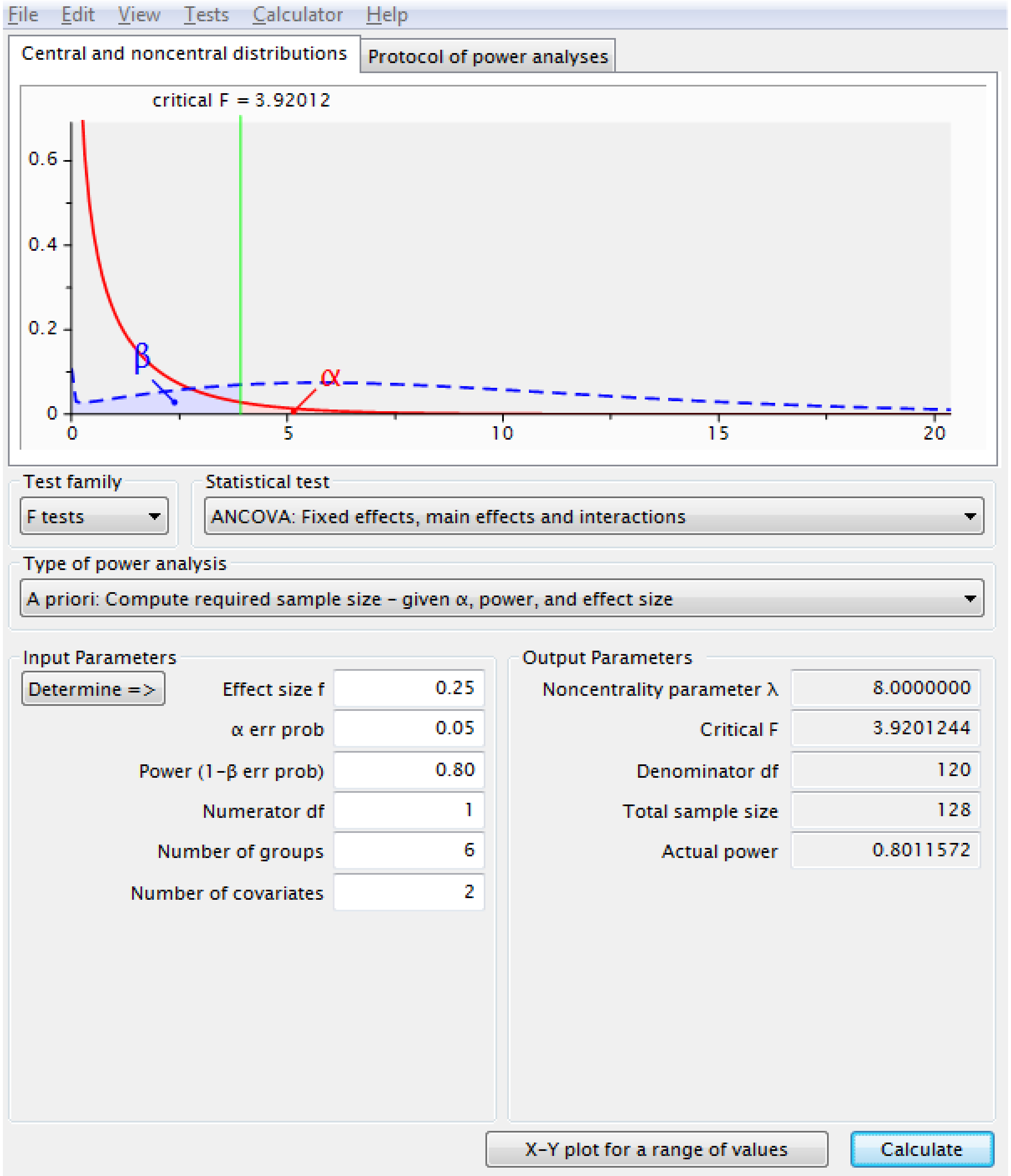
Conclusion

The aim of this research was to explore available tools for integrating intelligent tutoring communications in game-based learning platforms and to examine theory-based techniques for delivering explicit feedback in such environments. The primary tool influencing the design of this research was GIFT, a modular domain-independent framework that provides the tools and methods to author, deliver, and evaluate intelligent tutoring technologies within any training platform. Influenced by research surrounding SCT and CLT, the resulting experiment tested varying approaches for utilizing an EPA to function as a tutor during interaction in a game-based environment. Conditions were authored to assess the tradeoffs between embedding an EPA directly in the game environment, embedding an EPA in GIFT's browser-based TUI, or using audio prompts alone with no social grounding. Although not all predictions were supported by the resulting data, the application of using an EPA in the TUI to provide feedback during learning was found to be as effective as embedding the agent directly in the game environment.

This inference is based on evidence showing reliable differences across conditions on the metrics of performance and self-reported mental demand and feedback usefulness items. The overarching finding is that feedback, regardless of being delivered by an EPA, significantly improved performance in the training scenario. However, those assigned to an EPA condition were found to perform significantly better on transfer assessments when compared against subjects assigned to the audio alone condition (e.g. VoG). This finding supports previous research concerning the application of social agents in technology-based learning platforms. In addition, while using the TUI requires a game to be displayed in a windowed-mode, which was hypothesized to affect the level of immersion and mental demand a user experienced, data shows subjects to report the same level of flow and workload as those who interacted with an EPA directly in the game environment.

In conclusion, as the user community of GIFT increases with every version release, it is important system designers and developers are aware of the components available to them and the strengths/weaknesses they provide. More and more instructional designers are using serious games as domain practice environments, with a recognized need for identifying approaches to assist these games in facilitating the learning process while maintaining the benefit associated with their application. GIFT provides the tools to monitor performance in these environments in real-time, but no research was present for how best to interface communications back to the user based on performance outcomes. This research provides users with information linked to tactics for relaying training relevant explicit information to a user based on real-time performance that is most effective in terms of implementation requirements (i.e., cost and labor) and cognitive efficiency. Based on results from this research, GIFT provides a simplistic approach to include social EPAs as a communication mechanism for computer-based training applications.

APPENDIX A: POWER ANALYSIS WITH G*POWER3



Test family: F tests

Statistical test: ANCOVA: Fixed effects, main effects and interactions

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

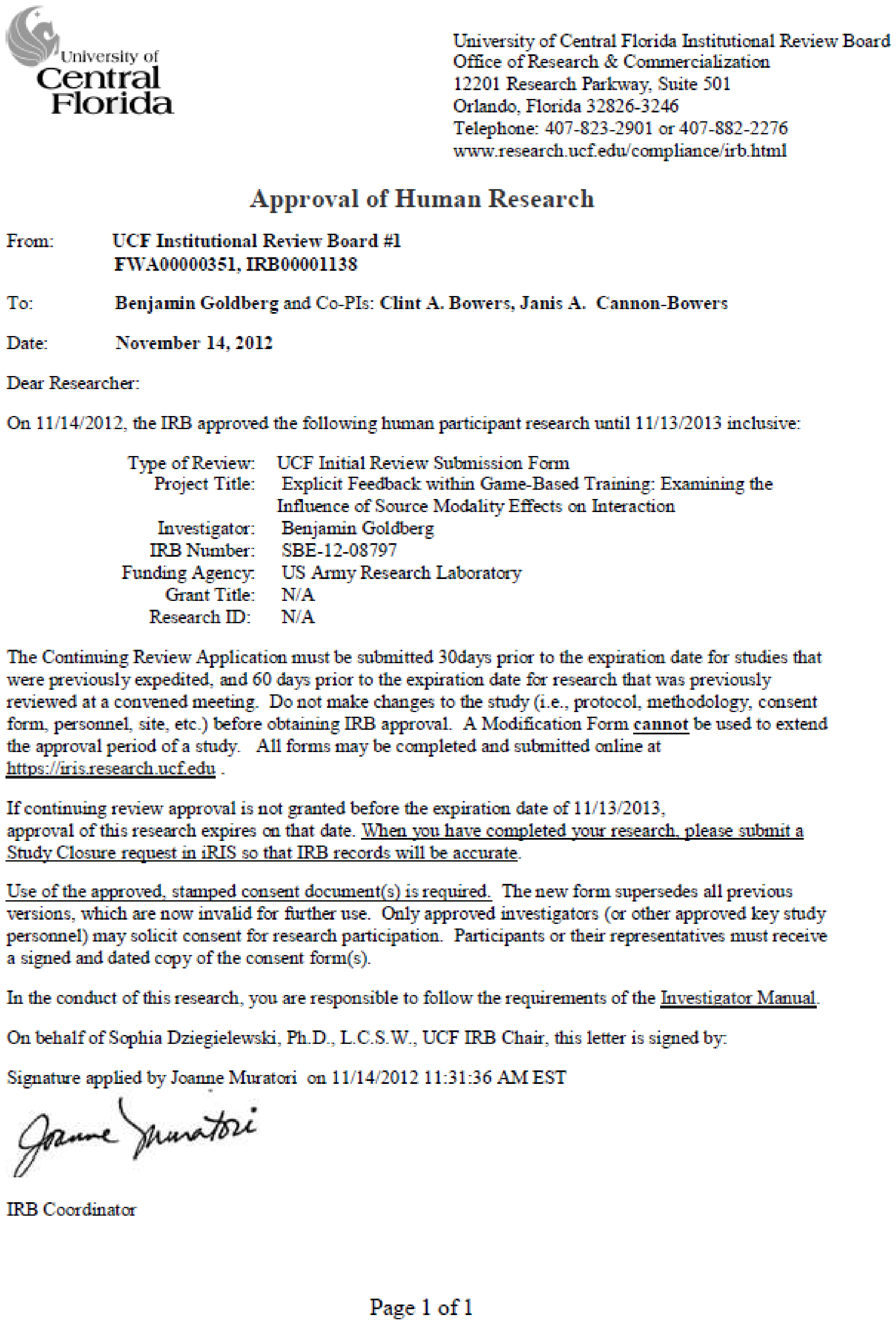
Input Parameters

Effect size f 0.25
α err prob 0.05
Power (1−β err prob) 0.80
Numerator df 1
Number of groups 6
Number of covariates 2

Output Parameters

Noncentrality parameter λ 8.0000000
Critical F 3.9201244
Denominator df 120
Total sample size 128
Actual power 0.8011572

APPENDIX B: IRB APPROVAL LETTERS



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Goldberg** and Co-PIs: **Clint A. Bowers, Janis A. Cannon-Bowers**

Date: **November 14, 2012**

Dear Researcher:

On 11/14/2012, the IRB approved the following human participant research until 11/13/2013 inclusive:

Type of Review:	UCF Initial Review Submission Form
Project Title:	Explicit Feedback within Game-Based Training: Examining the Influence of Source Modality Effects on Interaction
Investigator:	Benjamin Goldberg
IRB Number:	SBE-12-08797
Funding Agency:	US Army Research Laboratory
Grant Title:	N/A
Research ID:	N/A

The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 11/13/2013, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

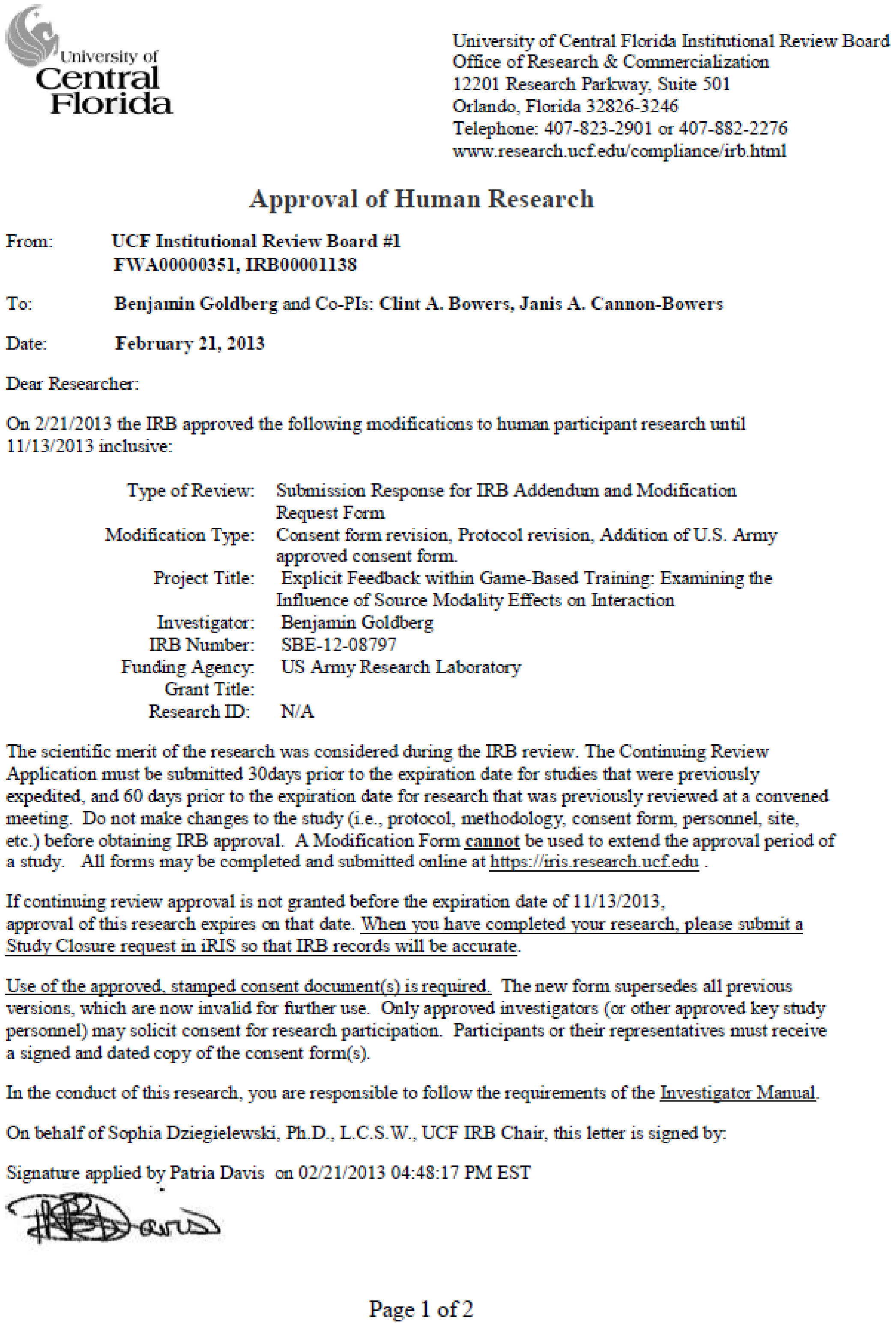
On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 11/14/2012 11:31:36 AM EST

Joanne Muratori

IRB Coordinator

Page 1 of 1



University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Approval of Human Research

From: **UCF Institutional Review Board #1**
FWA00000351, IRB00001138

To: **Benjamin Goldberg** and Co-PIs: **Clint A. Bowers, Janis A. Cannon-Bowers**

Date: **February 21, 2013**

Dear Researcher:

On 2/21/2013 the IRB approved the following modifications to human participant research until 11/13/2013 inclusive:

Type of Review:	Submission Response for IRB Addendum and Modification Request Form
Modification Type:	Consent form revision, Protocol revision, Addition of U.S. Army approved consent form.
Project Title:	Explicit Feedback within Game-Based Training: Examining the Influence of Source Modality Effects on Interaction
Investigator:	Benjamin Goldberg
IRB Number:	SBE-12-08797
Funding Agency:	US Army Research Laboratory
Grant Title:	
Research ID:	N/A

The scientific merit of the research was considered during the IRB review. The Continuing Review Application must be submitted 30days prior to the expiration date for studies that were previously expedited, and 60 days prior to the expiration date for research that was previously reviewed at a convened meeting. Do not make changes to the study (i.e., protocol, methodology, consent form, personnel, site, etc.) before obtaining IRB approval. A Modification Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at https://iris.research.ucf.edu .

If continuing review approval is not granted before the expiration date of 11/13/2013, approval of this research expires on that date. When you have completed your research, please submit a Study Closure request in iRIS so that IRB records will be accurate.

Use of the approved, stamped consent document(s) is required. The new form supersedes all previous versions, which are now invalid for further use. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Participants or their representatives must receive a signed and dated copy of the consent form(s).

In the conduct of this research, you are responsible to follow the requirements of the Investigator Manual.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Patria Davis on 02/21/2013 04:48:17 PM EST

Page 1 of 2

IRB Coordinator

Page 2 of 2

DEPARTMENT OF DEFENSE
UNITED STATES OF AMERICA

DEPARTMENT OF THE ARMY
U.S. ARMY MEDICAL DEPARTMENT ACTIVITY
WEST POINT, NEW YORK 10996-1197

REPLY TO
ATTENTION OF:

MCUD

8 November 2012

MEMORANDUM FOR Dr. Mike Matthews, United States Military Academy, Department of BS&L, West Point, New York 10996

SUBJECT: Initial IRB-approval for the KACH Protocol 13-004, IRBNet #382908-1, Explicit Feedback within Game-based Training: Examining the Influence of Source of Modality Effects on Interaction, PI: Mike Matthews, PhD

1. The subject protocol (version dated 16 October 2012), the Informed Consent (dated 16 October 2012) and the five Questionnaire/Surveys (versions dated 9 October 2012) was approved by the Chair of the Institutional Review Board (IRB) on 8 November 2012 via expedited review in accordance with 32 CFR 219.110 Federal Registry (FR) Category 4.

2. You may implement/start your study with the issuance of this KACH IRB approval letter, dated 8 November 2012. .

3. This minimal risk study is approved for a one-year period, 8 November 2012 – 7 November 2013. The study is approved for enrollment of up to 156 subjects.

4. In accordance with 32 CFR 219.109(e), the Principal Investigator must submit a continuing review report for this protocol to the KACH IRB. A continuing review report with a copy of the current protocol and informed consent must be submitted by 7 October 2013 to ensure approval on or before 7 November 2013.

5. The Principal Investigator is responsible for fulfilling reporting requirements to the KACH IRB.

MICHAEL F. DOYLE
COL, MC
Chair, KACH IRB

APPENDIX C: INFORMED CONSENTS

Explicit Feedback within Game-Based Training: Examining the Influence of Source Modality Effects on Interaction

Informed Consent

Principal Investigator(s):	Benjamin Goldberg Jan Cannon-Bowers, Ph.D. Clint Bowers, Ph.D.
Sub-Investigator(s):	Anne Sinatra, Ph.D.
Faculty Supervisor:	Clint Bowers, Ph.D.
Sponsor:	United States Army Research Laboratory—Human Research and Engineering Directorate—Simulation and Training Technology Center (ARL-HRED-STTC)
Investigational Site(s):	ARL-HRED-STTC's Learning and Intelligent Tutoring Environments Lab United States Military Academy at West Point (Approved USMA Informed Consent will be administered for all subjects at this investigational site)

Introduction: Researchers at the University of Central Florida (UCF) study many topics. To do this we need the help of people who agree to take part in a research study. You are being invited to take part in a research study which will include about 156 people at UCF and the United States Military Academy. You have been asked to take part in this research study because you are a student in a psychology course and a novice in casualty care procedures. You must be 18 years of age or older to be included in the research study.

The person doing this research is Benjamin Goldberg of the University of Central Florida and the U.S. Army Research Laboratory. Because the researcher is a graduate student, he is being guided by Dr. Clint Bowers, a UCF faculty supervisor in the psychology department.

University of Central Florida IRB
IRB NUMBER: SBE-12-08797
IRB APPROVAL DATE: 2/21/2013
IRB EXPIRATION DATE: 11/13/2013

What you should know about a research study:

- Someone will explain this research study to you.
- A research study is something you volunteer for.
- Whether or not you take part is up to you.
- You should take part in this study only because you want to.
- You can choose not to take part in the research study.
- You can agree to take part now and later change your mind.
- Whatever you decide it will not be held against you.
- Feel free to ask all the questions you want before you decide.

Purpose of the research study: The purpose of this study is to address how to best integrate feedback within game-based training events, and to determine how information delivered by embodied pedagogical agents (EPAs) in serious game environments (i.e., how the content is delivered) affects a user's performance and motivation/intention for future usage. Specifically, this research seeks to identify if embedding pedagogical agents for delivering feedback directly in a game-based environment improves training outcomes, reduces cognitive load required for interpreting information, and maintains a user's sense of presence in the virtual environment.

What you will be asked to do in the study:

- Upon arrival to the session, you will receive this informed consent which you will read. The purpose of the experiment will also be explained to you. You will be able to ask questions or voice any concerns you may have.
- Immediately after the consent process you will be fitted with the Q-Sensor, a wireless electrodermal activity sensor that wears just like a wristband.
- You will log-in to the Generalized Intelligent Framework for Tutoring (GIFT) system with a unique id assigned to you. You will complete a demographics questionnaire and the Immersive Tendencies Questionnaire (ITQ).
- You will complete a baseline knowledge assessment on hemorrhage control in the Tactical Combat Casualty Care (TC3) domain.
- You will complete a courseware package on hemorrhage control comprised of slides and video/audio presentations.
- You will interact with the game TC3Sim, a serious game used to train combat medics and combat lifesavers. You will first complete a tutorial scenario teaching the game controls. Following you complete two scenarios where you will be asked to apply the knowledge and skills presented in the courseware.
- You will complete a post-game knowledge assessment.
- You will receive additional information about the purpose of this study and have a chance to ask any questions you may have.

You do not have to answer every question or complete every task. You will not lose any benefits if you skip questions or tasks.

UCF
University of Central Florida IRB
IRB NUMBER: SBE-12-08797
IRB APPROVAL DATE: 2/21/2013
IRB EXPIRATION DATE: 11/13/2013

Location:

Learning in Intelligent Tutoring Environments (LITE) Lab at the Simulation and Training Technology Center (STTC) in UCF's Research Park. The address is 12423 Research Parkway, Orlando, FL 32826.

Time required:

We expect that you will be in this research study for 1 hour.

Funding for this study:

This research study is being paid for by the U.S Army Research Laboratory—Human Research and Engineering Directorate's LITE Lab.

Risks:

There are no reasonably foreseeable risks or discomforts involved in taking part in this study. Some of the courseware and game content is graphic in nature, and may be unsettling to some individuals. But this poses no more than minimal risk to participants. There is a small risk that people who take part will develop what is ordinarily referred to as simulator sickness. Side effects of VE (virtual environment) use may include stomach discomfort, headaches, sleepiness, dizziness and decreased balance. However, these risks are no greater than the sickness risks participants may be exposed to if they were to visit an amusement park such as Disney Quest (Disney Quest is a VE based theme park), Disney World or Universal Studios parks and ride attractions such as roller coasters. You will be given a 2 minute break during the exercise to lessen the chance that you will feel sick. If you experience any of the symptoms mentioned, please tell the researcher and remain seated until the symptoms disappear. Taking part in this research study may lead to added costs to you, such as transportation fees to and from the STTC facility.

Benefits:

We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include gaining insight and learning about research process and how psychological studies are conducted.

Alternatives:

If you choose not to participate, you may notify your instructor and ask for an alternative assignment of equal effort for equal credit. There will be no penalty.

Compensation or payment:

University of Central Florida IRB
IRB NUMBER: SBE-12-08797
IRB APPROVAL DATE: 2/21/2013
IRB EXPIRATION DATE: 11/13/2013

There is no direct compensation for taking part in this study. It is possible, however, that extra credit may be offered for your participation, but this benefit is at the discretion of your instructor. Participants may expect to spend 60 minutes performing experimental tasks, for which they will receive course extra credit for the amount of time they participate. Maximum course credit will be 90 minutes.

Anonymous research:

This study is anonymous. That means that no one, not even members of the research team, will know that the information you gave came from you.

Study contact for questions about the study or to report a problem: If you have questions, concerns, or complaints, or think the research has hurt you, talk to Janice Connor, Research Assistant, Learning in Intelligent Tutoring Environments Lab, Army Research Laboratory, (407) 208-3395 or Dr. Clint Bowers, Faculty Supervisor, Department of Psychology by email at clint.bowers@ucf.edu.

IRB contact about your rights in the study or to report a complaint: Research at the University of Central Florida involving human participants is carried out under the oversight of the Institutional Review Board (UCF IRB). This research has been reviewed and approved by the IRB. For information about the rights of people who take part in research, please contact: Institutional Review Board, University of Central Florida, Office of Research & Commercialization, 12201 Research Parkway, Suite 501, Orlando, FL 32826-3246 or by telephone at (407) 823-2901. You may also talk to them for any of the following:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You want to get information or provide input about this research.

If you are harmed because you take part in this study:

If you are made sick from taking part in this research study, medical care will be provided. Depending on the circumstances, this care may be provided at no cost to you. Contact the investigator for more information.

Withdrawing from the study:

If you decide to leave the research, you should tell the principle investigator as soon as possible. There will be no consequence for withdrawing from the study at any point, and extra credit will be given for the time spent. Withdrawing from the study at any time in no way brings risks to you losing your right to medical care. The person in charge of the research study or the sponsor can remove you from the research study without your approval. Possible reasons for removal include behavior deemed inappropriate for a lab environment and a lack of participant cooperation.

UCF
University of Central Florida IRB
IRB NUMBER: SBE-12-08797
IRB APPROVAL DATE: 2/21/2013
IRB EXPIRATION DATE: 11/13/2013

Your signature below indicates your permission to take part in this research.

DO NOT SIGN THIS FORM AFTER THE IRB EXPIRATION DATE BELOW

Name of participant

Signature of participant	Date
Signature of person obtaining consent	Date

Printed name of person obtaining consent

UCF
University of Central Florida IRB
IRB NUMBER: SBE-12-08797
IRB APPROVAL DATE: 2/21/2013
IRB EXPIRATION DATE: 11/13/2013

KELLER ARMY COMMUNITY HOSPITAL
IRB APPROVED
NOV 0 8 2012
APPROVAL EXPIRES
NOV 0 7 2013

UNITED STATES MILITARY ACADEMY
CONSENT TO PARTICIPATE IN RESEARCH

Study Title: **EXPLICIT FEEDBACK WITHIN GAME-BASED TRAINING: EXAMINING THE INFLUENCE OF SOURCE MODALITY EFFECTS ON INTERACTION**

You are asked to participate in a research study conducted at the United States Military Academy at West Point by Dr. Mike Matthews and Benjamin Goldberg. Your participation in this study is voluntary. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

PURPOSE OF THE STUDY

The purpose of this study is to address how to best integrate feedback within game-based training events, and to determine how information delivered by Virtual Humans in serious game environments affects a user's performance and motivation for future usage. Specifically, this research seeks to identify if embedding feedback directly in a game-based environment improves training outcomes.

EXPECTED DURATION OF PARTICIPATION

We expect that you will be in this research study for 1 hour.

PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Upon arrival to the session, you will receive this informed consent which you will read. The purpose of the experiment will also be explained to you. You will be able to ask questions or voice any concerns you may have.
- Immediately after the consent process you will be fitted with the Q-Sensor, a wireless electrodermal activity sensor that wears just like a wristband.
- You will log-in into the Generalized Intelligent Framework for Tutoring (GIFT) system with a unique id assigned to you. You will complete a demographics questionnaire and the Immersive Tendencies Questionnaire (ITQ).
- You will complete a baseline knowledge assessment on hemorrhage control in the Tactical Combat Casualty Care (TC3) domain.
- You will complete a courseware package on hemorrhage control comprised of slides and video/audio presentations.
- You will interact with the game TC3Sim, a serious game used to train combat medics and combat lifesavers. You will first complete a tutorial scenario teaching the game controls.

Following you complete two scenarios where you will be asked to apply the knowledge and skills presented in the courseware.

- You will complete a post-training knowledge assessment.
- You will receive additional information about the purpose of this study and have a chance to ask any questions you may have.

POTENTIAL RISKS AND DISCOMFORTS

There are no reasonably foreseeable risks or discomforts involved in taking part in this study. Some of the courseware and game content is graphic in nature, and may be unsettling to some individuals. But this poses no more than minimal risk to participants. There is a small risk that people who take part will develop what is ordinarily referred to as simulator sickness. Side effects of virtual environment use may include stomach discomfort, headaches, sleepiness, dizziness and decreased balance. However, these risks are no greater than the sickness risks participants may be exposed to if they were to visit an amusement park such as Disney Quest (Disney Quest is a virtual environment based theme park), Disney World or Universal Studios parks and ride attractions such as roller coasters. You will be given a 2 minute break during the exercise to lessen the chance that you will feel sick. If you experience any of the symptoms mentioned, please tell the researcher and remain seated until the symptoms disappear. Taking part in this research study may lead to added costs to you, such as transportation fees to and from the STTC facility.

ANTICIPATED BENEFITS

We cannot promise any benefits to you or others from your taking part in this research. However, possible benefits include gaining insight and learning about research process and how psychological studies are conducted.

MEDICAL CARE FOR RESEARCH RELATED INJURY

If you are made sick or have an adverse reaction from taking part in this research study, medical care will be provided. Depending on the circumstances, this care may be provided at no cost to you. Contact the investigator for more information.

CONFIDENTIALITY

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity.

This study is anonymous. You will be assigned a random participant number when you arrive. That means that no one, not even members of the research team, will know that the information you gave came from you. During the study, all personal data or information (such as informed consents and demographic data) will either be stored on a password protected laptop or secured under lock and key until destroyed. Any subject identification keys used during the study will be destroyed at the end of the study.

Authorized representatives of the U.S. Army Human Research Protection Office may need to review records of individual subjects. They may or may not see your identifiable information, if collected, but they are bound by rules of confidentiality not to reveal your identity to others.

COMPENSATION FOR PARTICIPATION

Participants may expect to spend 60 minutes performing experimental tasks, for which they will receive course extra credit for the amount of time they participate. This will result in 1.0 SONA credits when entered in by your associated psychology professor.

Payment Details (if applicable): If you decide to withdraw from this study at any time, you will receive credit for the time spent.

PARTICIPATION AND WITHDRAWAL BY YOU

Your participation in this research is voluntary. If you choose not to participate, that will not affect your relationship with investigators, the United States Military Academy or your right to health care or other benefits or services to which you are otherwise entitled. If you decide to participate, you are free to withdraw your consent and discontinue participation at any time without prejudice.

WITHDRAWAL OF PARTICIPATION BY THE INVESTIGATOR

The investigator may withdraw you from participating in this research if circumstances arise which warrant doing so. The investigator will make the decision and let you know if it is not possible for you to continue. The decision may be made either to protect your health and safety, or because it is part of the research.

NEW FINDINGS

During the course of the study, you will be informed of any significant new findings (either good or bad), such as changes in the risks or benefits resulting from participation in the research or new alternatives to participation, that might cause you to change your mind about continuing in the study. If new information is provided to you, your consent to continue participating in this study will be re-obtained.

POINTS OF CONTACT

In the event of a research related injury or if you experience an adverse reaction, immediately contact the following:

COL James Ness, 845-938-0239

If you have specific questions about the conduct of the research, please contact one of the investigators listed below.

Dr. Mike Matthews, Professor of Engineering Psychology
U.S. Military Academy
West Point, NY 10996
1-845-938-3696

If you have any questions about your rights as a volunteer in the research, please feel free to contact the KACH Human Protections Administrator at (845) 938-0761 or go to the USMA Human Research Protection website.

http://www.usma.edu/opa/hrpp/hrpp_home.html .

RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study.

SIGNATURE OF RESEARCH SUBJECT

I have read the information provided above. I have been given an opportunity to ask questions and all of my questions have been answered to my satisfaction. I have been given a copy of this form.

____
Name of Subject

____
Signature of Subject Date

____
Address

SIGNATURE OF WITNESS

My signature as witness certifies that the subject signed this consent form in my presence as his/her voluntary act and deed.

____
Name of Witness

____
Signature of Witness Date (same as subject's)

APPENDIX D: DEMOGRAPHICS SURVEY

Demographics Questionnaire

Participant ID: ___________________

1. What is your:
    Age_______
    Gender M F

2. Have you ever been in the military? Yes No

    If yes:
    Military Rank/Grade __________________
    Status (AD, Res, Ret) _______
    Primary MOS & description____
    Total Time in Service______years _______months

3. What is your class year?
    ____ Freshman - major____________
    ____ Sophmore - major____________
    ____ Junior - major_____________
    ____ Senior - major_____________

4. How much sleep did you get last night?
    ____

5. Do you have normal or corrected-to-normal vision:
    ____ Normal
    ____ Corrected (**Circle One**: glasses / contacts)
    ____ Problems
        Please describe____________________

6. What is your present level of energy? (1 through 5 with 1 = low and 5 = high)
    ____

7. What is your level of confidence in using a computer? (1 through 5 with 1 = low and 5 = high)
    ____

8. How would you describe your general level of gaming experience (i.e., playing video games)?
    ________ ***None*** (I have never played a video game).
    ________ ***Low*** (I have played a video game a few times in the past).
    ________ ***Moderately Low*** (I have played a video game a regularly in the past).

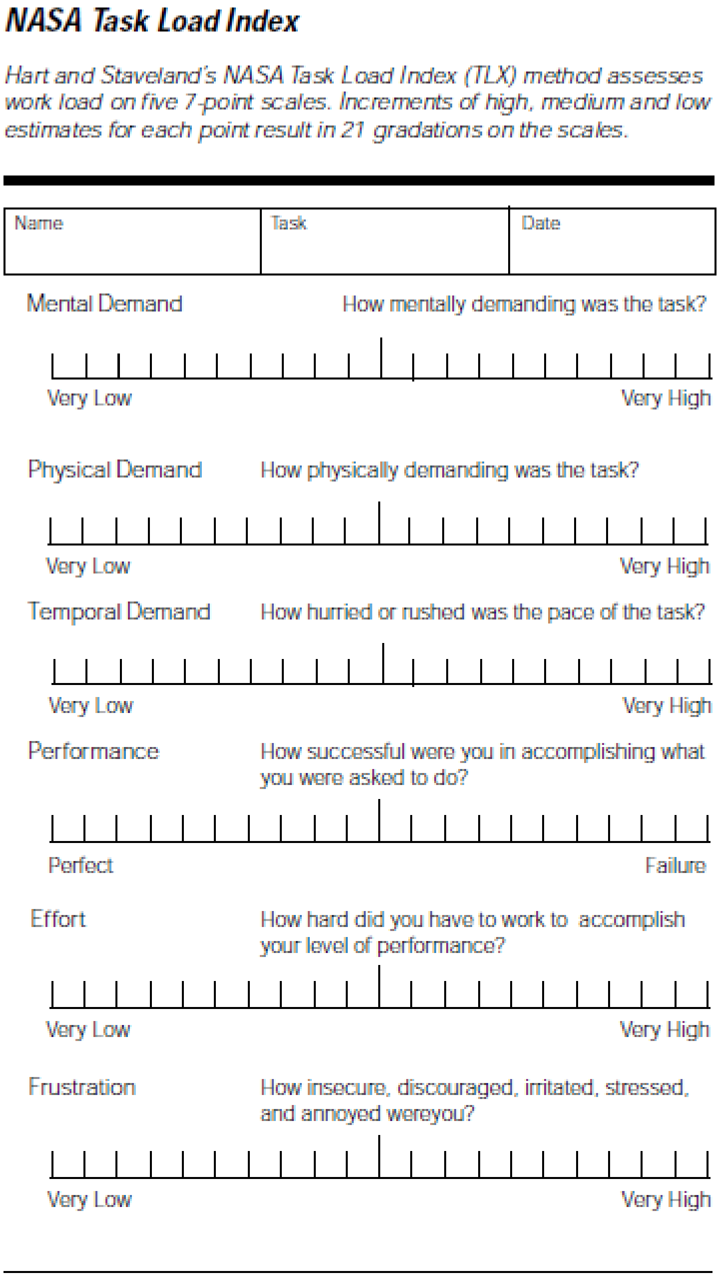
________ ***Moderately High*** (I currently play video games weekly).
________ ***High*** (I currently play video games daily).
________ ***Other*** (please explain) ____

9. Have you ever taken courses on First Aid and/or CPR?
Yes ___/No___ (If yes, please specify):____________

10. How would you rate your knowledge of First Aid? (1 through 5 with 1 = low and 5 = high)
____

11. How would you describe your skill level in performing First Aid procedures?
________ Novice
________ Experienced
________ Expert
________ Other (please explain) ____

APPENDIX E: NASA-TLX INSTRUMENT



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

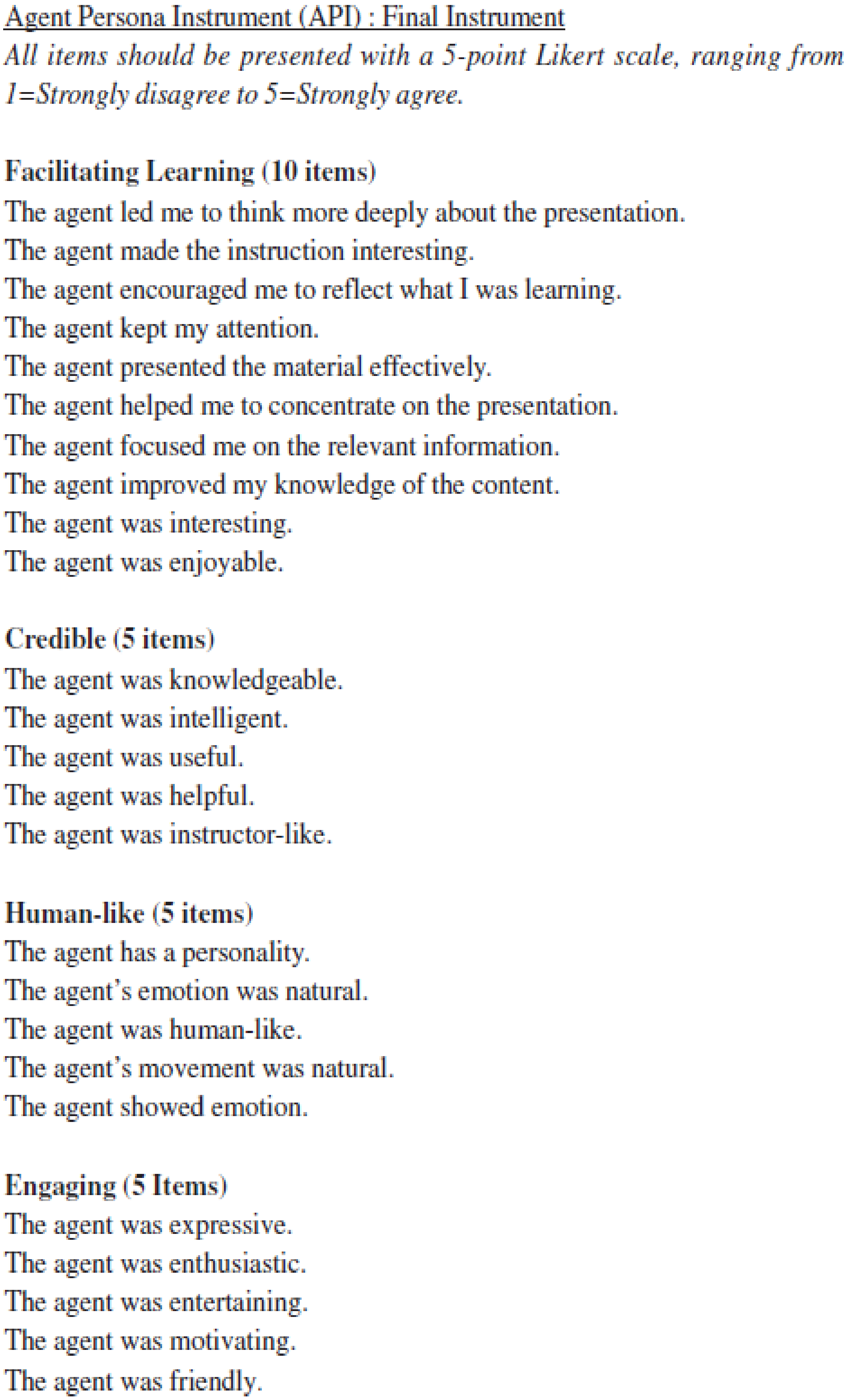
Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

APPENDIX F: AGENT PERSONA INSTRUMENT



Agent Persona Instrument (API) : Final Instrument

All items should be presented with a 5-point Likert scale, ranging from 1=Strongly disagree to 5=Strongly agree.

Facilitating Learning (10 items)

The agent led me to think more deeply about the presentation.
The agent made the instruction interesting.
The agent encouraged me to reflect what I was learning.
The agent kept my attention.
The agent presented the material effectively.
The agent helped me to concentrate on the presentation.
The agent focused me on the relevant information.
The agent improved my knowledge of the content.
The agent was interesting.
The agent was enjoyable.

Credible (5 items)

The agent was knowledgeable.
The agent was intelligent.
The agent was useful.
The agent was helpful.
The agent was instructor-like.

Human-like (5 items)

The agent has a personality.
The agent's emotion was natural.
The agent was human-like.
The agent's movement was natural.
The agent showed emotion.

Engaging (5 Items)

The agent was expressive.
The agent was enthusiastic.
The agent was entertaining.
The agent was motivating.
The agent was friendly.

APPENDIX G: IMMERSIVE TENDENCIES QUESTIONNAIRE

IMMERSIVE TENDENCIES QUESTIONNAIRE

(Witmer & Singer, Version 3.01)

Indicate your preferred answer by circling the number in the appropriate box of the seven point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.

1. Do you easily become deeply involved in movies or tv dramas?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

2. Do you ever become so involved in a television program or book that people have problems getting your attention?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

3. How mentally alert do you feel at the present time?

1	2	3	4	5	6	7
NOT ALERT			MODERATELY		FULLY ALERT	

4. Do you ever become so involved in a movie that you are not aware of things happening around you?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

5. How frequently do you find yourself closely identifying with the characters in a story line?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

7. What kind of books do you read most frequently? (CIRCLE ONE ITEM ONLY!)

Spy novels
Adventure novels
Westerns
Biographies

Fantasies
Romance novels
Mysteries
Autobiographies

Science fiction
Historical novels
Other fiction
Other non-fiction

8. How physically fit do you feel today?

1	2	3	4	5	6	7
NOT FIT			MODERATELY FIT			EXTREMELY FIT

9. How good are you at blocking out external distractions when you are involved in something?

1	2	3	4	5	6	7
NOT VERY GOOD			SOMEWHAT GOOD			VERY GOOD

10. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

11. Do you ever become so involved in a daydream that you are not aware of things happening around you?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

12. Do you ever have dreams that are so real that you feel disoriented when you awake?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

13. When playing sports, do you become so involved in the game that you lose track of time?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

14. How well do you concentrate on enjoyable activities?

1	2	3	4	5	6	7
NOT AT ALL			MODERATELY WELL			VERY WELL

15. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

16. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

17. Have you ever gotten scared by something happening on a TV show or in a movie?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

18. Have you ever remained apprehensive or fearful long after watching a scary movie?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

19. Do you ever become so involved in doing something that you lose all track of time?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

20. On average, how many books do you read for enjoyment in a month?

1	2	3	4	5	6	7
NONE	ONE	TWO	THREE	FOUR	FIVE	MORE

21. Do you ever get involved in projects or tasks, to the exclusion of other activities?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

22. How easily can you switch attention from the activity in which you are currently involved to a new and completely different activity?

1	2	3	4	5	6	7
NOT SO EASILY			FAIRLY EASILY			QUITE EASILY

23. How often do you try new restaurants or new foods when presented with the opportunity?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			FREQUENTLY

24. How frequently do you volunteer to serve on committees, planning groups, or other civic or social groups?

1	2	3	4	5	6	7
NEVER			SOMETIMES			FREQUENTLY

25. How often do you try new things or seek out new experiences?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

26. Given the opportunity, would you travel to a country with a different culture and a different language?

1	2	3	4	5	6	7
NEVER			MAYBE			ABSOLUTELY

27. Do you go on carnival rides or participate in other leisure activities (horse back riding, bungee jumping, snow skiing, water sports) for the excitement of thrills that they provide?

1	2	3	4	5	6	7
NEVER			OCCASIONALLY			OFTEN

28. How well do you concentrate on disagreeable tasks?

1	2	3	4	5	6	7
NOT AT ALL			MODERATELY WELL			VERY WELL

29. How often do you play games on computers?

1	2	3	4	5	6	7
NOT AT ALL			OCCASIONALLY			FREQUENTLY

30. How many different video, computer, or arcade games have you become reasonably good at playing?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|
NONE ONE TWO THREE FOUR FIVE SIX OR MORE

31. Have you ever felt completely caught up in an experience, aware of everything going on and completely open to all of it?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|
NEVER OCCASIONALLY FREQUENTLY

32. Have you ever felt completely focused on something, so wrapped up in that one activity that nothing could distract you?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|
NOT AT ALL OCCASIONALLY FREQUENTLY

33. How frequently do you get emotionally involved (angry, sad, or happy) in news stories that you see, read, or hear?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|
NEVER OCCASIONALLY OFTEN

34. Are you easily distracted when involved in an activity or working on a task?

|___1___|___2___|___3___|___4___|___5___|___6___|___7___|
NEVER OCCASIONALLY OFTEN

APPENDIX H: RETRO-FLOW SCALE

RETRO Flow Scale Items and Scoring with Subscales Color-Coded

Please answer the following questions by selecting the *most appropriate rating. Ratings are as follows:*

Never	Rarely	Sometimes	Frequently	Always
1	***2***	***3***	***4***	***5***
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

#	Item text						Source with subscale	RETRO Flow Subscale
1	Generally speaking, I can remain concentrated in the game.	**1**	**2**	**3**	**4**	**5**	EGame - Concentration	Concentration
2	Overall game goals were presented in the beginning of the game.	**1**	**2**	**3**	**4**	**5**	EGame – Goal Clarity	Mastery
3	I received feedback on my progress in the game.	**1**	**2**	**3**	**4**	**5**	EGame - Feedback	Feedback
4	The difficulty of challenges increased as my skills improved.	**1**	**2**	**3**	**4**	**5**	EGame - Challenge	Mastery
5	I lost track of time while I was playing the game.	**1**	**2**	**3**	**4**	**5**	Tychen Response – Time Distortion	Temporal Dissociation
6	The visual effects of the game allowed me to feel like I was part of the game and not just playing it.	**1**	**2**	**3**	**4**	**5**	Presense – Interface Qual	Visual Quality
7	The game was challenging, but I felt that I could meet that challenge.	**1**	**2**	**3**	**4**	**5**	DFS-2 - Balance	Mastery
8	I felt a sense of control over the game.	**1**	**2**	**3**	**4**	**5**	EGame - Autonomy	Mastery
9	I forget about time passing while playing the game.	**1**	**2**	**3**	**4**	**5**	EGame - Immersion	Temporal Dissociation
10	I played without thinking about how to play.	**1**	**2**	**3**	**4**	**5**	GEQ - Flow	Action-Awareness
11	I felt involved with the game.	**1**	**2**	**3**	**4**	**5**	EGame - Immersion	Action-Awareness

12	I became unaware of my surroundings while playing the game.	**1**	**2**	**3**	**4**	**5**	EGame - Immersion	Loss of SC
13	*The visual display quality interfered with me being able to get into the game.*	***1***	***2***	***3***	***4***	***5***	*Presense – Interface Qual*	*Visual Quality*
14	I knew what to do next in the game.	**1**	**2**	**3**	**4**	**5**	EGame - Autonomy	Mastery
15	My experience makes me want to play this game again.	**1**	**2**	**3**	**4**	**5**	DFS - Autotelic	Autotelic
16	I became involved in the game.	**1**	**2**	**3**	**4**	**5**	EGame - Immersion	Action-Awareness
17	I felt emotionally involved in the game.	**1**	**2**	**3**	**4**	**5**	EGame - Immersion	Action-Awareness
18	It was effortless to concentrate on the game -- I didn't even know I was so focused.	**1**	**2**	**3**	**4**	**5**		Concentration
19	I felt like I just couldn’t stop playing.	**1**	**2**	**3**	**4**	**5**	GEQ - Flow	Autotelic
20	I felt as if I were part of the game.	**1**	**2**	**3**	**4**	**5**	Refiana - Involvement	Loss of SC
21	I really enjoyed the experience.	**1**	**2**	**3**	**4**	**5**	DFS-2 - Autotetlic	Autotelic
22	I felt viscerally involved in the game.	**1**	**2**	**3**	**4**	**5**	GEQ - Immersion	Action-Awareness
23	The gaming equipment allowed me to play without interfering with my focus.	**1**	**2**	**3**	**4**	**5**	Presence – Interface Qual	Concentration
24	I was completely into the game, like I was experiencing it instead of playing it.	**1**	**2**	**3**	**4**	**5**	GEQ - Immersion	Loss of SC
25	If someone talked to me while I was playing, I probably would not have heard them.	**1**	**2**	**3**	**4**	**5**	GEQ - Flow	Concentration
26	I played longer than I meant to.	**1**	**2**	**3**	**4**	**5**	GEQ - Presence	Temporal Distortion
27	I lost track of where I was.	**1**	**2**	**3**	**4**	**5**	GEQ - Absorption	Loss of SC
28	I did things spontaneously and automatically without having to think.	**1**	**2**	**3**	**4**	**5**	DFS-2 - Merging	Action-Awareness
29	I received information on my success	**1**	**2**	**3**	**4**	**5**	GEQ -	Feedback

	(or failure) of intermediate goals immediately.						Feedback	
30	I loved the feeling of the performance and want to capture it again.	**1**	**2**	**3**	**4**	**5**	DFS-2 - Autotelic	Autotelic
31	*I felt the control devices interfered with my performance.*	***1***	***2***	***3***	***4***	***5***	*Presence – Interface Qual*	*Concentration*
32	I was absorbed in what I was doing while playing the game.	**1**	**2**	**3**	**4**	**5**	Tychen – Focused Imm	Action-Awareness
33	I knew how well I was playing the game.	**1**	**2**	**3**	**4**	**5**	FSS - Feedback	Feedback
34	Playing seemed automatic.	**1**	**2**	**3**	**4**	**5**	GEQ - Flow	Mastery
35	I learned new techniques that enabled me to improve my performance.	**1**	**2**	**3**	**4**	**5**	Mastery	

APPENDIX I: KNOWLEDGE PRE-TEST

#	**Question**	**Responses**	**Correct Answer**
1	What are the three most common medically preventable causes of death on the modern battlefield?	1-extremity hemorrhage, tension pneumothorax, airway obstruction 2-extremity hemorrhage, tension pneumothorax, gunshot wound 3-amputation of a limb, tension pneumothorax, gunshot wound 4-amputation of a limb, infection, airway obstruction	1
2	Pulse can be used to indicate the extent of blood loss	1-True 2-False	1
3	You are providing care under fire to a casualty. Which of the following actions can be performed before moving the casualty to a safe location?	1-Open the casualty's airway (head-tilt/chin-lift). 2-Perform needle chest decompression. 3-Apply a tourniquet to a limb with severe bleeding from a wound. 4-Insert a nasopharyngeal airway. 5-All listed actions can be performed before moving the casualty to a safe location.	3
4	A soldier has just had his forearm amputated slightly above the wrist. The bleeding from the amputation site is not severe. What should you do first?	1-Apply an Emergency Bandage to the wound. 2-Apply a tourniquet two inches above the amputation site. 3-Apply a pressure dressing to the stump. 4-Apply a tourniquet two inches above the elbow.	2
5	You are going to the aid of an injured soldier while under fire. What should be your first action upon reaching the soldier?	1-Check the soldier for responsiveness 2-Check the soldier's pulse 3-Check the soldier for breathing 4-Check the soldier for shock	1
6	You can move a casualty out of enemy fire and to a safe location. Should you also try to move the casualty's weapon to the safe location?	1-Yes 2-No	1
7	You have been wounded and are still under enemy fire. You are unable to return fire and there is no safe cover nearby. What should you do?	1-Call for help 2-Play dead	2
8	You are going to apply a tourniquet to an amputation that is about one inch below the elbow joint. Which of the following is an appropriate site for the tourniquet band?	1-Between the wound and the elbow. 2-Directly over the elbow. 3-A little above the elbow. 4-Two inches distal to the shoulder joint.	3
9	Hemorrhage control is the most important aspect of saving lives during Care Under Fire phase for what reasons?	1-A Soldier can go into shock and die quickly after injuring a large blood vessel 2-Hemorrhage is the easiest thing to treat on the battlefield 3-Hemorrhage is the leading cause of preventable death in combat 4-Hemorrhage rarely leads to infection	1 & 3

#	Question	Responses	Correct Answer
10	A soldier in your squad has been injured. You are in a tactical field care situation. When should you notify your unit leader of the soldier's injury?	1-As soon as you can 2-Only after you have performed a full examination of the casualty 3-Only after you have completed your treatment of the casualty 4-Only if the casualty requires evacuation	1
11	Which of the following statements are true? (Select all that apply)	1-Do not attempt to salvage a casualty's rucksack, unless it's critical to the mission 2-Always attempt to salvage a casualty's rucksack 3-Don't waste time taking a casualty's weapon and ammunition 4-Take the casualty's weapon and ammunition if possible	1 & 4
12	You applied a tourniquet to a soldier about 30 minutes ago, while under fire, in order to stop the bleeding from a serious wound on the soldier's forearm. The casualty and you have now reached a safe location. Which of the following statements is correct?	1-You can now safely remove the tourniquet. 2-You can now reevaluate the casualty's wound to see if other measures, such as a pressure dressing, would be more appropriate. 3-You cannot remove a tourniquet once it has been applied.	2
13	What has historically been a problem with requests for medical evacuation?	1-Proper classification. 2-Over classification. 3-Priority classification. 4-Routine classification.	2
14	You applied a tourniquet to a soldier about eight hours ago. The tactical situation now allows the casualty to be evacuated. Should you loosen the tourniquet and try to control the bleeding with a pressure dressing before evacuating the casualty?	1-Yes 2-No	2
15	You have controlled the bleeding from a wound on the casualty's thigh. The casualty lost a good deal of blood. Also, the casualty's skin appears to be pale, cool, and clammy. He is breathing faster than normal and he is acting agitated. The casualty is probably suffering from:	1-Blocked Airway. 2-Cardiac arrest. 3-Hypothermia. 4-Shock.	4

APPENDIX J: KNOWLEDGE POST-TEST

#	Question	Responses	Correct Answer
1	Which of the following is NOT part of care under fire?	1-Moving the casualty to safety 2-Checking the casualty's level of consciousness 3-Treating an open chest wound 4-Applying a tourniquet	3
2	Which of the following statements are true about "Care Under Fire"? (Select all that apply)	1-Medics should expect to return fire in a combat situation 2-Casualties should return fire if able 3-Airway management should be administered 4-Medics should direct the casualty to move to cover and apply self aid if able	1 & 2 & 4
3	Blood sweeps are performed prior to measuring blood pressure or taking the casualty's pulse.	1-True 2-False	1
4	The band of a Combat Application Tourniquet is being applied to a severely bleeding wound on the casualty's arm. Where should the tourniquet band be placed?	1-Six inches above the wound. 2-Two inches above the wound. 3-Directly over the wound. 4-Two inches below the wound. 5-Six inches below the wound.	2
5	Which of the following describes a combat lifesaver?	1-A nonmedical soldier who provides lifesaving measures as his primary mission. 2-A nonmedical soldier who provides lifesaving measures as his secondary mission. 3-A medical soldier who provides lifesaving measures as his primary mission. 4-A medical soldier who provides lifesaving measures as his secondary mission.	2
6	When should you plan how to move a wounded soldier out of enemy fire?	1-Before you leave your place of safety, to go to the wounded soldier 2-As soon as you reach the wounded soldier 3-As soon as you have treated the life-threatening conditions 4-As soon as you have treated all of the casualty's injuries	1
7	You and another soldier are in the open and separated when you both come under enemy fire. The other soldier is wounded, but is conscious and able to fire his weapon. What should you tell him to do?	1-Seek cover, return fire, play dead 2-Seek cover, return fire, administer self-aid 3-Play dead 4-Seek cover, return fire, administer buddy-aid	2
8	Which one of the following statements gives a proper rule for tightening a tourniquet?	1-A tourniquet should be loose enough so that you can slip two fingers under the tourniquet band. 2-A tourniquet should be loose enough so that you can slip the tip of one finger under the tourniquet band. 3-A tourniquet is to be tightened until the bright red bleeding has stopped and the distal pulse is gone; darker blood oozing from the wound can be ignored. 4-A tourniquet is to be tightened until both the bright red bleeding and the darker venous bleeding have stopped completely and the distal pulse is gone.	3

#	Question	Responses	Correct Answer
9	Once you have tightened an improvised tourniquet, you must:	1-Secure the windlass so that the tourniquet will not unwind. 2-Apply an Emergency Bandage over the windlass. 3-Remove the windlass and tie the tails in a nonslip knot.	1
10	How long can you leave a tourniquet on without having to worry about the loss of a limb?	1-10 Minutes 2-30 Minutes 3-1 Hour 4-2 Hours 5-5 Hours	5
11	How does evaluation and treatment of a casualty in a tactical field care situation (not under enemy fire) differ from that in a care under fire situation?	1-None of the below. 2-A tactical field care environment allows you to 3-focus more on the evaluation, treatment and evacuation of the casualty. 4-A tactical field care environment limits you to only to the treatment of life-threatening bleeding from a limb and movement to safety.	2
12	You are going to request medical evacuation. What should you say to notify the person receiving the message that you are going to make a MEDEVAC request?	1-Roger, Roger, I have a request for evacuation. Over. 2-Please dispatch (an air) (a ground) ambulance to the following location. (State location.) 3-I require medical assistance ASAP. Over. 4-I have a MEDEVAC request. Over.	4
13	You are crossing a battlefield after the fighting has stopped and the enemy has retreated. A soldier steps on a land mine and it explodes, giving the soldier a severe wound in his thigh. What type of care will you render to the soldier?	1-Tactical evacuation care 2-Tactical field care 3-Care under fire	2
14	You applied a tourniquet to a soldier's wounded leg before dragging him to a safe location. What should you do about the tourniquet once you and the casualty are safe?	1-Nothing. Leave the tourniquet in place 2-Examine the wound to see if it is bleeding and can be controlled using other means 3-Place another tourniquet above the first tourniquet and leave both tourniquets in place 4-Place another tourniquet above the first tourniquet and remove the first tourniquet	2
15	You have treated a soldier for wounds on his arms and have controlled the bleeding. The casualty remains conscious and is lying on his back. However, the casualty has developed sweaty and clammy skin, his breathing rate has become rapid, his lips look bluish, and his level of consciousness is decreasing. What should you do?	1-Flex the casualty's knees so that they are raised and his feet are flat on the ground. 2-Place a nasopharyngeal airway in each nostril. 3-Place a field pack or other object under his feet so that the feet are elevated slightly above the level of his heart. 4-Have the casualty drink a full canteen of warm, salted water.	3

APPENDIX K: TC3SIM MISSION BRIEFING SCRIPT

Mission Briefing

Task: Your Unit will be patrolling the main streets of Shakarat today. You are the acting medic for this unit. Navigate to the central village market. Your squad leader is tasked with locating Jamail, the village elder, to discuss opportunities for local support and humanitarian aid. Intel reports possible insurgent activity in the surrounding buildings, so keep your eyes peeled.

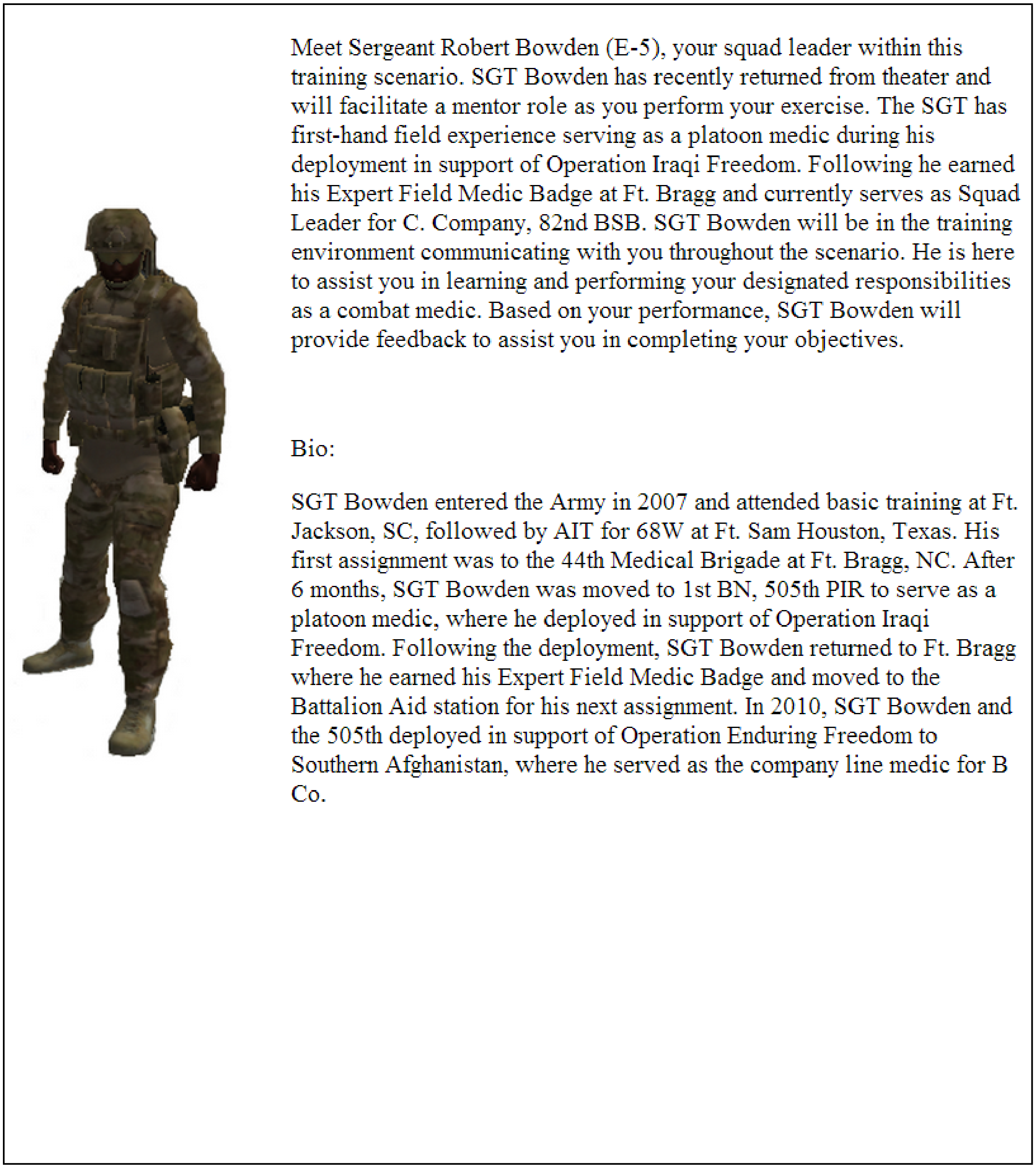
Conditions: A squad size plus element of ACF is suspected of operating in the village of Shakarat. This element consists of dismounted insurgents equipped with an assortment of small arms, including AK-47s, RPKs, and cell phone triggered IEDs. The neighborhood adjacent to the market has been a hot bed for insurgent activity over the past week, making it difficult for locals to purchase goods and services available from vendors, as well as causing them to shy away from American Forces. Your unit is to secure the market area while discussions are conducted on how to improve safety. You are to react as necessary to hostile contact. Current Rules of Engagement are in effect. If engaged, you are to perform all tasks associated with a combat medic’s role.

Standards:

1. Maintain situational awareness and keep a close proximity to your unit
2. Secure market and react to hostile personnel
3. Perform proper control of weapon and obey rules of engagement
4. Apply proper techniques of Care Under Fire and Tactical Field Care when appropriate
    a. Hemorrhage Control
    b. Casualty Movement
    c. Airway Management and Breathing
    d. Bleeding Control
5. Perform MEDEVAC procedures if required

APPENDIX L: EPA PROFILE BACKGROUNDS/BIOS

TC3Sim Embedded Peer Tutor Profile



Meet Sergeant Robert Bowden (E-5), your squad leader within this training scenario. SGT Bowden has recently returned from theater and will facilitate a mentor role as you perform your exercise. The SGT has first-hand field experience serving as a platoon medic during his deployment in support of Operation Iraqi Freedom. Following he earned his Expert Field Medic Badge at Ft. Bragg and currently serves as Squad Leader for C. Company, 82nd BSB. SGT Bowden will be in the training environment communicating with you throughout the scenario. He is here to assist you in learning and performing your designated responsibilities as a combat medic. Based on your performance, SGT Bowden will provide feedback to assist you in completing your objectives.

Bio:

SGT Bowden entered the Army in 2007 and attended basic training at Ft. Jackson, SC, followed by AIT for 68W at Ft. Sam Houston, Texas. His first assignment was to the 44th Medical Brigade at Ft. Bragg, NC. After 6 months, SGT Bowden was moved to 1st BN, 505th PIR to serve as a platoon medic, where he deployed in support of Operation Iraqi Freedom. Following the deployment, SGT Bowden returned to Ft. Bragg where he earned his Expert Field Medic Badge and moved to the Battalion Aid station for his next assignment. In 2010, SGT Bowden and the 505th deployed in support of Operation Enduring Freedom to Southern Afghanistan, where he served as the company line medic for B Co.

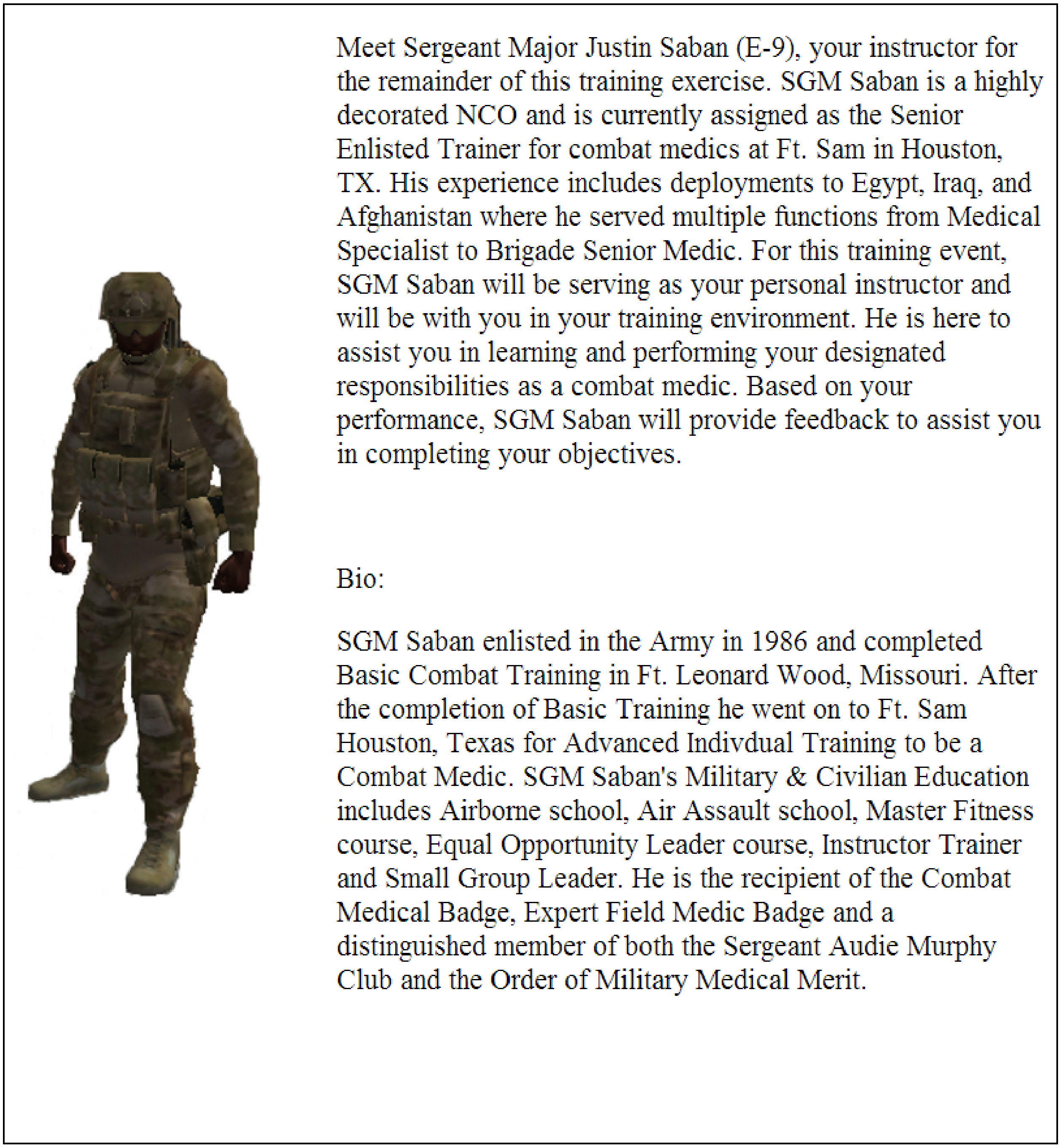
TUI Embedded Peer Tutor Profile

Meet Sergeant Robert Bowden (E-5), your squad leader within this training scenario. SGT Bowden has recently returned from theater and will facilitate a mentor role as you perform your exercise. The SGT has first-hand field experience serving as a platoon medic during his deployment in support of Operation Iraqi Freedom. Following he earned his Expert Field Medic Badge at Ft. Bragg and currently serves as Squad Leader for C. Company, 82nd BSB. SGT Bowden will be in the training environment communicating with you throughout the scenario. He is here to assist you in learning and performing your designated responsibilities as a combat medic. Based on your performance, SGT Bowden will provide feedback to assist you in completing your objectives.

Bio:

SGT Bowden entered the Army in 2007 and attended basic training at Ft. Jackson, SC, followed by AIT for 68W at Ft. Sam Houston, Texas. His first assignment was to the 44th Medical Brigade at Ft. Bragg, NC. After 6 months, SGT Bowden was moved to 1st BN, 505th PIR to serve as a platoon medic, where he deployed in support of Operation Iraqi Freedom. Following the deployment, SGT Bowden returned to Ft. Bragg where he earned his Expert Field Medic Badge and moved to the Battalion Aid station for his next assignment. In 2010, SGT Bowden and the 505th deployed in support of Operation Enduring Freedom to Southern Afghanistan, where he served as the company line medic for B Co.

TC3Sim Embedded Instructor Tutor Profile

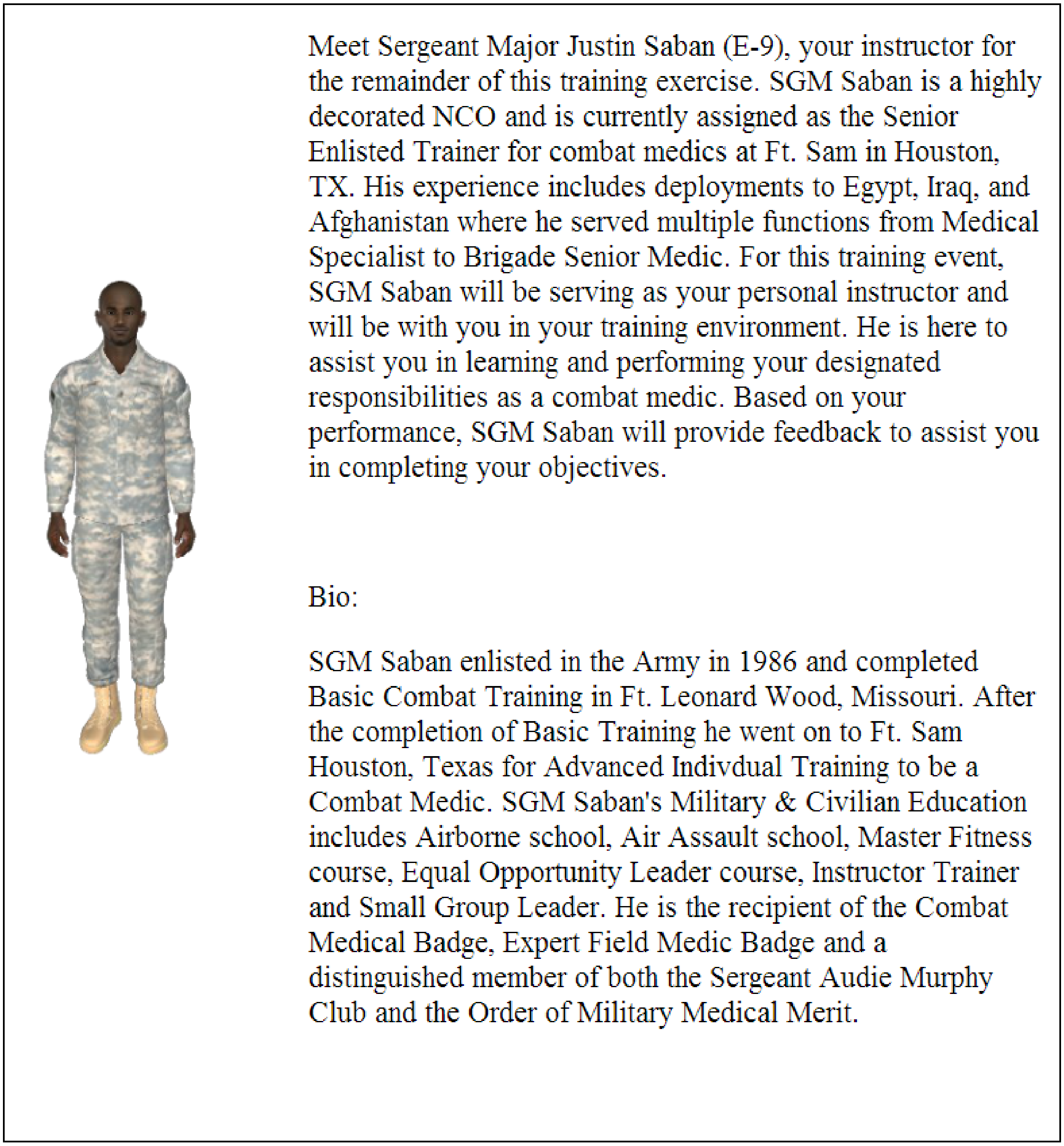


Meet Sergeant Major Justin Saban (E-9), your instructor for the remainder of this training exercise. SGM Saban is a highly decorated NCO and is currently assigned as the Senior Enlisted Trainer for combat medics at Ft. Sam in Houston, TX. His experience includes deployments to Egypt, Iraq, and Afghanistan where he served multiple functions from Medical Specialist to Brigade Senior Medic. For this training event, SGM Saban will be serving as your personal instructor and will be with you in your training environment. He is here to assist you in learning and performing your designated responsibilities as a combat medic. Based on your performance, SGM Saban will provide feedback to assist you in completing your objectives.

Bio:

SGM Saban enlisted in the Army in 1986 and completed Basic Combat Training in Ft. Leonard Wood, Missouri. After the completion of Basic Training he went on to Ft. Sam Houston, Texas for Advanced Indivdual Training to be a Combat Medic. SGM Saban's Military & Civilian Education includes Airborne school, Air Assault school, Master Fitness course, Equal Opportunity Leader course, Instructor Trainer and Small Group Leader. He is the recipient of the Combat Medical Badge, Expert Field Medic Badge and a distinguished member of both the Sergeant Audie Murphy Club and the Order of Military Medical Merit.

TC3Sim Embedded Instructor Tutor Profile



Meet Sergeant Major Justin Saban (E-9), your instructor for the remainder of this training exercise. SGM Saban is a highly decorated NCO and is currently assigned as the Senior Enlisted Trainer for combat medics at Ft. Sam in Houston, TX. His experience includes deployments to Egypt, Iraq, and Afghanistan where he served multiple functions from Medical Specialist to Brigade Senior Medic. For this training event, SGM Saban will be serving as your personal instructor and will be with you in your training environment. He is here to assist you in learning and performing your designated responsibilities as a combat medic. Based on your performance, SGM Saban will provide feedback to assist you in completing your objectives.

Bio:

SGM Saban enlisted in the Army in 1986 and completed Basic Combat Training in Ft. Leonard Wood, Missouri. After the completion of Basic Training he went on to Ft. Sam Houston, Texas for Advanced Indivdual Training to be a Combat Medic. SGM Saban's Military & Civilian Education includes Airborne school, Air Assault school, Master Fitness course, Equal Opportunity Leader course, Instructor Trainer and Small Group Leader. He is the recipient of the Combat Medical Badge, Expert Field Medic Badge and a distinguished member of both the Sergeant Audie Murphy Club and the Order of Military Medical Merit.

APPENDIX M: GIFT CONCEPTS AND SIMILE RULE CONDITIONS FOR TC3SIM SCENARIOS

	Simulation Activity	**Concepts**	**Concept ID**	**SIMILE Rule Condition (Pseudo-code)**	**Non-player Character Hints**
1	Exercise begins with the medic moving with their unit down the road of a village	The medic needs to be moving with their unit	stay_with_ unit	If any member of the unit and the medic player are more than 20 meters apart for 30 seconds then indicate that the stay_with_unit concept is below threshold.	Tutor says "you are part of a unit. You need to stay close to them."
		Medic needs to stay out of middle of street	move_ under_ cover	If the medic player and any polygons to the left or right is more than 2 meters but less than 10 meters then trigger that the move_under_cover concept is below threshold.	Tutor says "You are too exposed, get closer to the buildings and stay out of sight."
2	The unit moves to where the road makes a T with a large building at the end. In front of the medic an IED goes off wounding one of the medic's unit and signaling "Care Under Fire"	The medic should be seeking cover	seeking_ cover	If there are enemies firing weapons indicate that the medic is "under fire". If the medic player is "under fire" and there are no polygons between any enemy entity and the medic that are 2 meters from the medic for 30 seconds then indicate that the seeking_cover concept is below threshold.	Tutor yells at medic: "we're under attack, seek cover!"
3	A member of the unit goes out to attempt to drag the other casualty and is shot by the ensuing gunfire from the roof of the facing building	The medic should be returning fire	return_ fire	If the medic is "under fire" for 5 seconds and the medic is not firing their weapon then indicate that return_fire concept is below threshold.	Tutor yells at the medic "we need to neutralize the enemy forces. Return fire!"

	Simulation Activity	**Concepts**	**Concept ID**	**SIMILE Rule Condition (Pseudo-code)**	**Non-player Character Hints**
4	The medic continues with suppressive fire but also addresses the casaulties	The medic should be asking the casualty where they are hurt	communicate_with_casualty	If the medic is under fire for 15 seconds and has not used the communicate interaction "where are you hurt" then report the communicate_with_casualty concept is below threshold.	Tutor says "if you cannot reach the wounded, then communicate with them. Yell out some questions!"
5	The medic requests covering fire to move and help the most severe IED victim wounded with an amputation	The medic should be requesting cover fire	request_cover	If the firefight has gone on for 30 seconds and the medic has not moved to within 1 meter of the amputee then report the request_cover concept is below threshold	Tutor yells at the medic, "Do you need cover to go get him?"
6	The firefight continues for 10 seconds	The medic should apply tourniquet to amputee	apply_tourniquet	If the firefight has gone on for **30 seconds** and the medic has not used the apply tourniquet interaction on the amputee then report the apply_tourniquet concept is below threshold.	Tutor says "apply tourniquets to an amputation immediately. We only have a couple minutes before a bleed out"
7		The medic should move the amputee to a safe zone	move_to_save_zone	If the firefight has gone on for 45 seconds and the amputee is not in the "safe zone" then report the move_to_safe_zone concept is below threshold.	Tutor yells at the medic, "Get out of there before you get shot! Seek cover and get to the safe zone"
8	The firefight ends transitioning to "Tactical Field Care"	The medic should request a security sweep	request_security_sweep	If the firefight ends then indicate that the medic is now in "Tactical Field Care" If in "Tactical Field Care" for **20 seconds** and the medic has not used the "request security sweep" interaction then report the request_security_sweep concept is below threshold	Tutor yells at the medic, "Should I make sure the area is secure?"

	Simulation Activity	**Concepts**	**Concept ID**	**SIMILE Rule Condition (Pseudo-code)**	**Non-player Character Hints**
9		The medic should bandage the gunshot wound	apply_ bandage	If in "Tactical Field Care" for **25 seconds** and the medic has not used the "apply bandage" interaction on the gunshot wound then report the apply_bandage concept is below threshold	Casualty yells at the medic, "I've been shot! Are you going to do anything?" Tutor says "Do not forget to address your other wounded. A gunshot wound can result in severe blood loss and should be attended to"
10		The medic should roll the casualty to over to check for an exit wound	check_exit _wound	If in "Tactical Field Care" for **30 seconds** and the medic has not used the "roll" interaction on the gunshot wound then report the check_exit_wound	Tutor says "Always check for an exit wound. Overlooking another wound can lead to severe loss of blood."
11		The medic should move the gunshot wound to a safe location	move_ casualties_ to_ safety	If in "Tactical Field Care" for **60 seconds** and gunshot wound is not in the casualty collection point then report the move_casualties_to_safety concept is below threshold	Tutor yells at the medic, "Bad guys could still be around, Once their critical injury has been addressed move your casualties to a safe location."
12		The medic should request help to move the wounded to the collection point	request_ help	If in "Tactical Field Care" for **60 seconds** and the "request help" interaction is not used then report the request_help concept is below threshold	Tutor team asks, "Doc, do you need help moving these guys?"
13		The medic should remove the tourniquet on the amputee	reassess_in juries	If in "Tactical Field Care" for **90 seconds** and the "remove tourniquet" interaction has not been used on the amputee then report the reassess concept is below threshold	Tutor says "In tactical field care, you should reassess injuries and application of tourniquets"

	Simulation Activity	**Concepts**	**Concept ID**	**SIMILE Rule Condition (Pseudo-code)**	**Non-player Character Hints**
14		The medic should expose the amputated wound	expose_wound	If in "Tactical Field Care" for **90 seconds** and the "expose wound" interaction has not been used on the amputee then indicate that the expose_wound concept is below threshold	Tutor says "Remove clothing from a wound so that it does not contribute to infection."
15		The medic should apply another tourniquet	reapply_tourniquet	If the medic has used the "remove tourniquet" interaction more than 30 seconds ago indicate that the reapply_tourniquet concept is below theshold	Tutor says "If you remove a tourniquet and the bleeding has not been controlled by other means you will need to re-apply it."
16		The medic should apply a bandage	amputee_apply_bandage	If in "Tactical Field Care" and the apply_bandage interaction has not been used then report the amputee_apply_bandage concept is below threshold	Tutor says "to avoid infection and further blood loss, make sure to bandage all exposed wounds"
17		The medic should check the vitals of the amputee	amputee_check_vitals	If in "Tactical Field Care" for **180 seconds** and the "check airway", "check carotid pulse", "check distal pulse", "check blood pressure", "check breathing" interactions have been used then report the amputee_check_vitals concept is below threshold	Tutor says "check his vitals once bleeding is controlled. Low blood volume can result in shock."
18		If the blood volume of the amputee is low, the medic should administer hextend	amputee_administer_hextend	If the amputee's blood volume is below 2000 and the medic has not used the administer hextend interaction then report the amputee_administer_hextend concept is below threshold	Tutor says "Administer a hextend to increase blood volume and use something to warm him up"
19		The medic should check the vitals of the bullet wound	bulletwound_check_vitals	If in "Tactical Field Care" for **180 seconds** and the "check airway", "check carotid pulse", "check distal pulse", "check blood pressure", "check breathing" interactions have been used then report the bulletwound_check_vitals concept is below threshold	Tutor says "examine the vitals of all casualties. Make sure blood loss in controlled!"

	Simulation Activity	**Concepts**	**Concept ID**	**SIMILE Rule Condition (Pseudo-code)**	**Non-player Character Hints**
20		The medic should perform a blood sweep on the gunshot wound	bulletwound_blood_sweep	If in "Tactical Field Care" for **300 seconds** and the "blood sweep" interaction has not been used on the gunshot wound then report the bulletwound_blood_sweep concept is below threshold	Tutor says "Perform a bloodsweep to assure there are no wounds overlooked"
21		The medic should check the breathing of the bullet wound	bulletwound_check_breathing	If in "Tactical Field Care" for **360 seconds** and the "check breathing" interaction has not been used then report the bulletwound_check_breathing concept is below threshold	Tutor says "be sure to check the breathing among those who've been shot. Tension pneumothorax can result from a punctured lung"
22		The medic should request CASEVAC and fill out 9-line	request_casevac	If in "Tactical Field Care" for **480 seconds** and the "request CASEVAC" interaction has not been used then report the "request_casevac" concept is below threshold	Tutor asks, "Are you ready for me to call in that 9-line? Request CASEVAC once they are stable"

REFERENCES

(CALL), C. f. A. L. L. (2006). Handbook 06-18 Tactical Combat Casualty Care Handbook.

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*.

Ammar, M. B., Neji, M., Alimi, A. M., & Gouarderes, G. (2010). The Affective Tutoring System. *Expert Systems with Applications, 37,* 3013-3023.

Amory, A., Naicker, K., Vincent, J., & Adams, C. (1999). The use of computer games as an educational tool: Identification of appropriate game types and game elements. *British Journal of Educational Technology*, *30*(4), 311-321.

Anderson, D. I., Magill, R. A., & Sekiya, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of motor behavior, 33*(1), 59-66.

André, E., & Rist, T. (2001). Presenting through performing: on the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems, 14*(1), 3-13.

Andre, T. (1997). Selected micro instructional methods to facilitate knowledge construction: implications for instructional design. In R. D. Tennyson & F. Schott (Eds.), *Instructional Design: International Perspectives: Theory, Research, and Models* (Vol. 1, pp. 243-267). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Apt, C. (1970). *Serious Games*. New York: The Viking Press.

Army, D. o. t. (2009). Medical Education and Demonstration of Individual Competence (MEDIC) *Training Circular*. Washington, D.C.

Army, U. S. (2010). Combat Lifesaver Course: Student Self-Study *Army Correspondence Course Program* (C ed.).

Ashford, S. J. (1986). Feedback-seeking in individual adaptation: A resource perspective. *Academy of Management Journal*, 465-487.

Astwood, R. S., Van Buskirk, W. L., Cornejo, J. M., & Dalton, J. (2008). *The impact of different feedback types on decision-making in simulation based training environments*.

Ayres, P., & Gog, T. (2009). State of the art research into Cognitive Load Theory. *Computers in Human Behavior, 25*(2), 253-257.

Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research, 13*(2), 111-127.

Baddeley, A. (1990). *Human Memory: Theory and Practice.* Boston, Massachussetts: Allyn and Bacon.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4*(11), 417-423.

Baddeley, A. D. (2004). The psychology of memory. *The essential handbook of memory disorders for clinicians*, 1-13.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-89.

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction. *PRESENCE: Teleoperators and Virtual Environments*, *15*(4), 359-372.

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). *Detecting student misuse of intelligent tutoring systems.* Paper presented at the 7th International Conference on Intelligent Tutoring Systems, Brazil.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, N.J.: Prentice Hall.

Bandura, A. (2011). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology, 52*, 1 – 26.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Bannert, M. (2002). Managing cognitive load-recent trends in cognitive load theory. *Learning and Instruction, 12*(1), 139-146.

Barad, J. (2010). Games for Health 2010: Military TC3 Simulator Trains, Entertains Medics. *Military Medicine* Retrieved August 2, 2012, from http://medgadget.com/2010/05/games_for_health_2010_military_tc3_simulator_trains_entertains_medics.html

Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(1), 95-115.

Beal, C., & Lee, H. (2005). *Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction*.

Beck, J., Stern, M., & Haugsjaa, E. (1996). Applications of AI in Education. *ACM Crossroads, 3*(1), 11-15.

Bell, B. S., Kanar, A. M., & Kozlowski, S. W. J. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management, 19*(8), 1416-1434.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., . . . Olmstead, R. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction, 17*(2), 151-170.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., . . . Craven, P. L. (2007). EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation Space and Environmental Medicine, 78*(5), B231-B244.

Billings, D. R. (2010). *Adaptive Feedback in Simulation-Based Training*. Doctor of Philosophy, University of Central Florida.

Bilodeau, E. A. (1969). *Principles of Skill Acquisition*. New York: Academic Press.

Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: McGraw-Hill.

Bloom, B. S. (1984). The 2-Sigma Problem: The search for methods of group instruction as effective as one-to-one tutoring,. *Educational Researcher, 13*(6), 4-16.

Boulay, B. d., & Luckin, R. (2001). Modelling Human Teaching Tactics and Strategies for Tutoring Systems. *International Journal of Artificial Intelligence in Education, 12*, 235-256.

Brackbill, Y., Bravos, A., & Starr, R. H. (1962). Delay-improved retention of a difficult task. *Journal of Comparative and Physiological Psychology; Journal of Comparative and Physiological Psychology, 55*(6), 947.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press.

Bratt, E. O. (2009). Intelligent Tutoring for Ill-Defined Domains in Military Simulation-Based Training. *International Journal of Artificial Intelligence in Education, 19*, 337-356.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*, 624-634.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology, 42*(3), 361-377.

Bruner, J. S. (1966). Some Elements of Discovery. In L. S. Shulman & E. R. Keislar (Eds.), *Learning by Discovery: A Critical Appraisal* (pp. 101-113). Chicago, IL: Rand McNally.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53-61.

Burke, A. J. (1983). *Students' potential for learning contrasted under tutorial and group approaches to instruction.* University of Chicago.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Buttimer, J. (2003). *Research Review: Multiple Resource Theory*.

Cannon-Bowers, J., & Bowers, C. A. (2008a). Synthetic Learning Environments. In J. M. Spector, M. D. Merrill, J. v. Merrienboer & M. P. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 317-327). New York, NY: Lawrence Erlbaum Publishers.

Cannon-Bowers, J. A., & Bowers, C. A. (2008b). Synthetic Learning Environments: On Developing a Science of Simulation, Games, and Virtual Worlds for Training. In S. W. J. Kozlowski & E. Salas (Eds.), *Learning, training, and development in organizations*. Mahwah, NJ: Erlbaum.

Cannon-Bowers, J. A., Prince, C., Salas, E., Owens, J. M., Morgan, B. B., & Gonos, G. H. (1989). *Determining aircrew coordination training effectiveness*. Paper presented at the Interservice/Industry Training Systems Conference, Fort Worth, TX.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*(1), 113-117.

Carroll, M., Kokini, C., Champney, R., Fuchs, S., Sottilare, R., & Goldberg, B. (2011). Modeling Trainee Affective and Cognitive State Using Low Cost Sensors *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Orlando, FL.

Cassell, J., Sullivan, J., Prevost, S., & Churchhill, E. (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471-533.

Clariana, R. B. (1999). Differential Memory Effects for Immediate and Delayed Feedback: A Delta Rule Explanation of Feedback Timing Effects.

Clarke, D., & Duimering, P. R. (2006). How computer gamers experience the game situation: a behavioral study. *Computers in Entertainment (CIE), 4*(3), 6.

Conati, C., Gertner, A. S., VanLehn, K., & Druzdzel, M. J. (1997). On-line student modeling for coached problem solving using Bayesian networks. *COURSES AND LECTURES-INTERNATIONAL CENTRE FOR MECHANICAL SCIENCES*, 231-242.

Conkey, C. A. (2011). *Machinima and video-based soft skills training.* University of Central Florida.

Corbett, A. T., & Anderson, J. R. (2001). *Locus of Feedback Control in Computer-Based Tutoring: Impact on Learning Rate, Achievement and Attitudes*. Paper presented at the ACM CHI'2001 Conference on Human Factors in Computing Systems, New York.

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction, 2nd edition* (pp. 849-874). Amsterdam, The Netherlands: Elsevier Science.

Csikszenmihayi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper-Row.

Csikszenmihayi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*: Basic Books.

Cuevas, H. M., Fiore, S. M., Bowers, C. A., & Salas, E. (2004). Fostering constructive cognitive and metacognitive activity in computer-based complex task training environments. *Computers in Human Behavior, 2006*, 225-241.

Davis, W. D., Carson, C. M., Ammerter, A. P., & Treadway, D. C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *Human Performance, 18*, 409-426.

Delgado, K. M. (2005). *Interactive effects of feedback type and feedback propensities on task performance.* Unpublished master's thesis, Wright State University, Ohio.

Dorneich, M., Whitlow, S., Ververs, P. M., Carciofini, J., & Creaser, J. (2004). *Closing the loop of an adaptive system with cognitive state.* Paper presented at the International Conference of the Human Factors and Ergonomics Society (HFES).

Durlach, P. J., & Ray, J. M. (2011). Designing Adaptive Instructional Environments: Insights from Empirical Evidence: U.S. Army Research Institute for the Behavioral and Social Sciences.

Dyer, R. F., Matthews, J. J., Wright, C. E., & Yudowitch, K. L. (1976). Questionnaire construction manual: DTIC Document.

ECS. (2012a). SIMILE Retrieved July 30, 2012, from http://www.ecsorl.com/products/simile

ECS. (2012b). vMedic Retrieved August 5, 2012, from http://www.ecsorl.com/products/vmedic

Eggemeier, F. T., Crabtree, M. S., & LaPointe, P. A. (1983). *The effect of delayed report on subjective ratings of mental workload*.

Ericsson, K. A., & Ward, P. (2007). Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory: Toward a Science of Expert and Exceptional Performance. *Current Directions in Psychological Science, 16*(6), 346-350.

Estes, J. (1979). What’s In It For Me?”: Over, Under, and Around, Using a Computerized Business Simulation,”. *Journal of Experiential Learning and Simulation, 1*(1), 65-89.

Farmer, E., & Brownson, A. (2003). Review of Workload Measurement, Analysis and Interpretation Methods. In R. Gingell (Ed.).

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Field, A. (2009). *Discovering Statistics Using SPSS*. Thousand Oaks, CA: Sage Publications Ltd.

Fitts, P. M. (1962). Factors in complex skill training. In R. Glaser (Ed.), *Training Research and Education*. Oxford, England: University of Pittsburgh Press.

Fong, G. (2004). *Adapting COTS games for military simulation*.

Fournier, P. S., Montreuil, S., Brun, J. P., Bilodeau, C., & Villa, J. (2011). Exploratory Study to Identify Workload Factors that Have an Impact on Health and Safety A Case Study in the Service Sector. Université Laval: IRSST.

Fowler, S., Smith, B., & Litteral, C. S. M. D. J. (2005). *A TC3 game-based simulation for combat medic training*.

Fowlkes, J., Dickinson, S., & Lazarus, T. (2010). Blended Training for Combat Medics.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59-109.

Fu, F., Su, R., & Yu, S. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101-112.

Galvao, J. R., Martins, P. G., & Gomes, M. R. (2000, December 10-13). *Modeling realty with simulation games for a cooperative learning.* Paper presented at the 2000 Winter Simulation Conference, Orlando, FL.

Games, P. A. (1984). Data, transformations, power, and skew: A rebuttal to Levine and Dunlap. *Psychological Bulletin, 95*(2), 345-347.

Games, P. A., & Lucas, P. A. (1966). Power of the Analysis of Variance of Independent Groups on Non-Normal and Normally Transformed Data. *Educational and Psychological Measurement*, *26*(2), 311-327.

Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by computer-assisted instruction. *Journal of Educational Psychology; Journal of Educational Psychology, 60*(6p1), 503.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313-331.

Goldberg, B., Brawner, K. W., Sottilare, R., Tarr, R., Billings, D. R., & Malone, N. (2012). *Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012, Orlando, FL.

Goldberg, B. S., Holden, H. K., Brawner, K. W., & Sottilare, R. A. (2011). *Enhancing Performance through Pedagogy and Feedback: Domain Considerations for Intelligent Tutoring Systems*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2011, Orlando, FL.

Goldberg, B. S., Sottilare, R. A., Brawner, K. W., & Holden, H. K. (2011). Predicting Learner Engagement during Well-Defined and Ill-Defined Computer-Based Intercultural Interactions. In S. D. Mello, A. Graesser, B. Schuller & J.-C. Martin (Eds.), *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011), LNCS* (Vol. 6974, pp. 538-547). Berlin Heidelberg: Springer-Verlag.

Gonzalez, C. (2005). Decision support for real-time, dynamic decision-making tasks. *Organizational Behavior and Human Decision Processes, 96*(2), 142-154.

González, C., Burguillo, J. C., & Llamas, M. (2006). *A Qualitative Comparison of Techniques for Student Modeling in Intelligent Tutoring Systems*.

Graesser, A., & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45*(4), 234-244.

Graesser, A. C., Person, N. K., Harter, D., & Group, T. R. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12*(3), 257-279.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine, 22*(4), 39.

Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive Science Implications for Enhancing Training Effectiveness in a Serious Gaming Context. *ACM Journal of Educational Resources in Computing, 7*(3).

Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. *International Journal of Artificial Intelligence in Education, 14*(3, 4), 313-334.

Hanna, G. S. (1976). Effects of total and partial feedback in multiple-choice testing upon learning. *The Journal of Educational Research*, 202-205.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload, 1*, 139-183.

Hartog, C. (2009). *Scenario Design for Serious Gaming: Guiding Principles for the Design of Scenarios and Curricula in Military Job Oriented Training.* M.S., Delft University of Technology.

Hays, M., Lane, H. C., Auerbach, D., Core, M. G., Gomboc, D., & Rosenberg, M. (2009). *Feedback specificity and the learning of intercultural communication skills*.

Hays, R. T. (2005). The effectiveness of instructional games: A literature review and discussion: DTIC Document.

Heeter, C. (1992). Being there: The subjective experience of presence. *PRESENCE: Teleoperators and Virtual Environments, 1*(2), 262-271.

Hembree, R. (1988). Correlates, Causes, Effects, and Treatment of Test Anxiety. *Review of Educational Research, 58*(1), 47-77.

Hermann, G. (1969). Learning by Discovery: A Critical Review of Studies. *Journal of Experimental Education, 38*(1), 58-72.

Heylen, D., Nijholt, A., R., o. d. A., & Vissers, M. (2003). Socially Intelligent Tutor Agents. In T. Rist, R. Aylett, D. Ballin & J. Rickel (Eds.), *Proceedings of Intelligent Virtual Agents*

(IVA 2003), Lecture Notes in Computer Science (Vol. 2792, pp. 341-347). Berlin: Springer.

Höök, K., Persson, P., & Sjölinder, M. (2000). Evaluating users' experience of a character-enhanced information space. *AI Communications, 13*(3), 195-212.

ICT. (2013). Virtual Human Toolkit Retrieved May 20, 2013, from https://confluence.ict.usc.edu/display/VHTK/Home

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology, 64*(4), 349.

Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and Tracking Difficulty: Evidence For Multiple Resources in Dual-Task Performance. *Psychophysiology, 17*(3), 259-273.

Jackson, S. A., & Eklund, R. C. (2004). *The flow scales manual*. Morgantown, WV: Fitness Information Technology.

JISC. (2007). Game-Based Learning: JISC Retrieved July 25, 2012, from http://www.jisc.ac.uk/publications/briefingpapers/2007/pub_gamebasedlearningBP/pub_gamebasedlearningBP_content.aspx

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Active Learning: Cooperation in the College Classroom*. Edina, MN: Interaction Book Company.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*(1), 47-78.

Jones, D. M. (2002). *The 7±2 urban legend*.

Kalyuga, S. (2009). *Cognitive Load Factors in Instructional Design for Advanced Learners*. New York: Nova Science Publishers, Inc.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction.

Kassim, A. A., Kazi, S. A., & Ranganath, S. (2004). A Web-based intelligent learning environment for digital systems. *International Journal of Engineering Education, 20*(1), 13-23.

Katz, S., Lesgold, A., Eggan, G., & Gordin, M. (1992). Modelling the student in Sherlock II. *Journal of Artificial Intelligence in Education, 3*, 495-495.

Kelly, D., & Tangney, B. (2002). *Incorporating learning characteristics into an intelligent tutor*.

Kim, J., Hill, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., . . . Hart, J. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. *International Journal of Artificial Intelligence in Education, 19*, 289-308.

Kim, Y., & Baylor, A. (2006a). Pedagogical Agents as Learning Companions: The Role of Agent Competency and Type of Interaction. *Educational Technology Research and Development, 54*(3), 223-243.

Kim, Y., & Baylor, A. (2006b). A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. *Educational Technology Research and Development,, 54*(6), 569-596.

Kim, Y., & Baylor, A. L. (2006c). A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. *Educational Technology Research and Development, 54*, 569-590.

Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin, 119*(2), 254-284.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent Tutoring Goes To School in the Big City. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Kozlowski, S. W. J., & Bell, B. S. (2007). A theory-based approach for designing distributed learning systems.

Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology; Journal of Educational Psychology, 63*(5), 505.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279-308.

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Leemkuil, H., Jong, T., & Ootes, S. (2000). Review of educational use of games and simulations.

Lepper, M. R., Drake, M. F., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Scaffolding student learning: Instructional approaches and issues*. Cambridge, MA: Brookline.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A Cross-Media Presence Questionnaire: the ITC-Sense of Presence Inventory. *Presence, 10*(3), 282-297.

Lester, J. C. (2011). Affect, Learning, and Delight. In S. D'Mello, A. Graesser, B. Schuller & J.-C. Martin (Eds.), *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011), LNCS* (Vol. 6974, pp. 2). Berlin Heidelberg: Springer-Verlag.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997a). *The persona effect: affective impact of animated pedagogical agents*.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997b). *The persona effect: affective impact of animated pedagogical agents.* Paper presented at the SIGCHI conference on Human factors in computing systems

Liggett, K. K., Reising, J. M., & Hartsock, D. C. (1999). Development and evaluation of a background attitude indicator. *The International Journal of Aviation Psychology, 9*(1), 49-71.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*: Prentice-Hall, Inc.

Loftin, B., Mastaglio, T., & Kenney, P. (2004). Outstanding research issues in intelligent tutoring systems (Research Report No. N61339-03-C-0156) Retrieved October 23 2006, from http://www.mymicsurveys.com/site/files/pub_4.pdf

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer‐Mediated Communication, 3*(2), 0-0.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). *Measuring (Tele)Presence: The Temple Presence Inventory*. Paper presented at the Twelfth International Workshop on Presence, Los Angeles, California, USA.

Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics, 44*(3), 229-243.

Magerko, B., Stensrud, B. S., & Holt, L. S. (2006). *Bringing the Schoolhouse Inside the Box - A Tool for Engaging Individualized Training*. Paper presented at the 25th Army Science Conference, Orlando, FL.

Mangos, P. M., & Johnston, J. H. (2009). Performance measurement issues and guidelines for adaptive, simulation-based training. In D. A. Vincenzi, J. A. Wise, M. Mouloua & P. A. Hancock (Eds.), *Human Factors in Simulation and Training* (pp. 301-320). Boca Raton, FL: CRC Press.

Martens, A., Diener, H., & Malo, S. (2008). Game-Based Learning with Computers–Learning, Simulations, and Games. *Transactions on edutainment I,* 172-190.

Mason, B. J., & Bruning, R. (2001). *Providing Feedback in Computer-Based Instruction: What the Research Tells Us*. Center of Instructional Innovation, University of Nebraska-Lincoln.

Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., . . . Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion, 2*(4), 315.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). Stress, Arousal, and Performance *Human Performance: Cognition, Stress, and Individual Differences*. Philadelphia, PA: Psychology Press.

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.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107-119.

Mayes, T., & Freitas, S. d. (2004). Review of e-learning frameworks, models and theories. *JISC e-Learning Models Desk Study*, from http://www.jisc.ac.uk/epedagogy/

McAlinden, R., Gordon, A. S., Lane, H. C., & Pynadath, D. (2009). *UrbanSim: A game-based simulation for counterinsurgency and stability-focused operations*.

Meehan, M., Insko, B., Whitton, M., & Brooks Jr, F. P. (2002). *Physiological measures of presence in stressful virtual environments*.

Menaker, E., Coleman, S., Collins, J., & Murawski, M. (2006). *Harnessing Experiential Learning Theory to Achieve Warfighting Excellence.* Paper presented at the Interservice/Industry Training, Simulation & Education Conference, Orlando, FL.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*: Muska & Lipman/Premier-Trade.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional science, 32*(1), 99-113.

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31*(2), 186-207.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358.

Moreno, R., Mayer, R. E., & Lester, J. C. (2000). *Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction*. Paper presented at the World Conference on Educational Mulitmedia, Hypermedia and Telecommunications, Chesapeake, VA.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*(2), 177-213.

Mory, E. H. (2004). Feedback research revisited. In D. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 745-783). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. *Journal of Computer Assisted Learning, 18*(3), 253-261.

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.

Murphy, C. (2011). *Why Games Work: The Science of Learning.* Paper presented at the Interservice/Industry Training Simulation and Education Conference (I/ITSEC), Orlando, FL.

Murphy, C., Chertoff, D., Guerrero, M., & Moffitt, K. (2011). Creating Flow, Motivation, & Fun in Learning Games *The Design of Learning Games*: Springer-Verlag.

Murray, T., & Arroyo, I. (2002). *Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems*. Paper presented at the International Conference on Intelligent Tutoring Systems, Biarritz, France.

Murray, T., & Arroyo, I. (2003, July). *Toward an operational definition of the zone of proximal development for adaptive instructional software.* Paper presented at the 25th Annual Meeting of the Cognitive Science Society Boston, MA.

Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. v. Merrienboer & M. Driscoll (Eds.), *Handbook of Research on Educational Communications and Technology* (3rd ed., pp. 125-144). New York: Lawrence Erlbaum Associates, Inc.

Nicholson, D., Fidopiastis, C., Davis, L., Schmorrow, D., & Stanney, K. (2007). *An adaptive instructional architecture for training and education.* Paper presented at the HCII 2007, Berlin.

Nicholson, D., Fiore, S., Vogel-Walcutt, J. J., & Schatz, S. (2009). *Advancing the science of training in simulation-based training*. Paper presented at the Human Factors and Ergonomics Society (HFES) Annual Meeting.

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

Oser, R. L., Cannon-Bowers, J. A., Salas, E., & Dwyer, D. J. (1999). Enhancing performance in technology-rich environments: Guidelines for scenrio-based training. In E. Salas (Ed.), *Human/technology interaction in complex systems* (pp. 175-202). Stamford, CT: JAI Press.

Oviatt, S. (2006). *Human-centered design meets cognitive load theory: designing interfaces that help people think*.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(4), 737-743.

Pallant, J. (2007). *SPSS Survival Manual*. New York, NY: Open University Press.

Parkes, A. M., & Coleman, N. (1990). Route guidance systems: a comparison of methods of presenting directional information to the driver. In E. J. Lovesey (Ed.), *Contemporary Ergonomics* (pp. 480-485). London: Taylor & Francis.

Person, N. K., & Graesser, A. C. (2003). Fourteen Facts about Human Tutoring: Food for Thought for ITS Developers. In V. Aleven, U. Hoppe, J. Kay, R. Mizoguchi, H. Pain, F. Verdejo & K. Yacef (Eds.), *Artificial Intelligence in Education 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a View Toward the Classroom* (pp. 335-344). Sydney, Australia

Phye, G. D., & Andre, T. (1989). Delayed retention effect: Attention, perseveration, or both? *Contemporary Educational Psychology, 14*(2), 173-185.

Piaget, J., & Smith, L. (1995). *Sociological studies*: Psychology Press.

Pine, A. (2009). Simulation: The (almost) real thing. *Proceedings Magazine, 135*(12/1), 282.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61-86.

Porayska-Pomsta, K., Mavrikis, M., & Pain, H. (2008). Diagnosing and acting on student affect: the tutor's perspective. *User Modeling and User-Adapted Interaction, 18*(1), 125-173.

Prensky, M. (2007). *Digital game-based learning*. St. Paul, MN: Paragon House Publishers.

Pridemore, D. R., & Klein, J. D. (1995). Control of practice and level of feedback in computer-based instruction. *Contemporary Educational Psychology, 20*(4), 444-450.

Procci, K., Singer, A. R., Levy, K. R., & Bowers, C. A. (2012). Measuring the Flow Experience of Gamers: An Evaluation of the DFS-2. *Computers in Human Behavior, 287,* 2306-2312.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE TRANS. SYS. MAN CYBER., 13*(3), 257-266.

Refiana, L., Mizerski, D., & Murphy, J. (2005). *Measuring the state of flow in playing online games*. Paper presented at the ANZMAC 2005 conference: Marketing Research and Research Methodologies (quantitative).

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185-218). Amsterdam: Elsevier.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences, 13*(3), 273-304.

Reuters. (2003, May 17). Video game hones US soldiers' fighting skills, *The Straits Times*.

Roll, I., Baker, R., Aleven, V., McLaren, B., & Koedinger, K. (2005). Modeling students' metacognitive errors in two intelligent tutoring systems. *user modeling 2005*, 151-151.

Roth, P. (2003). 'America's Army'Is Big Hit, and Not Just with Civilians. *Wall Street Journal*.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology, 53*(1), 61-86.

Ryu, J., & Baylor, A. L. (2005). The psychometric structure of pedagogical agent persona. *TECHNOLOGY INSTRUCTION COGNITION AND LEARNING, 2*(4), 291.

Salas, E., & Cannon-Bowers, J. (2001). The Science of Training: A Decade of Progress. *Annual Review of Psychology, 52*, 471-499.

Salas, E., Priest, H. A., Wilson, K. A., & Burke, C. S. (2006). Scenario-based training: Improving military mission performance and adaptability. In A. B. Adler, C. A. Castro & T. W. Britt (Eds.), *The psychology of serving in peace and combat operational stress* (Vol. 2). Westport, CT: Greenwood Publishing Group, Inc.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009a). Performance Measurement in Simulation-Based Training. *Simulation & Gaming, 40*(3), 328-376.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009b). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming: An Interdisciplinary Journal,, 40*(3), 328-376.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Bulletin, 3*(4), 207-217.

Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. *Self-regulated learning and academic achievement: Theoretical perspectives, 2,* 125-151.

Seagull, F. J., Wickens, C. D., & Loeb, R. G. (2001). *When is less more? Attention and workload in auditory, visual, and redundant patient-monitoring conditions*.

Shute, V. J. (2007). Focus on Formative Feedback. Princeton, NJ: Educational Testing Service.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it - Or Can you? Evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education, 18*, 289-316.

Shute, V. J., Masduki, I., & Donmez, O. (2010). Conceptual Framework for Modeling, Assessing and Supporting Competencies within Game Environments. *Technology, Instruction, Cognition, and Learning, 8*(2), 137-161.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects*, 295-321.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*(1), 97-115.

Skinner, B. F. (1953). *Science and human behavior*: Free Press.

Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 13*(4), 484-493.

Slater, M., & Garau, M. (2007). The use of questionnaire data in presence studies: do not seriously Likert. *PRESENCE: Teleoperators and Virtual Environments, 16*(4), 447-456.

Sotomayor, T. M. (2008). *Evaluating tactical combat casualty care training treatments effects on combat medic trainees in light of select human descriptive characteristics.* Ph.D., University of Central Florida.

Sottilare, R. (2009). *USING STUDENT MOOD AND TASK PERFORMANCE TO TRAIN CLASSIFIER ALGORITHMS TO SELECT EFFECTIVE COACHING STRATEGIES WITHIN INTELLIGENT TUTORING SYSTEMS (ITS).* Ph.D., University of Central Florida.

Sottilare, R., & Goldberg, B. (2013). Designing Adaptive Computer-Based Tutoring Systems to Accelerate Learning and Facilitate Retention. *Cognitive Technology*.

Sottilare, R., Goldberg, B., Brawner, K. W., & Holden, H. (2012). *Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems*. Paper presented at the Interservicer/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Sottilare, R. A., & Gilbert, S. (2011). *Considerations for adaptive tutoring within serious games: authoring cognitive models and game interfaces.* Paper presented at the International Conference on Artificial Intelligence in Education (AIED), New Zealand.

Sottilare, R. A., Holden, H. K., Brawner, K. W., & Goldberg, B. S. (2011). *Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training*.

Stone, B. A. (2008). Human Factors Guidelines for interactive 3D and Games-based training Systems Design Retrieved from www.eece.bham.ac.uk/Default.aspx?tabid=154

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games–An overview. *Skövde: University of Skövde (Technical Report HS-IKI-TR-07-001).*

Sweller, J. (1999). Instructional design in technical areas.

Sweller, J. (2008). Human cognitive architecture. *Handbook of research on educational communications and technology*, 369-381.

Sweller, J., Merrienboer, J. J. G. v., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review, 10*(3), 251-296.

Sweller, J., Van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10* (3), 251–296.

Tabbers, H. K. (2002). *The modality of text in multimedia instructions: Refining the design guidelines*: Open Universiteit.

Thorndike, E. L. (1913). *Educational Psychology: The Psychology Of Learning*. New York: Teachers College Press.

Topolski, R., Leibrecht, B., Cooley, S., Rossi, N., Lampton, D., & Knerr, B. (2010). Impact of Game-Based Training on Classroom Learning Outcomes: DTIC Document.

TRADOC. (2011). The U.S. Army Learning Concept for 2015. TRADOC.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*(3), 358-381.

Tychsen, A., Newman, K., Brolund, T., & Hitchens, M. (2007). *Cross-format analysis of the gaming experience in multi-player role-playing games*. Paper presented at the Digital Games Research Associate Conference.

Ulicsak, M., & Wright, M. (2010). Games in education: Serious games. *Bristol, Futurelab. Retrieved November, 27,* 2010.

van der Hulst, A., Muller, T., Besselink, S., Coetsier, D., & Roos, M. C. (2008). *Bloody Serious Gaming–Experiences with Job Oriented Training*.

van Gorp, T. J. (2006). *Emotion, Arousal, Attention and Flow: Chaining Emotional States to Improve Human-Computer Interaction.* Masters, University of Calgary.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.

Van Mulken, S., & André, E. (1998). The persona effect: how substantial is it.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., . . . Wintersgill, M. (2005). *The Andes Physics Tutoring System: Five Years of Evaluations*. Paper presented at the International Conference on Artificial Intelligence in Education, Amsterdam.

VCom3D. (2013). VCommunicator Studio Retrieved May 20, 2013, from http://www.vcom3d.com/index.php?id=vcstudio

Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers & Education, 55*(2), 576-585.

Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, 52-91.

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

Waldman, A. (2009). Serious Gaming Retrieved April 30, 2012, from http://www.ngaus.org/content.asp?bid=14851

Weible, J. (n.d.). Preparing soldiers to respond knowledgeably to battlefield injuries. *MSTC U.S. Army* Retrieved August 10, 2012, from http://www.laerdal.com/hk/UserStories/42358816/Preparing-soldiers-to-respond-knowledgeably-to-battlefield-injuries

Whitaker, L., Peters, L., & Garinther, G. (1989). *Tank crew performance: Effects of speech intelligibility on target acquisition and subjective workload assessment*.

Wickens, C. D. (1991). Processing resources and attention. *Multiple-task performance*, 3-34.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science, 3*(2), 159-177.

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 25*(2), 227-248.

Widmayer, S. A. (2005). Schema theory: An introduction. *Retrieved April, 14*.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. *Multiple-task performance*, 329-360.

Wisher, R. A., Macpherson, D. H., Abramson, L. J., Thorton, D. M., & Dees, J. J. (2001). The Virtual Sand Table: Intelligent Tutoring for Field Artillery Training: U.S. Army Research Institute for the Behavioral and Social Sciences.

Witmer, B. G., & Singer, M. F. (1994). Measuring presence in virtual environments: DTIC Document.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Woolf, B. (1992). AI in Education. In S. Shapiro (Ed.), *Encyclopedia of Artificial Intelligence* (pp. 434-444). New York: John Wiley & Sons, Inc.

Woolf, B. P. (2009). *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning*. Burlington, MA: Morgan Kaufmann.

Xiao, Y., Wang, Z., Wang, M., & Lan, Y. (2005). The appraisal of reliability and validity of subjective workload assessment technique and NASA-task load index]. *Zhonghua lao dong wei sheng zhi ye bing za zhi= Zhonghua laodong weisheng zhiyebing zazhi= Chinese journal of industrial hygiene and occupational diseases, 23*(3), 178.

Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). *A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems.

Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing*, *23*(2), 102-110.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25-32.