Examining Motivational Feedback For Sensor-Free Detected Frustration Within Game-Based Learning

Jeanine A. DeFalco

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Examining Motivational Feedback For Sensor-Free Detected Frustration

Within Game-Based Learning

Jeanine A. DeFalco

Social interactions, decision-making, perceptions, and learning are all influenced by affect. Frustration, anxiety, and fear in particular can draw cognitive resources away from successful task completion, causing the learner to focus on the source of the emotion instead. Serious games offer an ideal environment to investigate how feedback influences student affect and learning outcomes, particularly when feedback is delivered via computer system detection. This dissertation discusses the results of an experiment run in September 2015 to investigate which motivational feedback condition yields the most significant correlation to positive learning gains when a computer system intelligently generates and delivers feedback based on the detection of frustration while participants played the serious video game, vMedic, a combat casualty care simulation which includes triage tasks. Of the three motivational feedback conditions examined (self-efficacy, social-identity, and control-value), the self-efficacy motivational feedback interventions yielded positive, statistically significant learning gains when compared to the social identity and control-value feedback conditions, as well as the non-motivational feedback control condition, and the no feedback control condition.

TABLE OF CONTENTS

LISTS OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

> Our life, as individual persons and as members of a perplexed and struggling race, provokes us with the evidence that it must have meaning. Part of the meaning still escapes us. Yet our purpose in life is to discover this meaning, and live according to it. We have, therefore, something to live for. The process of living, of growing up, and becoming a person, is precisely the gradually increasing awareness of what that something is.
>
> -- *No Man is an Island*, Thomas Merton

Without the wisdom, love, and unqualified support of my father, Dr. Anthony DeFalco, and my mother, Mary DeFalco, I would not have the privilege of writing this acknowledgement section. Two of the most important people in my life provided me with constant, unwavering faith in me and my abilities. They have made this road less travelled by a navigable one, making all to the difference between a world of intangible hopes and a world of substantive accomplishments. When faced with the daunting prospect of returning to school as a single divorced mother of four children -- to pursue a degree beyond the scope of anything I had previously ventured upon -- my hopes and dreams were only made possible through the faith, prayers, and support of my parents. I can never repay you for the blessing you have given me, but I will try.

And to my children: Andrew, Christopher, Eleanor, and Gwyneth Higgins. They have been my life's buoys to guide me back to shore when I felt overwhelmed, and despaired I would never find myself back on solid ground again. My children have reflected back to me their pride in my efforts, their faith in my abilities, and a demonstrable adoption of my values of hard work, integrity, and perseverance. They have emerged from this process as people I am proud to have shaped and inspired, inordinately pleased to know that the world is a better place because they

are in it. They will probably never know how much they kept me from going adrift from my path, but they have been bedrocks of purpose and meaning for me in the darkest and brightest hours of this journey.

And indeed there was an army – a virtual battalion – that walked with and supported me on this path. My siblings, Darian DeFalco, Loren DeFalco, and Alicia Salazar. Siblings unparalleled in their love and faith, they helped me in innumerable ways, both personally and professionally. They join the ranks of my best friends, Theresa O'Brien, Vasiliki Geourgoulas-Sherry, Francie Leader, and Amy Vanstone who were my cheerleaders, spiritual confidants, study buddies, research compatriots, and touchstones of sanity throughout this entire process. And to Jeff Bollier, whose love and faithfulness helped me see the path to a promise of a better tomorrow: you have been and continue to be a blessing to me.

So flanked by love, this battalion and so many others supported throughout my tenure at Columbia, beginning when I ventured into the great unknown in the fall of 2012, convinced the admissions committee made some kind of mistake. Assured no mistake had been made, I walked by faith into unbelievably good fortune when I met my epic advisor, Dr. Ryan Baker -- without whom I would not be where I am today. His brilliance and drive, his confidence in me, his unwavering support and uncompromising pursuit of perfection, made all things possible for me throughout the entirety of my doctoral studies.

Without Ryan's leap of faith in taking me on as a Doctoral Research Fellow, I would not have had the unparalleled opportunity to explore a world so rich, so full of possibilities. And I would not have had the unbelievably good fortune to have met Dr. Robert Sottilare of the Army Research Laboratory, who gave me his approval to work with Ryan from the very beginning. Not only did Dr. Bob take me under his wing in so many ways, including graciously serving on

my dissertation defense committee, but over and again Dr. Bob provided me with so many opportunities to grow and develop as a researcher. I cannot express enough my deepest gratitude for the opportunity to work with him, to have been able to learn from a man whose integrity and compassion will forever be a source of inspiration to me. Through Dr. Bob I also had the very good fortune to meet and work with the rest of our team at ARL: Dr. Ben Goldberg, Dr. Keith Brawner, Dr. Anne Sinatra. I will forever be thankful to these individuals who have all taken me under their wing and shared a beverage with me at some point.

I would be remiss without including my deepest thanks to Dr. John Black, who not only served on my dissertation committee but was the man responsible for accepting me into Columbia in the first place: I am forever in his debt. And to Dr. Dolores Perin and Dr. Stanley Hum, my other dissertation committee members: their support, wisdom, and insights fill me with deep gratitude and appreciation for all they have done to see me through the end of my journey.

Also, my sincerest appreciation to Vasiliki Georgoulas Sherry (yes, she warrants two notable mentions), Colonel James Ness, and Dr. Mike Matthews at the United States Military Academy at West Point, without whom my dissertation study would have been a far different experience, and not nearly as meaningful to me.

To Dr. Luc Paquette – there would have been no dissertation study without you. To Dr. Jaclyn Ocumpaugh – your training in BROMP, and your overall guidance and support have been so important to me over the years. To Dr. Michael Swart and Doug DiStefano – your friendship and support in my research efforts and my life more generally, were bright lights in my world on so many occasions; thank you both. To my lab mates at Columbia: Sweet San Pedro, Mia Almeda, Elle Yuan Wang, Yang Jiang, Miggy Andres, Stefan Slater, Shimin Kai, Aaron Hawn, Diego Luna Bazaldua – I am so grateful for the help and friendship over the years. And to the

North Carolina State constituents: Dr. Jonathan Rowe, Dr. James Lester, and Dr. Brad Mott working with you has been an honor and a privilege.

I have many more family members and friends to express my gratitude and appreciation: my Uncle Joseph DeFalco (your generosity and support made all the difference when I needed it most); my aunt, Fontane Rizzutto; my aunt, Dr. Amy Dunbar; my St. Anthony's friends (Nappi, may you rest in peace – you are so missed); and all my peers at Teachers College and Pace University. I am overwhelmed with the depth and breadth of gratitude owed to all who made this possible.

And last, but never least, to my God who never abandoned me: may I always be an instrument of your Peace and share the Good News that indeed no man nor woman is an island, but rather that we are "all members of one another."

DEDICATION

This dissertation is dedicated to my parents, Dr. Anthony and Mrs. Mary DeFalco; to my children: Andrew, Christopher, Eleanor, and Gwyneth Higgins; to my family (Darian, Loren, and Alicia, Uncle Joe, Aunt Fontane, Aunt Amy) and friends (Theresa O'Brien, Jeff Bollier, Wilhelm, Steve, Gary, Fink, OB, Kevin, O'Donnell, Brian, Walsh, Jenn, Christina, Mo, Francie Leader, Amy VanStone); and in memory of my childhood friend, John Nappi III.

Deo gratias pro gratia, per quem omnia possibilia sunt.

Chapter I: INTRODUCTION

Determining how to effectively respond to learner affect is important not only in face-to-face learning environments (Pekrun, Goetz, Titz, & Perry, 2002), but also within the field of intelligent tutoring systems (ITS) (Goldberg et al., 2012). This requires not only methods to accurately identify affect, but also developing a suite of accompanying interventions that can respond to learner affect (D'Mello et al., 2008; D'Mello, Craig, Fike, & Graesser, 2009; D'Mello, Lehman, & Graesser, 2011; Woolf et al., 2009).

In an effort to help learners regulate their affective states, some computer tutoring systems researchers have used interventional feedback messages to motivate the learner through a frustrated state (Rebolledo-Mendez et al., 2005; Robison, McQuiggan, & Lester, 2009). However, these researchers have noted that where frustration has been detected and feedback delivered, learners do not always respond positively to these interventions, but rather may react negatively to feedback provided by the system (Robison, McQuiggan, & Lester, 2009). This has given rise to the need to take a closer examination of the design of motivational feedback messages delivered to learners in a frustrated state (Arroyo et al., 2007; Robison, McQuiggan, & Lester, 2009) to determine the most effective approach for addressing learner frustration via interventional feedback messages.

Within this context, then, the gap addressed by the current work is understanding what kind of motivational feedback messages effectively ameliorates the affective state of frustration within a simulation-based training game, and promotes significant learning gains when delivered within an ITS. Three theories of motivation were targeted to design feedback messages: (1) theory of control-value (Pekrun, Elliot, & Maier, 2006); (2) theory of social identity (Tajfel &

Turner, 1979); (3) theory of self-efficacy (Bandura, 1977). These theories are distinct from each other in the way they target either a person's sense of what they value (control-value theory), who they are (social identity theory), or what a person believes they can achieve (self efficacy theory).

Statement of Problem

Prior research has demonstrated that motivational processes associated with affective states have a significant impact on memory, perception, attention, and categorization (Gable & Harmon-Jones, 2010; Harmon-Jones, Gable, & Price, 2013). However, the research on how frustration impacts cognition has yielded inconclusive and mixed results. Further, there is little to no research that identifies what kind of motivational feedback messages would produce the most significant learning outcomes while addressing the affective state of frustration in an ITS learning platform. As such, determining how to respond to learner frustration to yield robust positive learning outcomes is an important area of interest in ITS research.

Purpose of the Study

Given the central role affect, motivation, and cognition play in designing and implementing learning environments, research is needed to address the impact different motivational messages have on learning outcomes within game-based training environments. Therefore, the purpose of this dissertation is to examine the effect of motivational feedback messages delivered while participants are in a frustrated state while playing the serious video game vMedic, as part of a modified TC3Sim combat care course (National Association of

Emergency Medical Technicians, 2016) delivered by GIFT (Generalized Intelligent Framework for Tutoring) developed by the US Army Research Laboratory (Sottilare, Brawner, Goldberg, and Holden, 2012).

Using previously published sensor-free detectors of student frustration (Paquette et al., 2015), GIFT has been configured to automatically detect whether students were frustrated while playing vMedic, and uses these sensor-free detectors to trigger frustration adaptations in the form of feedback messages.

Within this context, this research aims to determine which of three types of motivational feedback message designs, delivered in response to sensor-free affect detection and based on three distinct theories of motivation (control-value, social identity, and self-efficacy), yield significant improvements to learning outcomes while participants are engaged in vMedic. The overarching purpose is to determine how motivational feedback messages can ameliorate frustration and support learning in an ITS learning environment.

Data will be collected across five conditions where the type and presence of motivational feedback is manipulated while the non-motivational content presented is held constant to determine the effect different kinds of motivational feedback has on learning. The results of this study will inform future work on game-based learning, specifically informing the ITS research community on the most effective feedback design to respond to frustration. The goal is to provide empirical evidence that theory-based designed motivational feedback messages have an effect on affect regulation and learning.

Research Questions

In order to accomplish the goal of this study, the research questions include examining which conditions yield statistically significant improvements in learning outcomes. Specifically, this research will examine whether greater learning gains are achieved in conditions with motivational feedback vs. no motivational feedback; in conditions with feedback vs. no feedback; which specific motivational feedback condition yields the most significant learning outcomes; whether environmental conditions such as a sense of presence impact learning across conditions; and whether the character trait of grit interacts with motivational feedback conditions and impacts learning outcomes.

Summary

This study asks whether interventional motivational feedback will promote statistically significant learning gains when participants are in a state of frustration. More specifically, this study examines whether motivational feedback delivered by an ITS upon the detection of a participant's high frustration, while engaged in the serious video game vMedic, would be correlated to greater learning gains. This study aims to determine if there is a difference between motivational conditions on learning gains, how frustration interacts with motivational messages and learning, and whether environmental contexts or character traits interact with these motivational conditions to yield different learning outcomes. The motivational feedback conditions examined include feedback designs based on the theories of (1) control-value theory; (2) social identity theory; (3) self-efficacy theory.

Overview of Dissertation and Profiles of Chapters

The five chapters of this dissertation begins with the first chapter that provides an introduction to the study, its purpose, rationale, and significance. The main theoretical constructs that have been investigated in the study are introduced in this chapter.

The second chapter reviews relevant literature and research on issues related to motivation, affect, and cognition; affect-sensitive computer tutoring systems; the complex relationship of frustration to cognition and learning; a review on intervention feedback designs used to address frustration in computer tutoring systems; considerations for designing feedback messages for a military population; and a review of the three motivational theories used to design feedback message interventions: control-value theory, social identity theory, and self-efficacy theory.

The third chapter discusses a prior study conducted in September 2013 that yielded data from which the sensor-free detectors of frustration were built and were subsequently used in this dissertation study. Additionally, this September 2013 study's results helped inform the main study's design.

The fourth chapter presents the summaries of the main research study design, results, and discusses the results in relation to the research question. The main study investigates the participant's learning gains when motivational feedback messages were delivered during the vMedic game upon the system detection of frustration. The results indicate that there was a statistical significant difference in positive learning gains between motivational feedback conditions and no motivational feedback condition. Further, the results indicate that there was an interaction effect of learning gains with levels of measured grit in the control-value condition

(condition 2) and there was an interaction of learning gains by frequencies of system detected frustration in the self-efficacy condition (condition 4).

Lastly, the fifth chapter discusses the theoretical and practical implications of the findings in addition with future directions.

Chapter II: LITERATURE REVIEW

Section 1. Overview

This chapter reviews relevant literature and research on issues related to motivation, affect, and cognition; affect-sensitive computer tutoring systems; the complex relationship of frustration to cognition and learning; a review on intervention feedback designs used to address frustration in computer tutoring systems; considerations for designing feedback messages for a military population; and a review of the three motivational theories used to design feedback message interventions: control-value theory, social identity theory, and self-efficacy theory. The chapter contains eight sections.

The first section presents an overview of the chapter. The second section presents briefly discusses the relationship of motivation, affect, and cognition. The third section presents literature on affect-sensitive computer tutoring systems. The fourth section reviews literature on the complex relationship of frustration to cognition and learning. The fifth section reviews literature on a review on intervention feedback designs used to address frustration in computer tutoring systems. The sixth section presents considerations for designing feedback messages for a military population. The seventh section reviews the three motivational theories used to design feedback message interventions: control-value theory, social identity theory, and self-efficacy theory. The eighth section contains a summary of the literature review.

Section 2. Motivation, Affect, and Cognition

Current research in cognitive psychology and neuroscience confirms the central role affect plays in mental processes and behavior. Cognition and affect have been identified as separate yet inextricably linked interactive aspects of brain organization (Barrett, 2006; Ciompi & Panksepp, 2004; Dalgleish & Power, 1999; Mandler, 1984; Panksepp, 2003b). Where cognition involves the neocortical processing of information largely from sensory input, affects are not encoded as information. Rather, affect is identified as diffuse global states generated by deep subcortical brain structures (Panksepp, 1998a, 1998b).

Defining motivation and understanding what makes a person behave in a certain way is a complex endeavor as well. Within the tradition of experimental psychology, there have been two intellectual traditions that have employed the concept of needs as a way to unpack motivation. Hull (1943) saw motivation as a directional response to a stimulus that addresses a drive or a need state, whereas Murray (1938) viewed motivation as the psychological needs that function as a force to organize perception and cognition. For the purpose of this paper, this author adopts a definition of motivation articulated by Deci and Ryan (2000) that categorizes motivation as psychological propensities and functions that inform intentional behavior to achieve goals as well as satisfy needs and interests. In this way, then, the constructs of self-efficacy, control-value, and social identity employed in this dissertation study can all be seen as psychological propensities and functions that inform behavior, and as such, should be considered as separate yet similar motivation variables.

Over the last decade, there has been a growing interest in the relationship between motivation and cognition, including perception, memory, task-switching, response inhibition, categorization, decision making, and selective attention (Braver, 2015; Locke & Braver, 2010; Maddox & Markman, 2010; Pessoa, 2009; Pessoa & Engelmann, 2010; Shohamy & Adcock, 2010). Further, theories of cognitive processing describe the interactive nature of motivation, affect, and cognition (Braver, 2015; Gable & Harmon-Jones, 2010; Harmon-Jones, Gable, & Price, 2013). According to these theories, supporting cognitive performance requires not only an understanding of an individual's traits and characteristics, but should also include in the analysis an individual's social context, including relevant linguistic control systems such as feedback (Buck, 1985; Locke & Braver, 2010; Maddox & Markman, 2010; Pessoa, 2009; Pessoa & Engelmann, 2010; Shohamy & Adcock, 2010).

Buck (1985) has argued for a model of interaction between motivation, affect, and cognition that highlights the use of language as a means through which culturally patterned systems of behavior are reinforced. Buck (1985) notes that what is unique to the human experience of motivation, affect, and cognition is that human behavior is functionally independent of biology, and instead is controlled by linguistic control systems that include logic, reasoning and social rules. It is through these linguistic control systems, Buck (1985) argues, that allows for the contemplation of goal strategies, decision making, and future planning.
As such, it is important to not only understand the implications of positive and negative affect on cognition and behavior, but to examine how linguistic control systems shape responses to motivational stimuli, guides our interactions with our environment, and enhances or impedes memory (Buck, 1985; Tooby & Cosmides, 2008; Keltner & Kring, 1998; Parrot, 1993).

Understanding this interactive effect is particularly important in the field of intelligent tutoring systems, where the primary aim of these systems is to not only support cognition and learning, but to do so in a manner that is comparable in sensitivity and effectiveness to a high-quality human tutor (Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009), including accounting for shifts in affect in students and modifying motivational stimuli and feedback to improve engagement and learning outcomes (Baker, D'Mello, Rodrigo, & Graesser, 2010; Corbett, Koedinger, & Anderson, 1997; D'Mello, Picard, & Graesser, 2007; Graesser, Conley, & Olney, 2012; Heffernan & Koedinger, 2002; Lehman, Matthews, D'Mello, & Person, 2008; Lepper, Woolverton, Mumme, & Gurtner, 1993; Robison, McQuiggan, & Lester, 2009; Woolf, Burleson, Arroyo, Dragon, Cooper, & Picard, 2009).

Motivation manipulations prime a person to exert a cognitive control to overcome obstacles in achieving their learning goals (de Wit & Dickinson, 2015; Marien, Aarts, & Custer, 2015). As such, when considering developing motivational feedback messages to be used as an intervention method in an intelligent computer tutoring system, it is important that messages are linguistically relevant and meaningful to the target population, and delivered in a manner that is timely, unobtrusive, and targeting affect that if left unaddressed, may result in a withdrawal of effort (Price, Handley, & Millar, 2011).

Section 3. Affect-Sensitive Computer Tutoring Systems

Computer tutoring system researchers have recognized the need to identify and address affective states that lead to disengagement in learning (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Lehman, & Graesser, 2011; D'Mello Strain, Olney, & Graesser, 2013; Forbes-Riley, Litman, Friedberg, 2011; Gee, 2004, 2007; Picard et al., 2004). Prior research has shown evidence that learners will likely remain in their current affective state – particularly negative affective states -- when interventions are not provided by ITSs (Baker et al, 2007; Baker et al., 2010; D'Mello et al., 2007). Also, providing interventions in the form of feedback messages has been shown to positively effect the learning of domain content in ITSs (Wagster, Tan, Wu, Biswas, & Schwartz, 2007; Roll, Aleven, McLaren, & Koedinger; 2011).

Some affective states have relatively uncomplicated relationships with student learning outcomes – engaged concentration appears to be positively associated (Craig et al., 2004; Pardos et al., 2014) while boredom is negatively associated (Craig et al., 2004; Pardos et al., 2014). However, research has shown that the affective state of frustration is more complex, where brief periods of frustration are not problematic, but extended frustration is associated with worse learning outcomes (D'Mello & Graesser, 2011; Liu et al., 2013; Robison, McQuiggan, & Lester, 2009). It is important to understand how intelligent tutoring systems (ITSs) can respond to learner frustration for future affect-sensitive learning environments (Picard et al., 2004).

Section 4. Frustration and Learning in Intelligent Tutoring Systems

As discussed above, the relationship between frustration and engagement is complex. In a review of the literature, frustration has been related to positive, null, negative, and mixed learning outcomes in ITSs.

Learning Outcomes and Frustration: Negative, Positive, Null, and Mixed Findings

Negative outcomes and frustration. In terms of the negative impact of frustration on learners, studies have demonstrated frustration can lead to gaming the system (Baker et al., 2006). There is also evidence that frustration can divert student attention from learning tasks (McQuiggan, Lee, & Lester, 2007), and lead learners to worrying about excessive failure (D'Mello & Graesser, 2011).

Positive outcomes and frustration. Research has shown that in some circumstances, positive learning outcomes are achieved by students in a frustrated state (Pardos et al., 2014). Using automated detectors of affect and behavioral engagement, researchers examined over a thousand students' actions in an entire year's log file data in the ASSISTments tutoring system to assess the predictive nature of affect and engagement and high-stakes test outcomes (Pardos et al., 2014). The findings of this examination showed a significant positive relationship between frustration and learning (Pardos et al, 2014).

Null outcomes and frustration. Baker, D'Mello, Rodrigo, & Graesser (2010) examined the rate of occurrence, persistence, and impact of students' cognitive-affective states during the use of three different computer-based learning environments. Their findings included that boredom and not frustration was the primary cognitive-affective state that led to an increase in likelihood of gaming behavior, which has in turn been associated with poorer learning (Baker et al., 2004; Cocea et al., 2009).

Further, in a study that investigated which observable affective states and behaviors of undergraduate freshmen in a computer programming class could be used to predict student achievement, frustration was found to not be a predictor of achievement (Rodrigo et al., 2009). Also, frustration was found to not be correlated with learning gains in any of studies of AutoTutor (D'Mello, Strain, Olney, & Graesser, 2013).

Mixed outcomes and frustration. Still other studies have showed mixed results in learning gains while students were in a frustrated state, depending on context. In a study by Liu, Pataranutaporn, Ocumpaugh, & Baker (2013), brief periods of frustration were associated with positive learning gains, but lengthier periods of frustration were associated with poorer learning gains.

Given the range and complexity of the impact of frustration on learning outcomes, then, further research is required to unpack the impact of frustration on learning and, equally important, how best to respond to an individual's frustrated state in an intelligent tutoring system (ITS); when to do so and how.

Section 5. Intervention Feedback Designs for Frustration

When a learner is in a frustrated state in ITSs, the range of solutions to address this frustration includes changing the elements in a system that elicits frustration, and supporting the learner in their ability to recover, manage, and persist in their task (Klein, Moon, & Picard, 2002; Kapoor, Burelson, & Picard, 2007). Amsel's (1992) frustration theory supports the notion that goal attainment includes overcoming emotional conflict rather than avoiding emotional conflict. Therefore, to encourage a learner to overcome frustration, while not changing the nature of the system elements, requires finding ways to help the learner recover, manage, and persist through frustration to persist in their learning tasks through the use of feedback messages (Kapoor, Burelson, & Picard, 2007).

The term "feedback" originated in cybernetics to denote processes by which information was delivered regarding the effects and consequences of actions (Wiener, 1948). In this context, then, feedback interventions are essentially a form of linguistic control that shapes the perception and understanding of an individuals actions with the added potential to influence the individual's future actions (Allwood, Nivre, & Ahlsen, 1992).

McQuiggan, Lee, & Lester (2007) maintain that ITSs should provide support, including the use of specific feedback, to help students cope with frustration to increase their tolerance of frustrating learning situations. Through diagnosis and detection of the affective state of frustration, ITSs can be configured to enact corrective affective scaffolding strategies that would facilitate specific feedback motivational messages (Robison, McQuiggan and Lester, 2009).

However, just as human-to-human assessment and feedback is a complicated endeavor, selecting interventions to respond effectively to learners in a frustrated state and provide the best possible feedback is similarly a complicated process. D'Mello, Strain, Olney, and Graesser (2013) note that a "one size fits all" approach to affective feedback is unlikely to regulate emotional experiences such as frustration, and that what is needed is an approach that coordinates cognition and emotions that is also adaptive to an individual's knowledge, goals traits, and moods.

Empathetic Feedback Approach

One approach to regulating frustration has been the use of empathetic feedback messages delivered to a learner in a frustrated state. In a study by Klein, Moon, and Picard (2002), empathetic feedback messages were delivered to frustrated participants playing a computer adventure game. This study found evidence that the empathetic messages relieved and aided participants in managing their frustrated state (Klein, Moon, & Picard, 2002).

Using the learning environment *Crystal Island*, Robison, McQuiggan and Lester (2009) investigated the consequences associated with an agent in *Crystal Island* providing empathetic responses via short, text-based responses when frustration of the participant was detected. In this investigation, two studies were conducted using two different designs of empathetic responses. The first study used empathetic feedback that paralleled the detected affective state of the participant, while the second study used affect-directed task-based feedback. For example, in the first study, the system would provide feedback that would say "Yes, I'm very frustrated as well!"

In the second study featuring affect-directed task-based feedback responses, messages were designed to be more motivational in nature, encouraging the participant to transition to a positive emotional state. For example, the system in this condition would provide a response saying, "I can understand why you are frustrated, but if you keep working, I'm sure you will figure it out," (Robison, McQuiggan, & Lester, 2009).

Using the data from self-reports of students' ratings on the quality of the agent's feedback messages, the results of both studies revealed that empathetic feedback interventions delivered while the participant was in a negative affective state were rated less favorably than when the students were in a positive affective state. The authors point out that it is possible other intervention messages might have rendered different results (Robison, McQuiggan, & Lester, 2009).

Motivational Feedback Approach

Another approach to designing feedback intervention messages has been motivationally designed feedback (Narciss, 2008). The motivational feedback model is contextualized within the theories of self-regulated learning, where the primary function of this feedback rests in guiding the learner to successfully regulate his or her learning process (Butler & Winne, 1995; Narciss, 2008). Narciss (2008) maintains that feedback that motivates learners to successful task completion can provide a mastery experience that supports the development of a learner's positive self-efficacy. This approach can be seen in an experiment by Burleson (2006) and his research with an affective agent built within an intelligent tutoring system.

In Burleson's (2006) experiment, "Stuck[1]" states were detected through sensors for groups of children (ages 11-13). An affective learning companion (ALC) delivered feedback interventions during a Towers of Hanoi[2] activity when children were in this "Stuck" state. In

---

[1] Burelson (2006) coined "Stuck," defining it as a state of non-optimal experiences that incorporates the feeling of frustration.

[2] The Tower of Hanoi, also called the Tower of Brahma or Lucas' Tower, is a mathematical puzzle.

Burelson's experiment, there were four conditions of support interventions: (1) affect support intervention triggered by sensor feedback, where the ALC empathetically mirrored the behavior of the learner; (2) affect support intervention by pre-recorded interactions with motivating feedback messages, (e.g., "Just remember that all of your effort does help you learn," (Burelson, 2006, p. 131); (3) task support intervention by mirroring behavior triggered by sensor feedback that also gave information on how to compete the task; (4) task support intervention by pre-recorded interactions that provided text on guiding the learner how to complete the Tower of Hanoi activity.

While there were no significant differences in any of the motivation measures between the two mirroring intervention groups (affect vs. task), when interventions were assessed, there was a main significant effect for differences in beliefs about strategies: girls reported that they believed they would be able to use the strategies provided in the interventions more than the boys did. Also, there was a significant correlation with interventions that indicated that regardless of level of frustration, both boys and girls who received affect support had self-reported higher values on a metacognitive/meta-affective scale, indicating higher levels of Flow and lower levels of "Stuck" (Burelson, 2006). As such, Burelson's (2006) study, while it did not measure learning outcomes, did demonstrate how motivational feedback was effective for at least one group of learners.

A review of the literature reveals evidence that the impact of motivational feedback differs according to groups distinguished between low ability and high ability, unmotivated and motivated. Meyer and Turner (2006) showed that perceived politeness, considered a motivational aspect of feedback, was better received by students who were not as experienced with computers than for those who were more experienced. Additionally, Rebolledo-Mendez et al., (2006) investigated the effect of motivationally scaffolding in the M-Ecolab tutoring system. The findings included that unmotivated students performed better with motivational feedback, whereas those students who were already motivated did not benefit from the motivational support.

Empathetic and Motivational Feedback Approach

D'Mello et al. (2010) conducted an experiment comparing affect-sensitive and non-affective sensitive versions of AutoTutor. When the affect-sensitive version of AutoTutor detected boredom, frustration, and confusion, the system would respond via empathetic and motivational responses regarding the material. Using pre-test data to split participants into low and high prior-knowledge groups, the results showed that low prior-knowledge participants learned significantly more from the supportive tutor than the regular tutor (d=.713), where as the students with more prior knowledge did not benefit from the supportive AutoTutor even when boredom, confusion, or frustration were detected.

Given that there is evidence that indicates effective motivational feedback messages should be designed for a specific target population, I will next examine how motivational theories can inform the design of feedback messages for an adult population within a military environment. The justification for this approach rests on the both the target population within which this study was conducted (cadets from the United States Military Academy at West Point) and that the results of this study will influence future intervention designs geared for both cadets and active members within the US Army.

Section 6. Motivational Feedback Messages Designed for a Military Population

What follows is an examination of unique traits and characteristics of a military population that warrant special consideration in designing motivational feedback messages.

Special Considerations: Military Population

The training and education of an adult learner within a military population, either as an enlisted private or as an officer-in-training enrolled at a military academy, is distinct from that of a civilian who attends post-secondary education institutions. While membership in an academic, university community is fluid and often self-directed, membership within a military community is much more rigid and cohesive (Johns et al., 1984). Military cohesion is a condition that causes members of the Armed Forces to subordinate self-interest and conform to standards of behavior dictated by the military (Johns et al., 1984). So while a civilian might have the freedom to attend a university or not depending upon what serves their own self-interest, once a soldier joins a military community, either as an enlisted soldier or as a member of a military academy, he or she is required to subsume their self-interests to support the organization to which they belong.

To that end, military personnel are required to provide service to their elected military branch either until their enlisted time expires, or for a period of years after graduating from a military academy. The United States Military Academy at West Point, for example, requires a minimum of five years of active duty after graduation, and three years in the Inactive Ready Reserve. In this way, the U.S. military population is different in terms of its traits and goals than a civilian population.

Once a person's obligated tenure is served in the military, there are individuals who do opt to continue their membership within the military community. Griffith (1988) examined why individuals are motivated to stay in the military, using sociological analyses of small-group behavior to illuminate this issue. Specifically, Griffith (1988) studied past and current attitudinal measures of morale and cohesion within a military unit. Results from his analyses showed group cohesion in military units is a complex social-psychological construct involving individual characteristics (e.g., pride, commitment, sense of purpose and meaning) and group characteristics (e.g., company combat confidence, leader concern for soldier welfare, confidence in small-unit leaders, mutual trust and caring, social support, cooperation, and teamwork among unit members) (Griffith, 1988).

Arguably, then, when designing motivational feedback messages for a military population, the theoretical constructs used to ground these designs should reflect the unique profile of this particular group. What follows are three theoretical constructs that seem suited to inform the strategies of designing motivational feedback for a military population. These constructs include theories of motivation based on control-value, social identity, and self-efficacy.

Section 7. Theories of Motivation:
Control-Value, Social Identity, and Self-Efficacy

Control-Value Theory

One path to intervening on frustration involves framing feedback messages for a military population within the context of control-value theory (Pekrun 2000, 2006). The objectives of this feedback are to: 1) seek to motivate learners to persist in the learning activity based on an implicit appeal to the learner's perceived controllability of achievement activities and their outcomes; 2) highlight the value and importance of the learning activities and outcomes (Artino, Holmboe, & Durning, 2012).

Control-value theory was developed by Pekrun (2000, 2006) as a comprehensive, integrative approach to understanding emotions in education. When individuals feel in or out of control of achievement activities and outcomes that are subjectively important to them, they experience specific achievement emotions (Pekrun, Frenzel, Goetz, & Perry, 2007). Achievement activities are mediated by emotions that influence cognitive resources, motivation, strategy choices, and intrinsic and extrinsic regulation of learning. The outcome of these achievement activities in turn influences students' emotions (Pekrun, Frenzel, Goetz, & Perry, 2007).

Control-value theory, then, states that student achievement emotions such as frustration can be influenced by changing the student's subjective perception of control and value through a shaping of the learning environment (Pekrun, Frenzel, Goetz, & Perry, 2007; Kim, 2010). By intervening with frustration feedback messages framed within the context of the control-value theory (Pekrun 2000, 2006), the objective is to motivate learners to persist in the learning activity based on an implicit appeal to the learner's perceived controllability of achievement activities and their outcomes, as well as highlighting the value and importance of the learning activities and outcomes (Artino, Holmboe, & Durning, 2012).

Where control appraisals relate to the perceived controllability of achievement activities and their outcomes, value appraisals pertain to the subjective value or importance of these activities and outcomes (Artino, Holmboe, & Durning, 2012). As such, the shaping of the learning environment of learners to be sensitive to the emotional components of learning and performance can be done according to five broad categories: cognitive quality; control and confidence; autonomy support; goal structures; and value (Artino, Holmboe, & Durning, 2012).

Cognitive quality refers to the cognitive quality of instructional activities such as their structure, clarity, and potential for cognitive stimulation (Pekrun and Stephens, 2010). Control and confidence speaks students' perceptions of personal control, where as autonomy support includes meeting the basic psychological need for autonomy (Artino, Holmboe, & Durning, 2012). Goal structures, includes supporting learners' development of behavior as purposeful, intentional, and directed toward the attainment of certain goals (Artino, Holmboe, & Durning, 2012). Lastly, and most pertinent to this literature review, the category of value includes clarifying the importance of specific learning activities and content, including utilizing authentic learning activities. For example, integrating course content with authentic, real-world cases to capture learner's immediate interest as well as foster an appreciation of the broader relevance and importance of the learners are learning (Artino, Holmboe, & Durning, 2012).

The perceived value of an activity in control-value theory is closely related to Keller's (1983) "R" in the ARCS model of motivation, a practical approach to defining motivation that includes attention, relevance, confidence, and satisfaction. For Keller (1983), relevance is a requirement to achieve motivation, and instructors and tutors should demonstrate the relevance of the material to the learner so the learner can perceive a degree of meaningfulness in what is being taught. In a review of the literature, there is one instance of using control-value theory principles to impact affect in an ITS, namely the work of Rebolledo-Mendez, du Boulay, and Luckin (2011), where factual-feedback messages that highlighted value or relevance of the learning tasks were used as part of an experiment using Ecolab II (Rebolledo-Mendez, du Boulay, & Luckin, 2011).

In Rebolledo-Mendez, du Boulay, & Luckin's work (2011), two feedback interventions were designed to address the negative affective states of frustration, anxiety, and boredom, delivered as a form of help in the system: (1) flattering feedback and (2) feedback based on facts. The feedback based on facts delivered text to the learners describing facts relevant to the task yet void of any praise or encouraging messages (Rebolledo-Mendez, du Boulay, & Luckin, 2011). The results showed that the young learners in the factual feedback condition completed a significantly larger number of learning activities than the participants who receive flattering feedback and were more interested in the learning activity (Rebolledo-Mendez, du Boulay, & Luckin, 2011).

While not explicitly linked to the control-value theory, the Rebolledo-Mendez, du Boulay, & Luckin (2011) study of pedagogical tactics to remediate negative affective states lays the groundwork to examine a motivational strategy for feedback messages designed according to control-value theory. Feedback messages designed according to the control-value theory use facts as a form of feedback messages, informing the learner of the value of their learning tasks, pointing out real-world facts relevant to the content being learned.

This articulation of the value and purpose of pursuing goals has been identified by military scholars as a fundamental element of effective in-person leadership that motivates military personnel to accomplish their mission (Shamir, House, & Arthur, 1993). As such, employing motivational feedback messages modeled similarly, but specifically through the lens of control-value theory, could prove effective when implemented in an ITS for a targeted military population, though there is no literature that addresses this approach directly.

Social Identity Theory

A second path for intervening on frustration involves framing feedback messages as a member of a group, for the purposes of this review, specifically as a member of the United States Army. This design capitalizes on Tajfel and Turner's (1979) social identity theory, which states our identities are formed through the groups to which we belong, creating some degree of uniformity of perception and action that exist among group members[3] (Stets & Burke, 2000). Social identity is aligned with the situated social cognition perspective that proposes cognition and action are not discrete entities but dynamically shaped by each other (Schwarz, 2007, 2010; Smith & Semin, 2004, 2007).

Developed by European social psychologists Tajfel and Turner (1979, 1986) social identity theory was developed as a theory of intergroup conflict (Kelly, 2009). In a series of experiments, Tajfel and colleagues sought to assess how, in the absence of a salient identity, individuals responded to group membership (Tajfel, 1970; Tajfel et al., 1971). Essentially, the basis of social identity theory is the idea that behavior exhibited by individuals is motivated in an individual's efforts to maintain a positive social identity within their group of membership

[3] It is important to note that the social identity theory discussed is within the context of an individualistic society, rather then a collectivist society. In a collectivist society, salient social identities evolve from where an individual is born (race, religion), whereas an individualistic society, social identities are marked by their self-elective properties, where one can join or leave a group at will (Oyserman & Destin, 2010).

(Kelly, 2009).

Social identity theory maintains that the self is reflexive and a person can categorize, classify, or name one's self in relation to social categories or classification (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Through this process of self-categorization and identification, one's identity is formed. Social identity theory is chiefly concerned with the idea that a person's social identity is in relation to maintaining membership in a group (Hogg & Abrams, 1988). In addition to self-categorization, social comparison is another element that constitutes social identity formation. Social comparison is the process through which a person accentuates those dimensions of one's self that will result in enhancing one's positive in-group judgment (Hoggs & Abrams, 1988).

Social identity theory posits that a person's sense of "simpatico" with their self-elected group defines one's sense of self, so that people who derive a strong sense of identity from their group are more likely to see things from the perspective of the group (Hogg & Abrams, 1988).

Cognitive manifestations of social identity include a person identifying themselves as prototypical of a group (Oakes, Haslam, & Turner, 1994). Attitudinal manifestations of social identity include instances when people make uniformly positive evaluations of the group of which they are a member, leading to a commitment to remain in the group even if a group's status might become diminished (Elemers, Spears, and Doosje, 1997). Behavioral manifestations of social identity includes using a group label to describe oneself, actively distinguishing one's self from people from outside their group, and concurrence in decision making conditions (Turner, Pratkanis, Probasco, & Leve, 1992).

Conformity, a hallmark of social identity, occurs when people behave according to expectations tied to one's social membership (Thoits & Virshup, 1997), usually as a way to maintain affiliation and fulfill self-concept goals (Cialdini & Goldstein, 2004). Thus, in order to maintain one's perceived membership in a group, individuals will often take actions to ensure conformity to group standards (Burke & Stets, 1999).

Noting the influence of social identity on behavior, a study by Oyserman and Destin (2010) examined the effect of providing at-risk children in school with identity-based motivation (IBM) interventions to effect a positive change in school outcomes. The IBM model articulated by Oyserman and Destin (2010) builds on the notion that identities are dynamically constructed in context. As such, Oyserman and Destin (2010) showed that when situations and difficulties are framed in a way that is identity congruent, interventions containing information about group norms and expectations generate identity-behavior links.

The results of Oyserman and Destin's (2010) study demonstrated that IBM interventions produced lasting positive change in school outcomes among children at risk. With ten sessions of interventions conducted over a seven-week period of time, researchers sought to shape children's future identities to include the notion that failures along the way in school were a normal part of a school-focused identity. Tracking the control and intervention students through two academic years, the researchers found their IBM interventions had significant direct effects on academic outcomes (Oyserman & Destin, 2010).

While not specifically identifying that the interventions mitigated against frustration, session 8 through 10 of the intervention sessions did focus on having students interpret and identify difficult experiences in school, including everyday, social, and academic problems: "The meta-theme was that all students care about these issues, that difficulties along the way are normative. This interpretation of difficulty implies that school-focused identities are important, not impossible," (Oyserman & Destin, 2010, p. 1028). As such, the IBMs were used to help the children identify themselves within the group of normal, achieving students, and link this identity to behaviors that included persistence in the face of difficulties.

Although the Oyserman and Destin (2010) study addresses using social identity to motivate *children* to persist in difficult circumstances, social identity theory has also been used to motivate *adult* human-to-human training to shape behavior and decision-making, including attitudes and value-orientations – particularly in the education and training of military cadets at West Point (Franke, 1997; Franke, 2000).

Social identification in the military has attracted interest since the 1960s, with more recent research examining the effects of collective identity, emphasis on shared values, and inclusive behaviors between staff members and soldiers (Shamir, Brainin, Zakay, & Popper, 2000). Further, positive correlations between a soldiers' identification with the army and evaluation of platoon leadership has been demonstrated (Mael & Alderks, 1993), as well the relationship of social identity with the Army and issues of attrition amongst new Army recruits (Mael & Ashforth, 1995).

As a way to capitalize on social identity in issues of leadership and management, Shamir, House, and Arthur (1993) have argued that leaders strengthen social identification through use of cultural symbols such as slogans, symbols, rituals, and ceremonies that highlight collective identity, superiority, and uniqueness. In a study by Shamir, Brainin, Zakay, and Popper (2000), leaders' behaviors were found to be positively correlated to social identification among unit members, particularly when an emphasis was placed on shared values and inclusive behaviors, mediated by the amount of cultural symbols in a unit and the social identification of staff members (Shamir, Brainin, Zakay, and Popper, 2000).

In terms of using social identity motivational feedback messages in tutoring systems, however, there seems to be a gap in the literature; this strategy has not been explicitly used or modeled in ITS or related systems for either a civilian or military population. Nevertheless, as tutoring systems have historically been modeled on the human mentor model (Merrill, Reiser, Ranney, & Trafton, 1992), it seemed reasonable that this study should examine motivational intervention feedback messages that capitalized on the a cadet's social identification with their elected group membership, namely, the US Army.

Social Identity and presence. For the purposes of this dissertation, measurements of presence was taken using the Presence Survey (Witmer & Singer, 1994, 2005) to determine whether there was an effect of presence on pre-post test scores across conditions. The Presence Survey is a multidimensional measurement devise to assess a participants' sense of presence while engaged with a virtual environment, and it provides a measurement of the participant's perceived level of realism – an element identified as needed for adequate learning and transfer (Witmer & Singer, 1994, 2005). In terms of reliability of the instrument, internal consistency measures (Cronbach's Alpha) yielded reliabilities of 0.75 and 0.81, and consistent positive correlations between the degree of presence and virtual environment task performance has established content validity (Witmer & Singer, 1998).

The objective of taking this measure was to determine whether presence would have a mediating effect across conditions on learning outcomes, capitalizing on the notion that if a participant experienced a sense of realism in the gaming-environment and identified with their avatar in the game as a combat soldier, they might be more receptive to motivational feedback messages that target their social identity.

Self-Efficacy Theory

The third theory of motivation that was chosen to develop motivational feedback messages involved creating messages framed by the theory of self-efficacy. Self-efficacy includes how the learner seems themselves as an individual, and their ability to succeed in a task if they persist (Bandura, 1986). Self-efficacy is known to correlate positively to academic performance and persistence rates (Bong, 2001; Kaun & Nauta, 2001; Multon, Brown, & Lent, 1991; Wood & Locke, 1987).

Bandura's (1986) socio-cognitive perspective on the role of self-efficacy theorizes that individuals are proactive and self-regulating. Bandura's social cognitive theory (1997, 2002) notes that perceived self-efficacy influences a person's motivation for tasks, actions towards goal achievement, perseverance on tasks, and responses to failures. Bandura's view highlights the notion that individuals possess self-beliefs that enable them to exert control over thoughts, feelings and actions, where behavior is inextricably linked to the beliefs that people have regarding their capabilities (Bandura, 1995). Indeed, predictions about behavior can be more accurately assessed according to an understanding of an individual's self-perception in what they are capable of accomplishing, or what Bandura termed self-efficacy (1986). Bandura (1977, 1997) defined self-efficacy as an individual's judgments pertaining to one's abilities to make decisions and implement a course of action to obtain a goal (Zimmerman, 2000).

Graham and Weiner (1996) note that perceptions of self-efficacy have been correlated to academic achievements and theories of academic motivation, where a student's beliefs about one's self and abilities compose the principle element of academic motivation. These self-beliefs are grounded in the idea that the view students develop and adopt about their personal efficacy is a vital force in academic success or failure. In short, a person's self-efficacy is evidenced in choices they make, efforts expended, persistence, and perseverance (Bandura, 1986).

Self-efficacy has been further situated within a larger theory of personal and collective agency that operates with other socio-cognitive factors in regulating attainment and well-being (Bandura, 1997), and self-efficacy measures have largely focused on performance capabilities (Zimmerman, 2000). Respondents to self-efficacy measures judge their capabilities to fulfill a given task demand as opposed to how they feel about themselves, and are assessed before performance on a relevant activity (Zimmerman, 2000).

Bandura (1997, 2002) identifies four paths towards creating strong self-efficacy perceptions: performance accomplishment, verbal persuasion, emotional arousal, and vicarious experience. Verbal persuasion occurs when leaners are persuaded to believe they have the necessary skills to succeed (Bandura, 1994). For the purposes of this dissertation work, supporting a learner's sense of self-efficacy was addressed by using verbal persuasions. This approach was chosen based on prior research showing that audio messages of a tutoring system can have positive and meaningful impact on student engagement and learning (Grafsgaard et al., 2014, Vail et al., 2014).

In sum, self-efficacy motivational feedback messages in tutoring systems have been used in ITSs, as seen in the work of Burleson's (2006) experiment that included self-efficacy feedback messages for 11-13 year olds: "Just remember that all of your effort does help you learn," (Burelson, 2006, p. 131]). However, there is a gap in the literature in studying how self-efficacy based feedback messages compare to feedback messages based on other theories of motivation, how self-efficacy based feedback messages impact learning gains for a military population, and whether levels of grit moderate outcomes of motivational intervention responses based on self-efficacy.

Self-efficacy and grit. It is important to note that self-efficacy is distinct from other constructs such as outcome expectancies, self-concept, and perceived control (Shell, Murphy, & Bruning, 1989; Hattie, 1992; Pajares & Miller, 1994). Convergent validity has been demonstrated in how self-efficacy beliefs influence academic motivation in regards to choice, level of effort, persistence, and emotional reactions (Bandura, 1997; Bandura & Schunk, 1981).

Recent research has shown a positive correlation between measures of self-efficacy and the measures of grit (Slack, 2014), though these two measures are distinct from each other. In a study by Slack (2014), results revealed that students' score on the Short grit Scale (Duckworth & Quinn, 2009) were positively correlated to the subscales of the Academic Self-Efficacy Scale (ASES) based on Bandura's (1997, 2002) social cognitive theory. Grit has been defined as resilience and effort in the face of failure that may require an extensive period of time to overcome (Duckworth & Quinn, 2009; Perkins-Gough, 2013).

The eight-item Short Grit Scale (see A: SHORT GRIT SCALE) (Duckworth and Quinn, 2009) measures "trait-level perseverance and passion for long-term goals" (p. 166). This instrument has been shown to predict achievement in avocational, academic, and vocational domains (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009; Duckworth, Quinn, & Seligman). Comparing this to self-efficacy, Bandura (1994) has maintained that persistence, often encouraged through verbal persuasion, leads to perceived self-efficacy. However, where grit seems to be used as a predictor of long-term success, self-efficacy seems to measure an individual's belief in their ability to succeed in the present moment.

A scale comparable to ASES is the General Self-Efficacy Scale (GSES) (see Figure 1 for excerpt of GSES) (see APPENDIX A: GENERAL SELF-EFFICACY SCALE for full survey) (Schwarzer & Jerusalem, 1995). The GSES is an instrument that focuses on perceived self-efficacy relating to *effort, persistence, and goal setting*.

Measure		
	1	I can always manage to solve difficult problems if I try hard enough.
	2	If someone opposes me, I can find the means and ways to get what I wa
	3	It is easy for me to stick to my aims and accomplish my goals.
	4	I am confident that I could deal efficiently with unexpected events.
	5	Thanks to my resourcefulness, I know how to handle unforeseen situatio
	6	I can solve most problems if I invest the necessary effort.
	7	I can remain calm when facing difficulties because I can rely on my copi abilities.
	8	When I am confronted with a problem, I can usually find several solution
	9	If I am in trouble, I can usually think of a solution.
	10	I can usually handle whatever comes my way.

Figure 1. Excerpt of GSES, (Schwarzer & Jerusalem, 1995).

It has been used internationally for over two decades, and in samples from 23 nations has a Cronbach's alpha ranging from 0.76-0.90 with the majority in the high 0.80's (Schwarzer & Jerusalem, 1995). When comparing the GSES (Schwarzer & Jerusalem, 1995) to the Short Grit Scale (Duckworth and Quinn, 2009), there seem to be an overlap between persistence and effort self-assessment items.

For example, in the 8-Item Grit Scale the second item reads: "Setbacks don't discourage me," (see Figure 2), which is comparable to the first entry on the General Self-Efficacy Scale (GSES) of Schwarzer & Jerusalem (1995): "I can always manage to solve difficult problems if I try hard enough." However, a closer examination of these instruments yields a distinction between self-assessments regarding present and future beliefs vs. self-assessments of patterns of prior behavior: GSES present and future beliefs ("*If* someone opposes me", "I *could* deal efficiently", "how to handle *unforeseen* situations") (see Table 1) vs. the Grit Scale patterns of past behavior ("I *have been* obsessed", "I *often* set a goal", "I *finish whatever I begin*") (see Table 2).

3. I have been obsessed with a certain idea or project for a short time but later lost interest.*
   - Very much like me
   - Mostly like me
   - Somewhat like me
   - Not much like me
   - Not like me at all

4. I am a hard worker.
   - Very much like me
   - Mostly like me
   - Somewhat like me
   - Not much like me
   - Not like me at all

5. I often set a goal but later choose to pursue a different one.*
   - Very much like me
   - Mostly like me
   - Somewhat like me
   - Not much like me
   - Not like me at all

6. I have difficulty maintaining my focus on projects that take more than a few months to complete.*
   - Very much like me
   - Mostly like me
   - Somewhat like me
   - Not much like me
   - Not like me at all

Figure 2. Excerpt Short Grit Scale, (Duckworth & Quinn, 2009).

Consequently, while Slack (2014) has demonstrated a positive correlation between the constructs of self-efficacy through the ASES and the Short Grit Scale, there have been no studies to date that have examined the correlational relationship between GSES and the Short Grit Scale. Irrespective of this gap in the research, however, it is important to note that grit is considered a stable personality trait (Duckworth, Peterson, Matthews, & Kelly, 2007) while self-efficacy is considered to be more transitory (Bandura, 1992). This distinction is seen most clearly in the GSES instrument that measures self-assessment of beliefs about present and near-future competencies, while the Short Grit Scale measures self-assessment of patterns of past behavior that are in turn used to predict future behavior.

For the purposes of this dissertation, while the language of self-efficacy was employed in motivational feedback messages for one of the intervention conditions, the Short Grit Scale was administered to participants to take measures of the personality trait of grit. In part, this was to see whether there would be an impact on the direction of learning gains in the self-efficacy condition, as Slack (2014) had previously demonstrated a positive correlation between self-efficacy and the Short Grit Scale. This work also had the goal of determining whether grit functioned as a moderator of learning across motivational feedback conditions.

Section 8. Summary

This chapter reviewed literature issues of motivation, affect, and cognition, the issues of addressing frustration in a affect-sensitive intelligent tutoring system, the complexity of frustration in terms of its relationship to cognition and learning, addressing unique trait characteristics of a military population relevant for designing motivational messages, and three theories of motivation (control-value theory, social identity theory, and self-efficacy theory) that were targeted to design feedback message interventions.

Overall, the review of the literature indicates that designing motivational feedback for an intelligent tutoring system (ITS) is a complex endeavor. Messages should be designed to encourage the learner to persist through their frustration to accomplish their learning goals, and in a manner that is timely, i.e., delivered on the detection of high frustration, and not during brief moments of frustration.

Further, these messages need to be designed with the target population in mind, taking into consideration what the individual -- as part of a larger, distinct community -- may respond to and find meaningful. Employing motivational theories as a framework in which to design feedback messages builds on prior affect detection and feedback research. However, this review illuminates a gap in the literature as to which motivational theory would be the most effective theory to design an intervention feedback message for a military population that address high frustration while engaged in a simulated training video game delivered by intelligent tutoring system.

Chapter III: BACKGROUND

Section 1. Overview

This chapter presents the research design and results of a prior, baseline study conducted in September 2013 to collect data on affect and behavior while participants engaged in the simulated training video game, vMedic, that is part of the Tactical Combat Casualty Care (TCCC) Training developed by the US Army (Sotomayer, 2010). vMedic was deployed to participants through the Generalized Intelligent Tutoring Framework (GIFT), . This data was used to then identify frustration as the affect most significantly related to negative learning gains within this context. Subsequently, sensor-free affect detectors were built around the model of frustration and later embedded in the Generalized Intelligent Tutoring Framework (GIFT) to enable detection of high frustration of participants while playing vMedic and upon this detection, for GIFT to deploy intervention messages while participants' played vMedic in the main, dissertation study conducted in September 2015.

Section 2. Background and Description

The September 2013 study was conducted at the United States Military Academy at West Point (USMA) to investigate the correlation between affect, behavior, and learning in the vMedic video game (also known as TC3Sim) as delivered within a course for combat medical care delivered by the Generalized Intelligent Framework for Tutoring system (GIFT), a modular and service-oriented, reusable tutoring tool (Sottilare, Goldberg, & Holden, 2012). This September 2013 study yielded important results regarding of the relationship between frustration and learning outcomes, as well as provided the baseline data to create the sensor-free affect detectors for frustration that were used for this current dissertation study.

Description of September 2013 USMA Study and GIFT

The September 2013 USMA study investigated the relationship between affect and learning for a subset of the West Point cadet population using vMedic, a combat medical training video game, which is part of a larger tutoring course used by the US Army to provide training in tactical field care and care under fire. The September 2013 study's data was used in the development of affect detectors for the computer system detection of frustration that was subsequently embedded into GIFT (Baker, DeFalco Ocumpaugh, & Paquette, 2015), and used for this dissertation study to high detect frustration.

GIFT and vMedic. The Generalized Intelligent Framework for Tutoring (GIFT) developed by the Army Research Laboratory in Orlando, FL, is a modular and service-oriented platform that contains the components of sensor, trainee, pedagogical, learning management system, and domain modules. GIFT is representative of a reusable tutoring tool (Sottilare, Goldberg, & Holden, 2012).

vMedic, also known as TC3Sim, is a serious game and training simulation developed for the U.S. Army by ECS Orlando (Sotomayer, 2010). A serious game has been defined as a game in which education rather than entertainment is the primary goal, (Michael & Chen, 2005). Abt (1987) described the objective of serious games as a way to get players to learn something while possibly, though not necessarily, having fun while doing it. Further, Abt (1987) notes that serious games are highly motivating because they give dramatic representations to subjects or problems studied, and allow for students to assume roles, formulate strategies, engage in decision making, and get immediate consequential feedback from actions taken without real world errors or costs (Abt, 1987). And it is this -- the affordances of engaging in real world problems and decision making without incurring real world damage or costs -- that have made simulating training experiences a valuable training tool for the US Army. Specifically, the *Army Learning Model 2015* (TRADOC, 2011) required an increase in the use of technology-driven platforms including gaming environments in their efforts to develop a new learning model for future training systems. vMedic fits well within that paradigm.

As such, vMedic has been integrated into GIFT through its Domain module function (Sottilare, Goldberg, & Holden, 2012), and can be imported into any course built through the GIFT platform. Figure 3 below illustrates the domain modules of GIFT and how they interact to receive and pass information to the learner/participant.

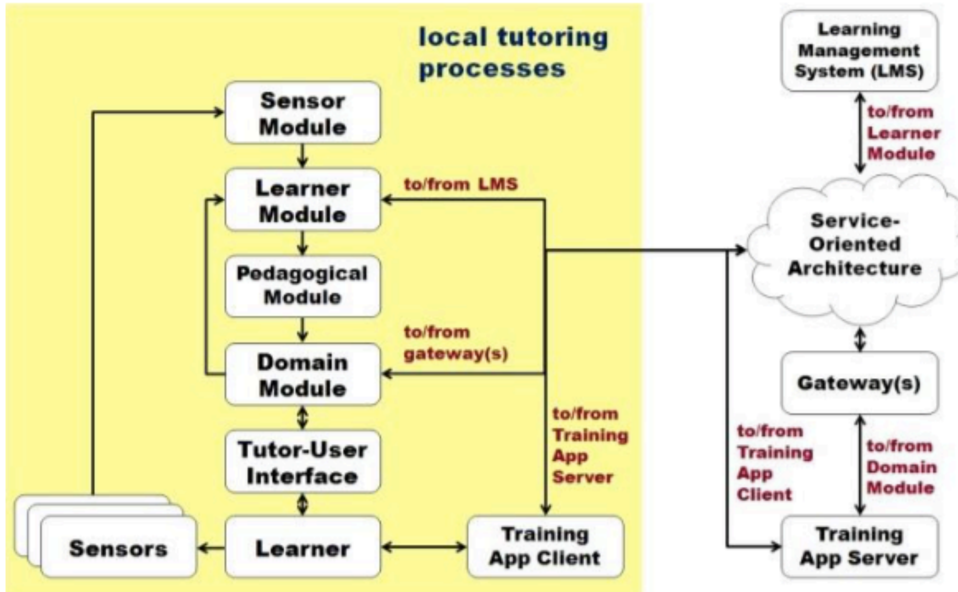


Figure 3. GIFT and vMedic integrated as an instructional framework (Sottilare, Goldberg, Brawner & Holden, 2012).

vMedic teaches combat medic competencies including assessing casualties, performing triage, providing emergency treatment, and evacuating a casualty from a battlefield, and is representative of a reusable tutoring tool. The interactions of the participant in vMedic are recorded in log files within GIFT and can be later extracted for analysis and used in combination with pre-post test measures, and field observations using the HART application for the Android that records affect and behavior using the Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) (Ocumpaugh, Baker, & Rodrigo, 2015).

The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP). The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) is a protocol for Quantitative Field Observations (QFOs) of student affect and behavior. First developed ten years ago by Dr. Ryan Baker and Dr. Ma. Mercedes T. Rodrigo, BROMP is a holistic coding procedure that has been used in thousands of hours of field observations of students from kindergarten to undergraduate populations. BROMP-trained coders must be certified to achieve inter-rater reliability of 0.6 (Cohen's Kappa) or better before being allowed to conduct field observations in research. As the primary coder for the September 2013 study, I had been trained five months prior to the 2013 study under Dr. Jaclyn Ocumpaugh and subsequently certified by Dr. Ryan Baker.

BROMP has become an increasingly common practice for field observations related to interaction-based detection of affect (Baker et al., 2012; Baker, DeFalco Ocumpaugh, & Paquette, 2015; Pardos et al., 2014) and has been used for several years in educational settings to study behavior and affect (Baker, D'Mello, & Rodrigo, 2010; Baker et al., 2012; Rodrigo & Baker, 2009). Whereas in typical qualitative field observations the researcher has to devise a method of recording and appropriate codes for their observations, in the BROMP protocol affect and behavior are coded using the Human Affect Recording Tool (HART) developed for the Android platform (Baker, & Rodrigo, 2012; Ocumpaugh, et al., 2015). The HART application for this study was pre-programmed with affect and behavior categories created specifically for the cadet population that was to be observed. The categories for affect included: frustration, confusion, engaged concentration, boredom, surprise and anxiety. The codes for behavior included: on-task, off-task behaviors, Without Thinking Fastidiously, and intentional friendly fire.

For the September 2013 study, cadets were observed individually, in a pre-determined sequence. Affect and behavior recorded using the HART application HART requires a strict coding order determined at the beginning of each session, and coders trained to rely on peripheral vision and side-glances to minimize observer effects code learners individually. Coding must take place within the first 20 seconds of the observation, during which time the coder has to categorize each trainee's behavior and affect, recording the first observable behavior and affect.

Sensor-free detectors. The purposes of conducting the BROMP observations was to obtain ground-truth labels of affect that would be used to create the sensor-free affect detector for frustration. Sensor-free detectors are computational models that automatically detect learners' affective states. Powerful tools for investigating the interplay of affect and learning used for over the past decade, these affect detectors recognize learners' affective states at run-time using behavior logs and sensor data. As such, these computation model detectors are used to detect affect differently than human detection, and in some cases, render results superior to human detection (Hoque, McDuff, & Picard, 2012)[4].

The results section of this chapter explains how the affect detector models were constructed and how they are used to detect frustration.

[4] Hoque, McDuff, and Picard (2012) from the MIT Media Lab discovered not only that smiling is quite common when test subjects are frustrated, but also that software is better than humans at differentiating frustration smiles from happiness smiles. Indeed, for frustrated smiles, humans performed below chance, whereas computer algorithms were able to detect frustration better than 90 percent.

Section 3. Method

Design

There was only one condition in this initial, baseline study. All participants filled out demographic information, took a pre-test, and went through a modified version of a PowerPoint describing hemorrhage control, tactical field care, and care under fire. After the PowerPoint, participants engaged in a tutorial in vMedic and three training scenarios. Following vMedic, participants took the post-test, and a survey.

Participants

The population of interest for the September 2013, baseline USMA study included future Army officers who will likely engage with training systems with intelligent tutors in future military training conditions. Over a five-day period at USMA, 161 cadets participated in the study (83% male, 17% female). These cadets were predominantly first year students enrolled in USMA's PL100 Intro to Psychology course and recruited through West Point's SONA System. The age of the cadets ranged between 18-22.

In total, there were 161 participants, though only 127 participants' data could be used due to an error in consent forms that occurred during the experiment. The error in consent form occurred when the project manager of this study from the Army Research Lab, Dr. Keith Brawner, printed out the wrong consent forms and handed these consent forms to the participants before this baseline study. The consent forms printed were textually identical to the correct consent forms approved by the TC and West Point IRBs; however, the official stamp was missing. This error was not detected until months after the experiment, at which time participants were re-consented by the research psychologist at West Point, Vasiliki Georgoulas, as per the instructions of the West Point IRB office. Once participants were re-consented, the IRB allowed us to continue to analyze the data yielded from this baseline study.

The lab space was located in the basement of Thayer Hall at USMA and was arranged to run ten participants at a time.

Experimenters

There were approximately five to seven members providing support, proctoring, or conducting live observations at any one time. Those working on this 2013 study included Dr. Ryan Baker and the current author from TC, Dr. Jonathan Rowe and Dr. James Lester of NCSU, and Dr. Robert Sottilare and Dr. Keith Brawner of ARL.

As a certified BROMP coder (see section 2 of this chapter for an explanation of the BROMP certification process), the current author conducted field observations according to the BROMP methodology (Ocumpaugh, Baker, & Rodrigo, 2015) using the HART application on the Android hand held smartphone, as well as trained Dr. Rowe in the BROMP protocol. Dr. Rowe, once trained and subsequently certified, conducted field observations in tandem with me during later sessions of the study.

Experimental test bed

Experimental tasks. The GIFT platform manages the experimental tasks that include pre-tests for prior knowledge, questionnaires, surveys, training courseware via a PowerPoint presentation, the vMedic training scenarios, and administering a post-test (see Table 1 for a timeline of experimental procedures). As GIFT executed this experimental procedure, there was no other interaction needed between proctor and participant for the duration of the experiment, unless there was a technical error such as a system crash that would require a manual system reboot by the proctors.

The tasks in this experiment were related to standard, simulated combat medic training. The training materials provided to the participants pertained to the knowledge and procedures on hemorrhage control during care under fire and tactical field care, components of the Tactical Combat Casualty Care (TC3) Training developed by the US Army (Sotomayer, 2010).

Training materials. All participants received the same training in this study. Participants were presented a PowerPoint presentation on declarative and procedural knowledge pertaining to hemorrhage control, care under fire, and tactical field care that was modified from a previously developed from a TC3 training program (see Figure 4). The PowerPoint had text, audio, and pictures related to the aforementioned domain.

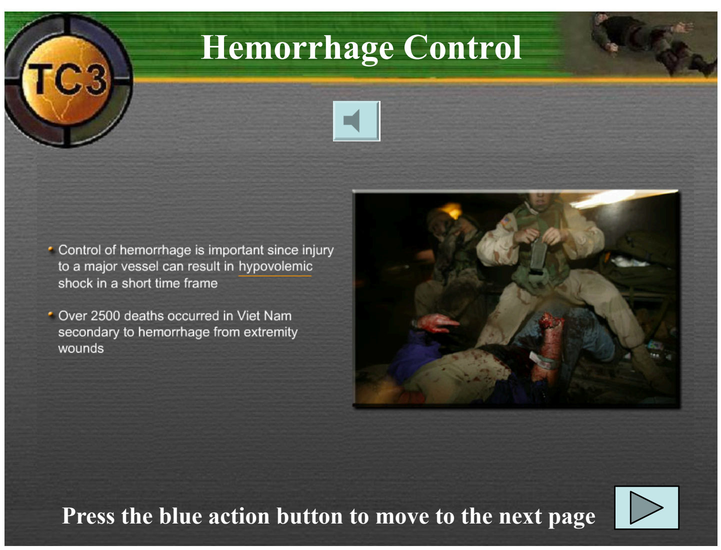


Figure 4. Example of PowerPoint used during September 2013 study

Following the PowerPoint, all participants went through a tutorial on the game controls of vMedic (see Figure 5).

Figure 5. Screenshot of beginning of vMedic

After the tutorial in vMedic, the participants went through four scenario-based training events.

Questionnaire and surveys. Prior to logging on to the laptops to begin the experiment, participants were given a paper consent form by Dr. Keith Brawner that explained the study and provided the participants with an opportunity to withdraw from the study without penalty. Once these consent forms were signed, participants were instructed to log on as a new user within GIFT to begin the experiment.

There was a demographic questionnaire administered once a participant logged into GIFT to collect data on age, sex, education level, and computer game experience (see APPENDIX C). This demographic questionnaire was based on items used in prior experiments that related to combat emergency medical care (Goldberg, 2013; Carroll et al., 2011; Sottilare, Brawner, & Holden, 2011).

Following the demographics questionnaire, the Presence survey (Witmer & Singer, 1994) was administered. This survey is a multidimensional measurement devised to assess a participants' subjective experience related to a target activity in a laboratory experiment The Presence survey was administered through GIFT after participants were engaged with the serious game, vMedic. This survey collected data on participants' inclination to experience a sense of presence while engaged with a mediated environment.

Performance metrics. The performance measures included the learning gains based on the administered pre- and post-tests assessing knowledge levels in hemorrhage control. Log file data that captured the actions of participants when they interacted with vMedic was extracted from GIFT. Lastly, affect and behavior measures were collected as recorded via BROMP (Baker & Rodrigo, 2012; Ocumpaugh, Baker, & Rodrigo, 2015) through the HART application deployed through an Android phone.

Field observations using BROMP 1.0 (Baker & Rodrigo, 2012) were conducted in a pre-chosen order to balance observations across trainees and to avoid bias towards more noteworthy behaviors or affect. Observations were used using quick side-glances in order to be less obtrusive during observations. Coding included recording the first behavior and affect displayed by the participant within 20 seconds of the observation, choosing from a predetermined coding scheme. As previously mentioned, the affect coded included: frustration, confusion, engaged concentration, boredom, disdain, and surprise. Behavior coding included: on-task behavior, off-task behavior, intentional friendly fire, and WTF ("without thinking fastidiously") behavior, where the participant's actions have no relation to the scenario (Sabourin, Rowe, Mott, & Lester, 2011).

Procedures

Pre-test, surveys, training, post-test. After reading and signing the paper consent forms, participants were fitted with Q-sensors and positioned so that the Kinect depth sensors could record depth-map images to support recognition of postural positions. Q-sensors are wristbands that are detect and measure slight electrical changes to the skin that result from stress and excitement (Engadget, 2010). The Kinect is a motion sensor that is equipped with a depth sensor image that can capture physiological and behavioral manifestations of emotion. This data can later be extracted to build sensor-models that can automatically detect learner affect using a range of physical indicators including posture (Paquette et al., 2015).

Once the sensors were affixed and centered, participants logged into GIFT using a unique ID based on laptop station and time/day of experiment. After logging in, GIFT managed all experimental procedures and sequencing of the experiment, and the interaction logs of all participants were backed up onto hard drives.

Once the participants were logged into GIFT, GIFT prompted participants to complete the demographic questionnaire and a pre-test (see APPENDIX D: KNOWLEDGE PRE-TEST and POST-TEST 2013). When completed, the participant was presented the course materials on hemorrhage control, care under fire, and tactical field care through a PowerPoint presentation.

Following this tutorial, the participants played the vMedic game that was launched within GIFT. Participants engaged in a tutorial session for about three minutes to learn the interface of vMedic and general navigations through the vMedic simulated environment. Next, participants played four scenarios, each targeting a different wound (arm, leg, chest, multiple wounds). For the purposes of eliciting a strong emotional response, the final scenario was designed by the principal investigators of this study so that no matter what actions the participant took the injured soldier would expire. This scenario was coined *Kobayashi Maru* after a “no-win” training exercise to test cadets and purposefully elicit frustration (named after the film *Star Trek II: The Wrath of Kahn*).

After participating in these scenarios, the participants took the Presence survey (see APPENDIX E), and a final post-test (see APPENDIX D), and logged out of the system. Once logged out, the sensors were removed from the participants and participants were debriefed outside the testing room.

Table 1

Timeline for Experimental Procedures for Baseline September 2013 Study

05:00 - 10:00 minutes; Consent form given to participant. Participant signs paper form and logs into GIFT. Demographic survey launched by GIFT. This process took about 5 minutes to complete.
10:00 - 20:00; After the demographic survey, GIFT presented a ten-item pre-test on hemorrhage control on the hemorrhage control, tactical field care, and care under fire. The pre-test will be used to determine the effectiveness of the hemorrhage control course. This entire process took approximately 10 minutes to complete.
20:00 - 25:00; After the pre-test, GIFT launched a PowerPoint presentation on hemorrhage control, tactical field care, and care under fire. It is estimated participants spent an average of 5 minutes with this content.
25:00 - 35:00; Following the PowerPoint, GIFT launched vMedic and participants had a tutorial on the game controls. This introduction reviewed interface components and allow participants to interact with environment elements prior to the start of the scenario-based training event. Following the tutorial, the participant went through three scenarios and applied what they knew of hemorrhage control, tactical field care, and care under fire in the serious video game vMedic. This lasted approximately 10 minutes.
35:00 - 50:00; After completing the three scenarios in vMedic, GIFT launched the ten-item post test, and once that was completed, the Presence questionnaire (Witmer & Singer, 1994). This took approximately 15 minutes.
50:00 - 55:00; At the completion of the Presence questionnaire, a message was displayed telling the participant that the study was over and they should notify one of the proctors. Once the proctor removed the Q-sensors and insured the participant had logged out of the study, they were brought out of the lab into the hall for a debriefing, including an opportunity to ask any questions about the study. This took approximately 5 minutes.

Section 4. Analysis and Results

Pre- and Post-Test Analysis

Students' answers on the pre and post-tests were extracted from the log files using GIFT's Event Reporting Tool (ERT). Each test was composed of 10 multiple-choice questions, and graded on a scale from 0 to 10 with each correct answer worth 1 point. Test answers were collected for 107 students. Out of those, 7 students did not have post-test answers.

Pre-test scores ranged from 4 to 10 with a mean score of 7.45 and a standard deviation of 1.15. Post-test scores ranged from 4 to 10, with a mean score of 7.39 and a standard deviation of 1.14. On average the difference between the post-test scores and the pre-test scores was -0.08, meaning that, on average, students performed 8% worse on the post-test than on the pre-test. This suggests that the tests may not have been aligned completely with the learning that occurred in vMedic, and that students may have been somewhat disengaged while taking the exam itself.

A paired t-test did not show any significant difference (t (1,99) = 0.783; p = 0.436) between the pre-test (M = 7.47; SD = 1.09) and post-test scores (M = 7.39; SD = 1.14) for the 100 students that completed both tests. The effect size for this test was very small (d = 0.07), also suggesting that there were no learning gains between on the pre- and post-tests.

Table 2

Paired T-Tests of Pre and Post Test For All Participants, One Condition (n =100)

Measures	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$	[SD]	$t_{(1,99)}$	Sig.	d
All participants	7.45	[1.15]	7.39	[.185]	-.08	[1.14]	0.783	.436	0.07

Background on Affect Detectors

After the data was collected, Dr. Ryan S. Baker and Dr. Luc Paquette built sensor-free affect models to detect frustration and integrated these detectors into GIFT (Paquette et al., 2015). In terms of interactive affect detectors, sensor-free models have been developed that can infer confusion (Pardos, Baker, San Pedro, Gowda, & Gowda, 2014; Baker et al. 2012), boredom (Baker et al., 2012; D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2008; Sabourin, Rowe, Mott, & Lester et al., 2011), frustration (Baker et al. 2012; Paquette et al., 2014), and engaged concentration (Baker et al., 2012, D'Mello, Craig,Witherspoon, McDaniel, & Graesser, 2008; Sabourin et al., 2011), among other constructs. In addition, there has been prior work assessing and addressing learner frustration with sensors within ITSs, which has included the use of galvanic skin response and blood volume pressure (Fernandez & Picard, 1998); the use of a pressure sensitive mouse (Qi et al., 2001); haptic sensors (McLaughlin et al., 2004); and multi-modal sensors (Mota & Picard, 2003; Haro et al., 2000; D'Mello et al., 2005). Using multimodal sensors, D'Mello and colleagues (2005) were able to detect a student's affective state 42% of the time, and detected frustration 78% of the time. Kappor et al. (2007), also using multimodal sensors, could identity frustration with an 79% accuracy.

Sensor-free automated detectors have the ability to scale to retrospective log files, and have thus been used to conduct basic research on the conditions and impacts of engagement and affect on learners, including research on the relationship between these constructs and learning (Baker, Corbett, & Koedinger, 2004; Cocea, Hershkovitz, & Baker, 2009; Pardos et al., 2014), student goal orientation (Baker, 2007; Baker et al., 2008; Hershkovitz, Baker, Gobert, Wixon, & Sao Pedro, 2013), and student attitudes towards mathematics (Arroyo et al., 2009; Baker, 2007; Baker et al., 2008). Once a detector is developed and validated, it can be applied at scale to additional data, potentially enabling very large-scale analyses (see discussion in Hollands & Bakir, 2015).

Building Interaction-Based, Sensor-Free Affect Detectors

The interaction-based, sensor-free affect detectors were built using the baseline September 2013 data that included data from the GIFT logs of learner interactions with vMedic, as well as the affect data collected through BROMP field observations (Baker, DeFalco, Ocumpaugh, & Paquette, 2015). The on-location affect data labels were obtained by using Quantitative Field Observations (QFOs), using BROMP (Ocumpaugh, Baker, & Rodrigo, 2015).

In total, 3066 BROMP observations were collected by the current author and Dr. Rowe of NCSU, starting our observations from the time the cadets logged into the GIFT system until the cadets logged out of the system (Paquette et al., 2015). The results of those observations included: 735 (97.35%) coded as the cadet being on-task; 19 (2.52%) as off-task; 1 (0.13%) as Without Thinking Fastidiously; and 0 as intentional friendly fire. For affect, the results included 435 observations (57.62%) coded as concentrating; 174 (23.05%) as confused; 73 (9.67%) as bored; 32 (4.24%) as frustrated[5]; 29 (3.84%) as surprised; and 12 (1.59%) as anxious (see Table 3).

[5] The observed frustration measures were comparable to other BROMP frustration observations detected in prior research: 5.4% observed frustration with 8th grade students using ASSISTments program (Pardon et al., 2014); 4-6% observed frustration in three distinct computer-based learning environments with undergraduate students (Baker, D'Mello, Rodrigo, & Graesser, 2010).

Table 3

Mean percentages of observed affect September 2013 study

Affect	Mean percentage of observations
Concentrating	57.62
Confused	23.05
Bored	9.67
Frustrated	4.24
Surprised	3.84
Anxious	1.59

Participants' actions within the vMedic log files were pulled from GIFT and synchronized to on-location affect BROMP field observations collected using the HART application. This was done to generate training data for the interaction-based affect detectors. Features were then generated to summarize the behavior of students and the current state of the vMedic game. Then, machine learning was used to create detectors that identify the relationship between the best combination of features and the observed frustration (see Table 4).

Table 4

Features that indicated frustration actions within vMedic

How many times blood pressure changed in the last 20 seconds
The maximum value of the heart rate of the casualty in the last 20 seconds
How many times participant conducted blood sweep actions on the casualty in the last 20 seconds
How many times the participants' avatar was out of cover during hostile enemy fire in last 20 seconds
How many times participants dragged an injured casualty out of hostile crossfire in last 20 seconds

Detectors were built separately for each affective state by Dr. Luc Paquette. Dr. Paquette validated each detector using 10-fold participant-level cross-validation[6]. In this process, the participants are randomly separated into 10 groups of approximately equal size and a detector is built using data for each combination of 9 of the 10 groups before being tested on the 10th group. By cross-validating at this level, confidence is increased that detectors will be accurate for new participants. Oversampling (through cloning of minority class observations) was used to make the class frequency more balanced during detector development. However, performance calculations were made with reference to the original dataset.

Dr. Paquette built the detectors in RapidMiner 5.3 (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006) using six machine learning algorithms that have been successful for building similar detectors in the past (Baker et al., 2012; Pardos, Baker, San Pedro, Gowda, & Gowda, 2014): J48, JRip, NaiveBayes, Step Regression, Logistic Regression and KStar. The detector with the best performance was selected for each affective state.

Detector performance was evaluated using two metrics: Cohen's Kappa (Cohen, 1960) and A' computed as the Wilcoxon statistic (Hanley & McNeil, 1982). Cohen's Kappa assesses the degree to which the detector is better than chance at identifying the modeled construct. A Kappa of 0.5 indicates that the detector performs at chance, and a Kappa of 1 indicates that the detector performs perfectly. A' is the probability that the algorithm will correctly identify whether an observation is a positive or a negative example of the construct (e.g. is the learner frustrated or not). A' is equivalent to the area under the ROC curve in signal detection theory (Hanley & McNeil, 1982). A model with an A' of 0.0 performs at chance, and a model with an A' of 1.0 performs perfectly. A' was computed at the observation level.

---

[6] Meaning, cross validated ten times (10 folds).

When fitting models, feature selection was performed using forward selection on the Kappa metric. Performance was evaluated by repeating the feature selection process on each fold of the participant-level cross-validation in order to evaluate how well models created using this feature selection procedure perform on new and unseen test data. The final models were obtained by applying the feature selection to the complete dataset.

Affect detectors results. Performance of the interaction-based detectors was highly variable across affective states. The detector of boredom achieved the highest performance (Kappa = 0.469, A' = 0.848) while some of the other detectors achieved relatively lower performance. This was the case for the confusion detector that performed barely above chance level (Kappa = 0.056, A' = 0.552). Detectors of frustration and surprise achieved relatively low Kappa (0.105 and 0.081 respectively), but good A' (0.692 and 0.698 respectively). Performance for engaged concentration achieved a Kappa closer to the average (0.156), but below average A' (0.590).

Once the results on the performance of the affect detectors were obtained, Pearson correlations were computed between the frequencies of each affective state as obtained through the BROMP data, and the learning gains of the pre- and post-test that were administered to the participants during the study. The relationships between learning outcomes and confusion and surprise were not significant (confusion, r= -0.107, n= 100, p=0.286; surprise, r= -0.134, n=100, p=0.180). However, frustration was marginally significantly negatively correlated with learning outcomes (frustration, r= - 0.169, n=100, p= 0.092).

Given the negative correlation between the measures of frustration and learning gains, it was decided to study the effect of providing feedback to frustrated students in order to help them persevere through the learning activity, and mitigate the effect of frustration on learning gains.

This decision, then, became one of the justifications for this current dissertation study, which is discussed in the next chapter, Chapter IV: “Main Study.”

Chapter IV: MAIN STUDY

Section 1. Overview

This chapter presents the research design and results of the main study of this dissertation. Whereas the September 2013 baseline study had only one condition, this main dissertation study had five conditions: three motivational feedback intervention conditions and two control conditions. The purpose of the study discussed in this chapter was to investigate the effect of three motivational feedback conditions delivered during the serious video game vMedic on adult participants' learning while engaged in a modified TC3 Training course on hemorrhage control, tactical field care, and care under fire.

Accordingly, this chapter contains sections describing the main study's rationale, background, and significance research questions and hypotheses, method (design, participants, materials, and procedures), results, and discussion.

Section 2. Introduction

In addition to the findings from the September 2013 USMA study where the affect of frustration was determined to yield the best model for an affect detector, a survey of the literature indicated that there was a gap in the literature in terms of the directional relationship between frustration and learning gains, as well as the interactive effect of frustration, motivation, and learning. As such, this dissertation study focused on examining whether motivational feedback messages delivered during the game vMedic, upon the detection of high frustration, positively impacted learning outcomes of USMA cadets. What follows is the methodology employed to compare interventions of each of the three types of motivational feedback messages with two control conditions (one condition with non-motivational messages, and one condition with no messages delivered at all) to see which conditions yielded the greatest positive learning outcomes.

Section 3. Methodology

Participants

An a priori power analysis was conducted using the G*Power3 application for the purpose of calculating an estimate sample size to attain statistical power (Faul, Erdfelder, Lang, & Buchner, 2007). The inputs used were: (1) large estimated effect size of $f=0.25$; (2) $\alpha=0.05$; (3) desired power level = 0.80; (4) numerator df (df=degrees of freedom) = 4; and (5) number of groups = 5 (see APPENDIX F: POWER ANALYSIS WITH G*POWER3). The estimated sample size required to achieve a power level of 0.80 for Repeated Measures, within-between ANOVA was 90 participants, or 18 per condition.

After obtaining IRB approval at both Teachers College and the United States Military Acaemy (USMA) (see APPENDIX G: IRB APPROVAL), 141 volunteer participants from the Corps of Cadets were recruited using SONA Systems, a human management tool for participants at USMA. This was a population of interest because this group consists of future Army officers who will likely engage with training systems with intelligent tutors in future military training conditions, and this study fit within a larger study investigating the development of sensor-free affect detectors for integration into GIFT.

141 participants showed up over the course of the three days at USMA during which time the experiment was run. There was no deception used in this study; participants were told they were being asked to participate in a study to test the effectiveness of feedback messages. Participants were not told there would be different conditions, nor that the study was designed to elicit and detect frustration.

None of the participants needed to be dismissed or excused from participation in this study due to any fainting or queasiness as a result of the graphic images rendered in vMedic. Of the 141 participants, 17 did not complete the study due to laptop crashes not allowing enough time for cadets to restart the experiment from the beginning. A running tab was kept to insure any deficits of participants by condition due to system failures were remedied by making sure the random assignment included balancing the numbers between condition. This effectively rendered the experiment as a randomized block design so to control the nuisance factor of the crashed computers. In this, we were able to reduce the contribution to experimental error contributed by this nuisance factor. As such we were able to analyze the effect of varying levels of the primary factor (between subjects) within each block of the experiment (between conditions). In total, then, there was total of 124 complete data sets that were comprised of five blocks (five conditions) that were subsequently analyzed.

The ages of the participants (N=124) ranged from 17 to 25. Analyses were run on the complete data sets of the 14 females and 110 males that participated in this study. While the gender ratio is unbalanced, this is a reflection of overall gender difference and current ration of gender at USMA. In 2014, out of the 4,591 cadets at USMA, 83% of the cadets were males, and 17% were female. This study's sample included 11% female participants, and the rest males, similar to the current population at USMA. However, while participants were randomly assigned to conditions, on closer analysis of the gender distribution between conditions, the second condition, the social identity condition, had 8 female participants, as compared to the control value, self-efficacy, and the non-motivational control conditions that only had a single female participant, and the control condition of no messages had three female participants (see Figure 6).

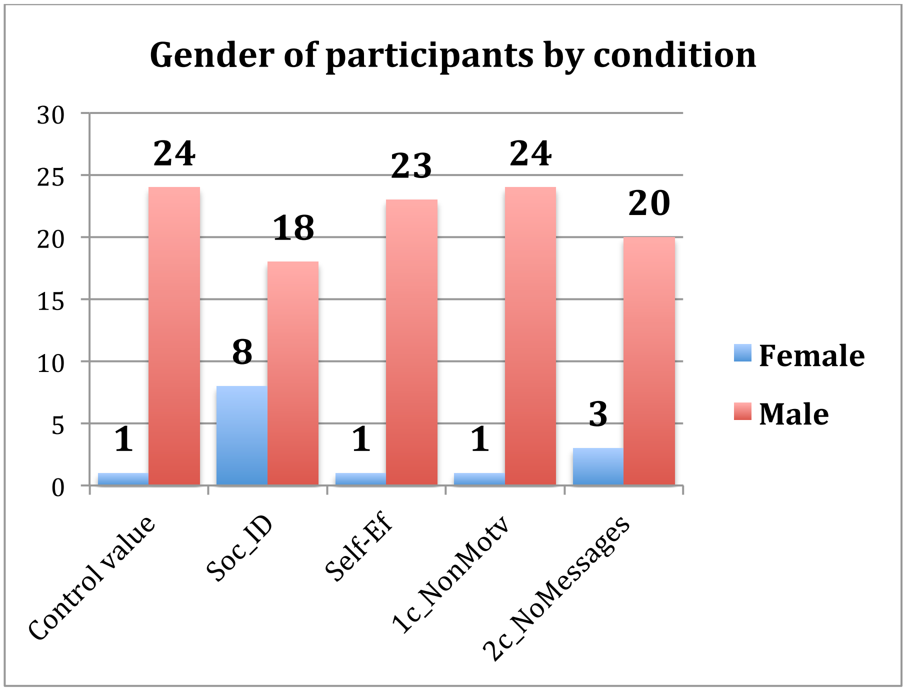


Figure 6. Gender distribution by condition.

Most participants (N=65) were freshmen in their first year of USMA while the rest of the participants were juniors in their third year (N=48) and fourth year seniors comprised the rest of the participants (N=8) with three participant not having indicated their year in school, (see Figure 7). Analyzing conditions by cadets' year in school, all conditions except the social identity condition had between 52%-62% freshman and between 29-42% juniors; the social identity condition, on the other hand had 35% freshmen and 58% juniors.

Also, 109 of the participants had not previously served in the military, while 15 participants reported that they had been active members of the military prior to joining USMA, (see Figure 8). While all cadets complete a basic training course that includes first aid training during the first summer at USMA, only a small part of this training includes materials related to TC3. When asked to rate their first aid skill, only one participant claimed to have expert knowledge, 30 had some experience and knowledge, and 93 participants self-reported that they were novices (see Figure 9).

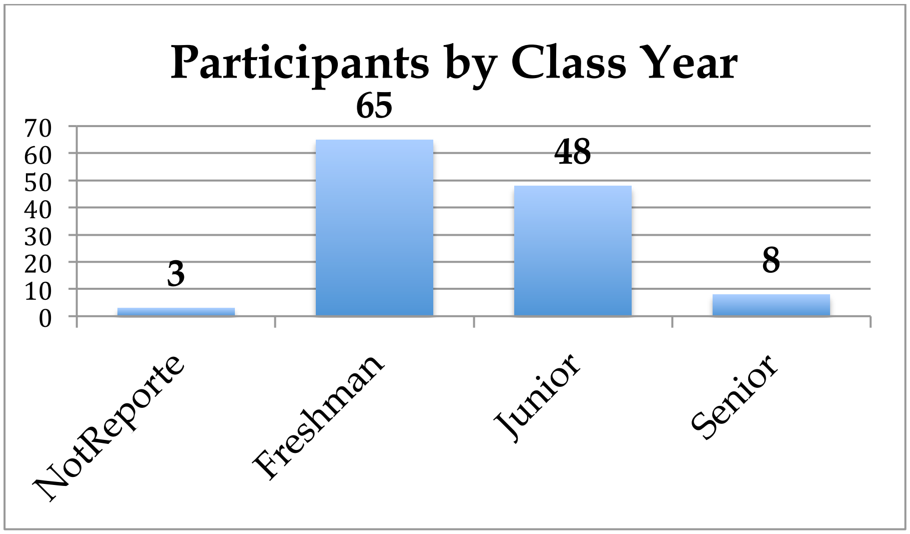


Figure 7. Breakdown of participants by year in school.

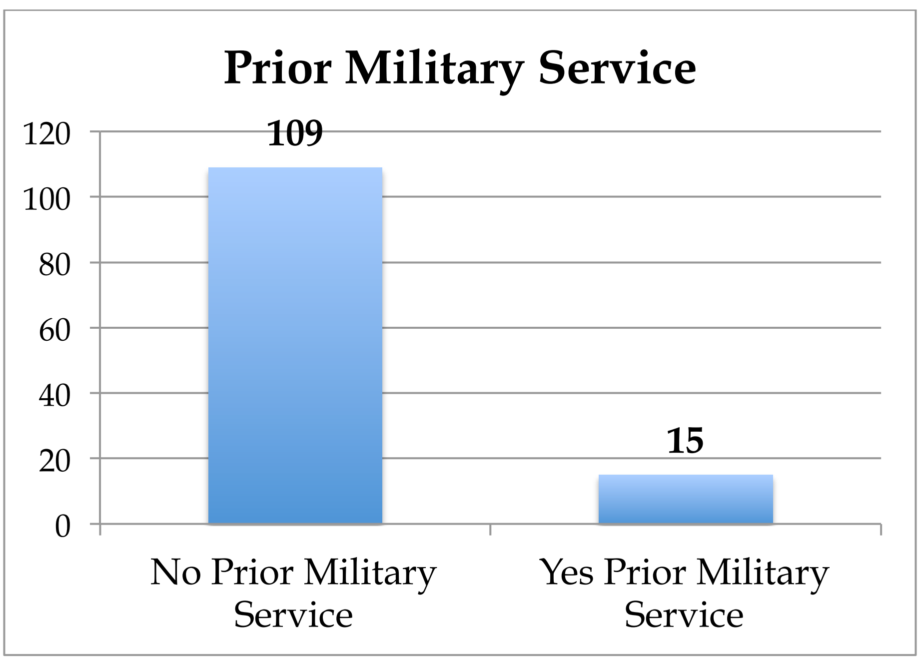


Figure 8. Prior military service.

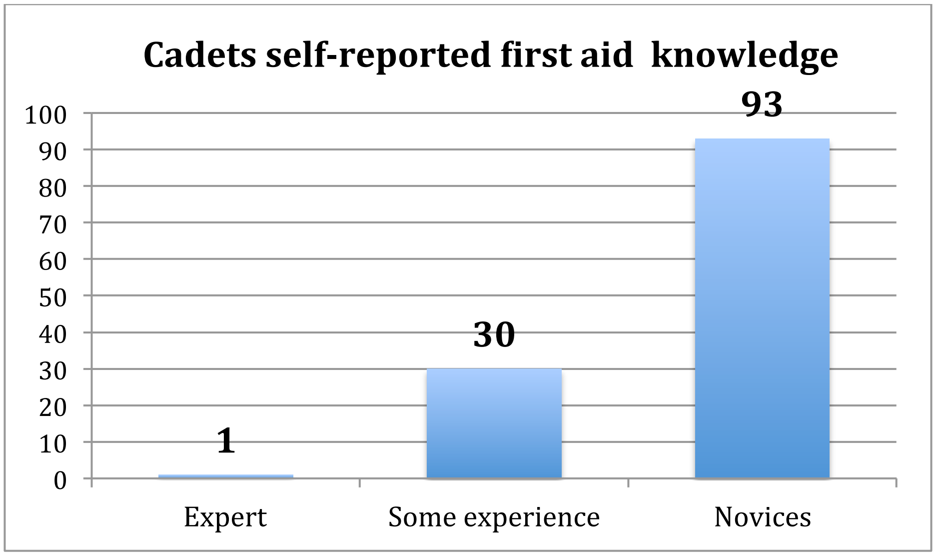


Figure 9. Cadets self-reported first-aid knowledge

The experiment was run in a lab space located in the basement of Thayer Hall at USMA, and was arranged to run ten participants at a time. Vasiliki Georgoulas-Sherry, Research Psychologist at USMA and a PhD student at Teachers College, was on hand to help with proctoring, handing out and collecting consent forms (APPENDIX H: CONSENT FORM), as well as randomly assigning participants, and assisting in rebooting individual laptops when there was a technical failure. The current author was present during the entire experiment and assisted in launching the PowerPoint, vMedic, and surveys, as well as rebooting individual laptops when necessary. In addition, she collected affect and behavior using the BROMP 2.0 field observation protocol using the HART Android application (Ocumpaugh, Baker, & Rodrigo, 2012).

Data collection was conducted over a three-day period at USMA: five sessions on the first day of the experiment (September 30, 2015); six sessions on the second day of the experiment (October 1, 2015); and seven sessions on the third day of the experiment (October 2, 2015). Each participant took part in exactly one session. All sessions on all days had participants randomly assigned to one of the five feedback conditions.

Experimental Test Bed

Domain. The domain for this experiment was Tactical Combat Casualty Care (TC3), and used the same domain content as the September 2013 study. As noted in the September 2013 study, TC3 is care rendered to a casualty in an active combat environment before hospital care can be obtained (National Association of Emergency Medical Technicians, 2016; Sotomayer, 2010), (see Figure 4, p. 44).

As the training of this critical mission is a stressful life or death situation, conducting live exercises to train combat care medics is costly and difficult to implement. As such, use of simulation training has become an optimal way to train combat medics, and the serious game vMedic was developed to assist in skill develop and allow trainees to practice treatments and execute protocol (Goldberg, 2013).

All elements of the experimental test bed were built into a course using the Generalized Intelligent Tutoring Platform (GIFT) as developed by the Army Research Laboratory (see Figure 10) and were subsequently delivered to all participants by logging into GIFT using a unique user ID (see Chapter III, section 2 for description of GIFT).

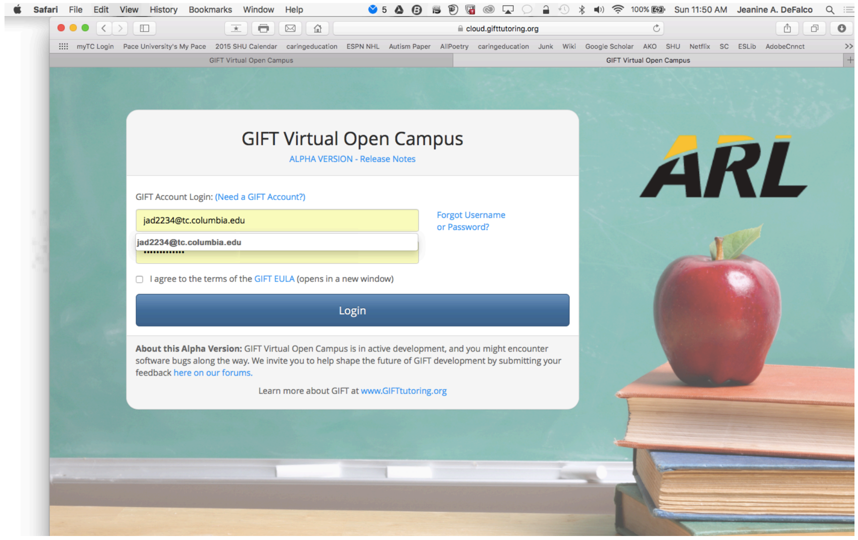


Figure 10. Screen shot of GIFT platform.

Experimental Tasks: Overview

The experimental tasks included the introduction of knowledge and procedures for combat medical care, focusing on hemorrhage and bleeding control, and were built as a course in GIFT (see Figure 11). All tests and surveys were input into the GIFT database by the current author and then administered through GIFT during the experiment.

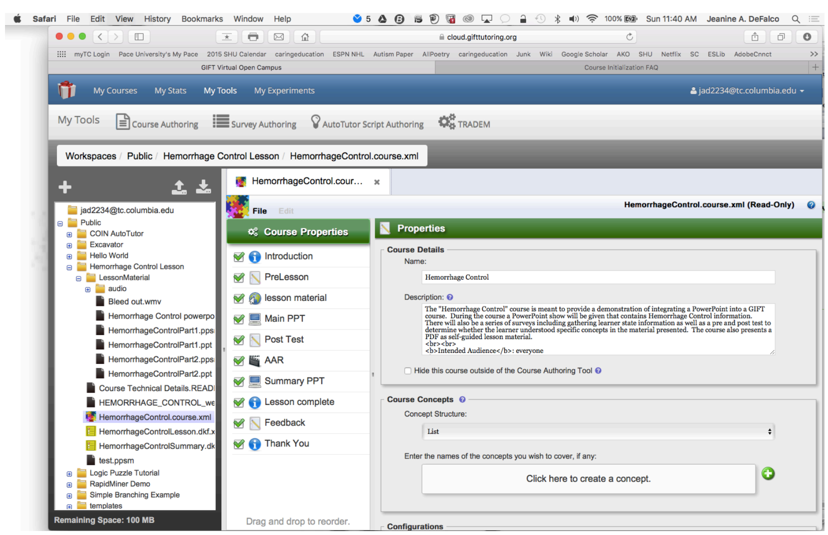


Figure 11. Screen shot of GIFT course authoring tool.

The pre-test was given to measure prior knowledge on these tasks, and this was followed by training courseware in the form of a PowerPoint. Following this PowerPoint, participants demonstrated the application of this knowledge in the simulated virtual environment: vMedic. After a tutorial that demonstrated the navigation tools of vMedic, participants engaged in five scenarios that allowed them to apply the knowledge and skills of hemorrhage and bleeding control, tactical field care, and care under fire. A post-test was delivered after vMedic to determine learning gains. The pre-test and post-test were authored in GIFT using the survey authoring tool (see Figure 12) using a bank of questions pulled from a database within GIFT.

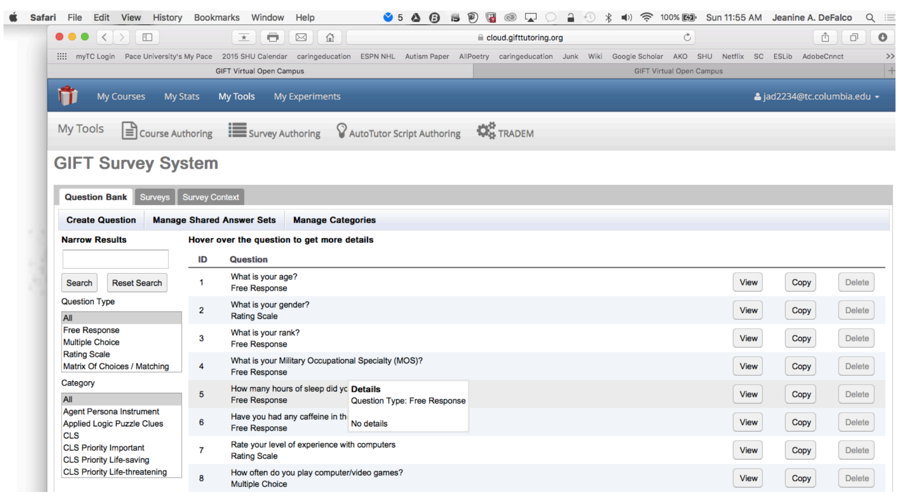


Figure 12. Screen shot of survey authoring tool in GIFT.

Experimental Design

This experiment used a random assignment, pre- and post-test, control group design, i.e. an experimental design. The design for this experiment was a one-factor between-subjects experiment with five levels. The independent variable was motivational feedback delivered upon the detection of frustration in the game vMedic. Repeated measures ANOVAs were conducted, with pre and post as the two time points to establish learning gains.

Five Conditions. The independent variable, motivational condition, had five levels:

(1) Control-value motivational messages,

(2) Social identity motivational messages,

(3) Self-efficacy motivational messages,

(4) Non motivational feedback messages,

(5) No messages.

Participants were randomly assigned to one of the five conditions. In all conditions except for the fifth condition, one message was delivered per scenario in vMedic to participants upon the system detection of high frustration.

In the first condition, the control-value motivational messages were designed around the idea that achievement emotions such as frustration can be influenced by changing the student's subjective perception of control and value through a shaping of the learning environment. As such, messages in this condition included, "Tourniquets began to gain acceptance in military medical care in the 1990s when special forces in Somalia found that the correct use of tourniquets saved lives," (DePillis, 2013) (see APPENDIX I for full bank of messages).

In the second condition, the social-identity condition messages capitalized on the notion that the cadets were members of the military, and under the social identity theory, people prefer identity-congruent to identity-incongruent actions. As such, the messages in this condition were direct quotes from Army Generals such as, "As General Patton said, 'An army is a team. It lies, sleeps, eats, and fights as a team,'" (Patton, 1944) (see APPENDIX I for full bank of messages).

In the third condition, the self-efficacy condition messages were designed to persuade the learner that they had the necessary skills to succeed, such as, “Your best outcomes will be achieved if you persist,” (see APPENDIX I for full bank of messages).

For the fourth condition, which was the first control condition, messages were non-motivational in nature and presented mere factoids to the learner. An example of the non-motivational feedback message is, “As of 10 September 2001, the unreliable, World War II–era U.S. Army tourniquet was the only widely fielded tourniquet in the U.S. military,” (Kragh et al., 2015) (see APPENDIX I for full bank of messages).

The fifth condition was the second control condition and no messages were delivered in this condition.

Message Designs. The intervention messages were audio files delivered by GIFT within the vMedic scenarios upon the detection of frustration by the embedded sensor-free affect detectors. The recording of the messages consisted of using the voice a professional, 60-year old male actor. This actor recorded all the messages using the GarageBand application on a MacBook Air laptop.

Subsequent to the recording of the messages, audio messaged were spliced with a recording with another sound file: #1 Action Movie Soundtrack (Instrumental) (Royalty Free Music Factory, #1 Ambient Soundscapes, Movie Soundtracks, & Sound Effects, Vol. 2, Royalty Free Music Factory). This design was done to seamlessly integrate the audio messages into the overall audio ambiance of the game.

Before the audio message was delivered, a very short music sound effect byte from the beginning of #1 Ambient Soundscapes played to alert the participant that a message was going to be delivered. #1 Ambient Soundscapes played in the background the entire time the spoken feedback messages was played, though the volume of the sound effect was drawn down to not impede the quality or clarity of the spoken intervention message. This overall design of splicing #1 Ambient Soundscapes with the spoken messages was duplicated across each messaged condition, differing only in length and text due to the specific message that was being delivered under each condition.

Prior to the execution of the study, all conditions were tested to ensure that audio messages would be triggered upon the detection of frustration and delivered during the scenarios in vMedic. The configuration of connecting the sensor-free affect detectors to the actual audio messages was done by Dr. Luq Paquette, at the time a post-doc in Dr. Baker's lab who assisted me in the configuration of the GIFT course used for this dissertation study.

Equipment and Materials

Apparatus. The experiment was conducted on ten separate research stations configured to collect data simultaneously; each station was used by one cadet at a time. Each station consisted of an Alienware laptop and headphones that ran the GIFT platform. All participants wore headsets throughout the experiment in order to listen to the audio components of the experiment and minimize any noise in the room.

Training Materials. The content that participants first interacted with was edited content from the TC3 training program developed by the US Army. This consisted of a multimedia PowerPoint that reviewed hemorrhage and bleeding control during Care Under Fire and Tactical Field Care. The PowerPoint used in this dissertation study was the same as used in the prior September 2013 study (see Figure 11).

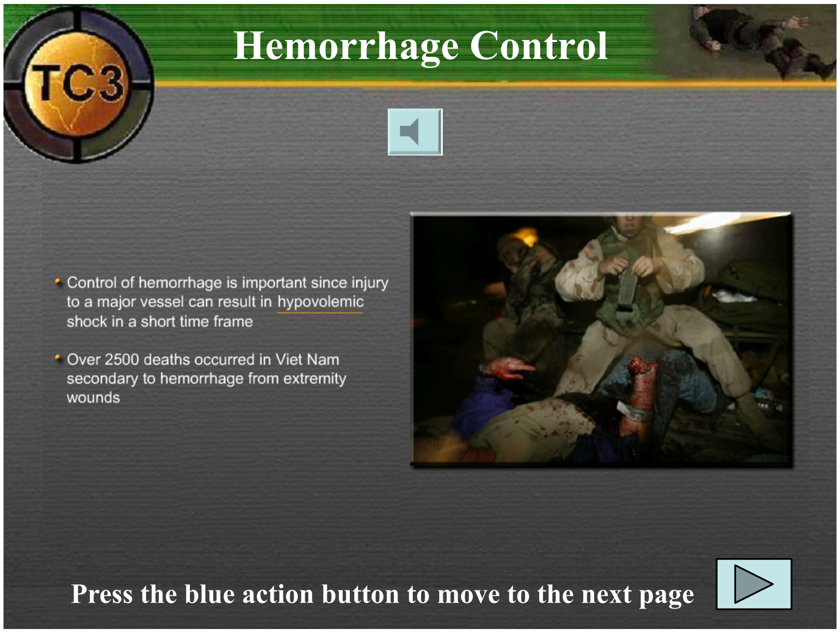


Figure 13. Example of PowerPoint used during September 2015 study

vMedic. Following the PowerPoint, participants went through a tutorial on the game controls of vMedic (see Figure 14). To review, this game aims to train the participant in the tactics, procedures and techniques required to perform emergency medical care of an Army Combat Medic and Lifesaver (CLS). Short, goal-oriented exercise are used to train tasks such as assessing casualties, performing triage, and preparing the injured for evacuation.

Figure 14. Screenshot of vMedic with navigation dial to diagnose casualty

After watching the initial tutorial in vMedic, the participants went through five scenario-based training events: (1) a leg amputation scenario; (2) the *Kobayashi Maru;* (3) repeat leg amputation scenario; (4) complex village scenario with added elements of enemy fire and loud explosions; (5) *Kobayashi Maru* again.

The sequence of these scenarios began with a simple, easy to solve and win scenario, the leg amputation that required the application of a tourniquet. This was devised so participants would be primed with the idea that this was a winnable game, if they could only devise the correct medical procedure. The second scenario, *Kobyashi Maru* — multiple hemorrhage -- was devised so that the fallen soldier that required medical attention had multiple wounds and would expire quickly – no matter what actions the participant took. In short, in the *Kobyashi Maru* scenario, it was an impossible scenario to solve. The objective for this was to elicit high frustration.

The third scenario, the leg amputation again, was to again raise the expectations of the participants with providing them an easy, winnable scenario. The objective was to imply that as participants proceeded through scenarios, they were improving in their problem solving skills and mastering the game. However, the fourth scenario, the village scenario, added the element of enemy fire and very loud explosions. This scenario, while winnable, was placed here to add to the complexity of the game, raise the stakes for the participant, and ideally increase their stress level while still allowing them to succeed. The final scenario coming right after this theoretically stressful scenario, was the *Kobyashi Maru* again in which the fallen soldier in the game had multiple wounds, expired quickly, and was a no-win situation. It was anticipated that sequencing the scenarios in this manner would increase the frequency of frustration among the participants.

Surveys. There were three surveys used in this study, all built through and administered by GIFT. A demographic questionnaire was the first assessment measure administered by GFIT to collect data on age, sex, education level, prior military experience, first aid experience, first aid expertise, computer game experience, and amount of sleep from the night before (see APPENDIX J: DEMOGRAPHICS QUESTIONNAIRE). This demographic questionnaire was based on items used in prior experiments by the Army Research Lab (Goldberg, 2013; Carroll et al., 2011; Sottilare, Brawner, & Holden, 2011) and the September 2013 study described in chapter three of this dissertation.

The second survey was administered by GIFT after the participant went through the vMedic scenarios was the Short Grit Scale (Duckworth & Quinn, 2009) (see APPENDIX A); the third survey administered was the Presence questionnaire (Witmer & Singer, 1994) (APPENDIX E) that was administered directly after the Grit survey. As discussed earlier, this survey collects data on participants' inclination to experience a sense of presence while engaged with a mediated environment (Conkey, 2011).

Dependent measures. Repeated measures ANOVAs were conducted, with pre and post tests used as the two time points of measurement to establish dependent measure of learning gains. The pre- and post-tests had 20 items each that were generated from a bank of questions located in GIFT that assessed clinical skills (e.g., asses, diagnose, treat, and evacuate) – questions that had been previously used in in the September 2013 study (see Figure 15).

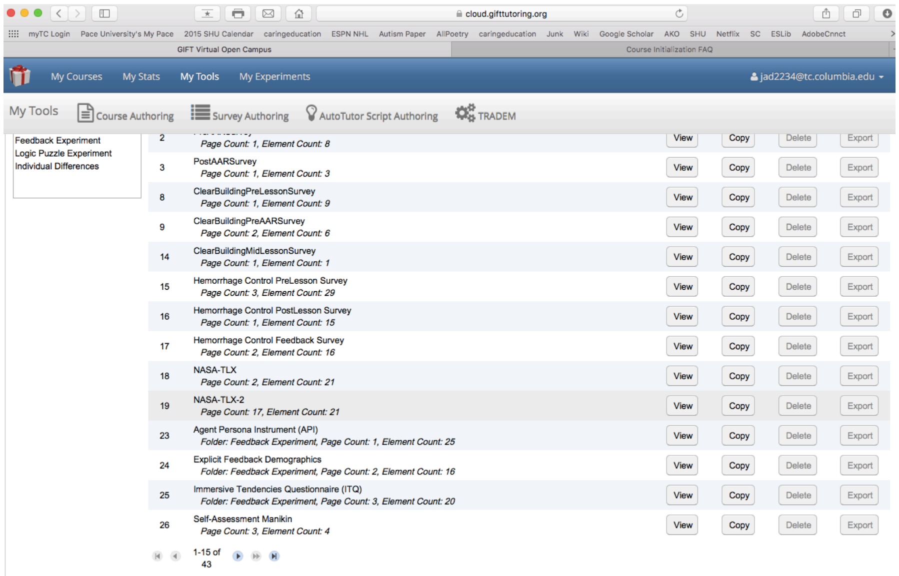


Figure 15. Screen shot of GIFT database of pre-post test questions

Moderating and mediating metrics

System detected frustration. Additional metrics used in the final data analysis included the system detected high frustration measures distilled using the detector from the log files extracted from GIFT. These log files had time stamps that were then correlated to the deployment of each vMedic scenario to ensure that feedback messages were delivered in the feedback message conditions (conditions 1, 2, 3, and 4). This system detected frustration was also calculated to obtain the mean frequencies with which the system detected a participant's level of high frustration per scenario per condition. This measure was subsequently used as a covariate in the rANOVA analyses.

Grit and presence metrics. Measures of participant's grit and presence were obtained using the Short-Item Grit Scale (Duckworth, Peterson, Matthews, & Kelly, 2007) and the Presence Survey (Witmer & Singer, 1994) respectively. The instruments were selected because both have established face and construct validity, and address potential moderating and mediating effects on pre-post test outcomes between conditions.

Research has shown a positive correlation between measures of self-efficacy and measures of grit (Slack, 2014), though these two constructs are distinct from each other. Grit is considered a stable personality trait (Duckworth, Peterson, Matthews, & Kelly, 2007) while self-efficacy is considered to be more transitory (Bandura, 1992). The choice to take measurements of grit was to examine the relationship between levels of grit and participants' response to motivational messages, particularly those messages that employed the language of self-efficacy. Specifically, the author was interested to see if there would be an interaction effect between grit and condition on learning gains, and if so the directionality of that interaction. This notion rested on the idea that a significant interaction effect would give some evidence that grit functions as a moderator variable across motivational feedback conditions (Baron & Kenny,1986).

Measurements of presence were also collected to determine whether there was an effect of presence on pre-post test scores across conditions. The Presence Survey is a multidimensional measurement devise to assess a participants' subjective experience related to a target activity in a laboratory experiment (Witmer & Singer, 1994). Specifically, presence is interpreted as the sense of being in the virtual environment depicted by computer-generated displays and the perceived ability to act in that environment (Usoh, Catena, Arman, & Slater, 2000). The objective of taking this measure was to determine whether presence would have a mediating effect across conditions on learning outcomes. Also, it was theorized that a participant's sense of presence might also indicate a participant relating more closely to the identity of their avatar, a combat soldier. As such, the author wanted to investigate whether there was a relationship between presence in the social identity condition, and whether presence had a mediating effect on learning outcomes.

Section 4. Experimental Hypotheses

Based on previous research and the results of the initial September 2013 study, the following hypotheses were generated for testing the motivational conditions. Accordingly, these hypotheses were defined around the experimental manipulations and the effect on the dependent measures (pre-post tests learning gains).

Hypothesis 1

The first hypothesis states that there will be a statistically significant difference between motivational feedback vs. non-motivational conditions when addressing frustration in a game-based learning environment. This is based on research that correlates motivational manipulations with greater cognitive processing (Locke & Braver, 2010; Maddox & Markman, 2010; Pessoa, 2009; Pessoa & Engelmann, 2010; Shohamy & Adcock, 2010), and the impact affect has on cognition (Barrett, 2006; Ciompi & Panksepp, 2004; Dalgleish & Power, 1999; Mandler, 1984; Panksepp, 2003b).

Prediction. All groups receiving motivational messages will significantly outperform the two control conditions on the dependent measure of learning gains when factoring in the mediation effect of frustration. It is expected that participants receiving motivational feedback will show greater learning gains than those participants who do not receive motivational feedback messages. This prediction is linked to a comparison of motivational feedback conditions (conditions 1, 2, & 3) vs. no motivational feedback conditions (conditions 4 & 5).

Hypothesis 2

It is hypothesized that there will be a statistically significant difference between conditions that provide intervention messages to address frustration vs. the condition where no messages are provided.in a game-based learning environment. This is based on prior research that gave evidence that providing interventions in the form of messages has been shown to positively effect the learning of domain content in ITSs (Wagster, Tan, Wu, Biswas, & Schwartz, 2007; Roll, Aleven, McLaren, & Koedinger; 2011).

Prediction. All groups receiving messages will significantly outperform the control condition on the dependent measure of learning gains. It is expected that participants receiving feedback will show greater learning gains than those participants who do not receive feedback messages. This prediction is linked to a comparison of feedback conditions (conditions 1, 2, 3, 4) vs. no message condition (condition 5).

Hypothesis 3

It is hypothesized that there will be statistically significant differences between motivational feedback conditions when addressing frustration in game-based learning environment. This is based on prior research that a "one size fits all" approach to affective feedback is unlikely to regulate emotional experiences such as frustration (D'Mello, Strain, Olney, & Graesser, 2013), and that when a participant is frustrated, their response to motivational feedback messages will be different depending upon the kind of motivational messages they receive.

Prediction. Learning gains between motivational conditions will be significantly different when controlling for the mediating effect of detected frustration, and participants receiving motivational feedback in the self-efficacy condition will show greater learning gains than those participants who do not receive self-efficacy feedback messages. This prediction is based on the notion that the messages in this condition will promote participant's efforts to persist through a frustrating learning task by supporting the participant's beliefs that they can succeed through greater immediate effort.

Further, it is expected that the self-efficacy condition will yield statistically significant better results than the social-identity issue condition particularly if the majority of the participants have not been active military personnel prior to the experiment. The messages in the social identity condition would arguably be less effective with participants who have not fully aligned their identity to being military personnel and would subsequently be less motivated by messages that address an identity with which they have not fully integrated.

Lastly, the self-efficacy condition is expected to yield statistically significant better results than the control-value theory, as the control-value messages were designed to address the more long term value in persisting through frustration, something that may not be motivational immediately in a frustrating moment.

These predictions are linked to a comparison of all conditions.

Hypothesis 4

The fourth hypothesis states that a participant's perceived presence in a game-based learning environment will influence the differences in learning outcomes between motivational feedback conditions. The rationale behind this hypothesis rests on the idea that participants who score high on the presence survey are more likely to have a sense of realism about the game environment and will identify more personally with the simulation (Witmer & Singer, 1994). Subsequently, it is possible that the social identity motivational feedback messages delivered to participants with high scores of presence might have a mediating effect in addressing in pre-post test outcomes.

Prediction. It is predicted that presence scores will mediate the learning outcomes, yielding a statistically significant difference between motivational feedback conditions. This prediction is linked to prior research that people respond more favorably to feedback messages that target their similar social-identity contexts (Schwarz, 2007, 2010; Smith & Semin, 2004, 2007), and that high levels of presence in a game-simulation indicate a greater sense of realism felt by the participant (Witmer & Singer, 1994).

Hypothesis 5

The fifth hypothesis claims that there will be a difference on learning outcomes between motivational feedback conditions based on a person's level of grit. The rationale behind this hypothesis is that grit may have an effect on a participant's receptivity to motivational feedback. This is based on prior research that demonstrates that the impact of motivational feedback differs according to groups who differ in terms of other factors such as ability and motivation (Burelson, 2006; Meyer & Turner, 2006; Rebolledo-Mendez et al., 2006).

Prediction. It is hypothesized that grit will moderate learning outcomes between conditions. This prediction is linked to an overall analysis comparing all conditions when controlling for grit, and analyzing differences in groups divided between high and low levels of grit.

Section 5. Procedure

Pre-Test, Survey, and Training

After reading and signing consent forms (see APPENDIX G: IRB APPROVAL; APPENDIX H: CONSENT FORM), participants were randomly assigned to one laptop for one of the five conditions. In this experiment, ten laptops were used allowing for two laptops to run one condition for each session for hour at a time of the experiment.

After logging in, GIFT managed some of the experimental procedures, such as the launching of the demographic questionnaire and pre-test. However, proctors had to step in and manually advance the system to launch the PowerPoint and vMedic, and occasionally the questionnaires and post-test. The interaction logs of all participants were backed up onto hard drives and then extracted at the end of the experiment for further analysis.

Once the participants were logged into GIFT, participants were prompted to complete the demographic questionnaire (APPENDIX J). Following the demographic questionnaire, participants took a pre-test of 20 questions (APPENDIX K) that consisted of questions on hemorrhage and bleeding control, tactical field care and care under fire. These questions had previously been pulled from a larger bank of questions that were already built in GIFT by the Army Research Lab to create a unique pre-test for this experiment. Upon completion of the pre-test, the participant was presented the course materials on hemorrhage control, care under fire, and tactical field care through the modified TC3 PowerPoint presentation.

Following the PowerPoint, vMedic would launch beginning with a tutorial on how to navigate the controls of this serious game. Participants engaged in a tutorial session for about three minutes to learn the interface of vMedic and general navigation through the vMedic simulated environment. After this tutorial, the training scenarios were launched, taking approximately 15-25 minutes to compete entirely. Each scenario targeted a different wound (arm, leg, chest, multiple wounds). However, for the purposes of eliciting frustration, the scenario that had been previously designed so that no matter what actions the participant took, the injured soldier would expire (coined the *Kobayashi Maru*) ran twice: once after the second scenario and again as the last scenario of the experiment.

After participating in these scenarios, the participants took the Grit Questionnaire (APPENDIX A), the Presence Questionnaire (see APPENDIX E), and a final post-test (APPENDIX L). Following this, the participant logged out of the system and were finally debriefed outside the testing room by Georgoulas-Sherry.

Post-Test and Surveys

After the participant used vMedic, GIFT administered the Short Scale Grit questionnaire (Duckworth & Quinn, 2009) (see APPENDIX A), and then the Presence questionnaire (Witmer & Singer, 1994, 2005) (see APPENDIX E). After these questionnaires, a 20 question post-test was administered by GIFT (see APPENDIX L). The questions in this post-test were pulled from the same bank of TC3Sim questions as the pre-test questions and designed in a similar manner. The post-test questions were parallel in nature to test for content mastery and avoid any recall contamination.

Participant Debrief

Once participants completed the experiment, GIFT launched a message saying that the experiment was complete. Participants were then brought outside the lab and Vasiliki Georgoulas-Sherry asked the participants if they had any questions about the experiment.

Following this brief conversation, a debrief form was handed to the participants, which was an unsigned copy of the consent form the participants had signed at the beginning of the experiment (see Table 5 for breakdown of experimental procedures for all conditions).

Table 5

Experimental Procedures for All Conditions

00:00 – 5:00 After the consent form was signed and participants logged in, the demographic survey was launched by GIFT. This took approximately 5 minutes.
5:00-15:00 After the demographic survey, GIFT presented to the participant a pre-test on hemorrhage control. There were 20 questions on the pre-test that included material on hemorrhage control, tactical field care, and care under fire. The pre-test was used to determine the effectiveness of the hemorrhage control course. This entire process in estimated to take 10 minutes to complete.
15:00 -- 20:00 After the pre-test, GIFT launched a PowerPoint presentation on hemorrhage control, tactical field care, and care under fire. It was estimated participants spent an average of 5 minutes with this content.
20:00 -- 40:00 Following the PowerPoint, GIFT launched vMedic and participants had a tutorial on the game controls. This introduction reviewed interface components and allowed participants to interact with environment elements prior to the start of the scenario-based training event. Following the tutorial, the participant went through five scenarios and applied what they knew of hemorrhage control, tactical field care, and care under fire in the serious video game vMedic. This took approximately 20 minutes.
40:00 – 45:00 After completing the five scenarios in vMedic, GIFT launched the first questionnaire, the Short-Scale Grit questionnaire (Duckworth & Quinn, 2009). After the Grit questionnaire was completed, GIFT launched a second questionnaire, the Presence questionnaire (Witmer & Singer, 1994, 2005).

45:00-55:00

Once this second questionnaire was completed, GIFT launched a post-test consisting of 20 questions. This took approximately 10 minutes.

55:00 – 60:00

When the experiment was finished, students logged out of GIFT and were debriefed outside the lab in the hallway, allowing participants an opportunity to ask questions they may have had about the experiment. This took approximately five minutes.

Chapter V: MAIN STUDY -- ANALYSIS AND RESULTS

Section 1. Overview of Statistical Analyses Performed

Statistical analyses were performed on the data using SPSS Statistics 21. An alpha value of .05 was used for all tests that included one-way ANOVA's, repeated measures ANOVA's (rANOVA), and two-way mixed design rANOVA's. Prior to conducting hypotheses testing, the data was examined to determine whether the independent variables were highly correlated, and the dependent measures met the assumptions applicable for simple ANOVA analyses and repeated measures ANOVA analyses. Exploratory analyses were performed to obtain descriptives – including histograms, Q-Q plots, and box plots -- to determine if the data appeared to be multivariate normal. Assumptions of sphericity were not applicable for rANOVA's as there were only two repeated measures employed in the experiment: one pre-test and one post-test.

The assumptions of normality were also met for both pre- and post-test score data. Once the assumptions of normality had been met, two way mixed design rANOVA's were performed to test hypotheses using SPSS GLM procedure with condition as the between-subjects fixed factors, and the pre- and post-tests as the within-subjects factors. System detected frustration, presence, and grit measures each were used as covariates to run independently as second independent, within-subject factors to test for interactions that would indicate any mediating or moderating effect of the variables (Verma, 2016). Since ANOVAs are a special case of a multiple regression model, and multiple regressions do not make any assumptions about the distributions of the explanatory variables/covariates, the system detected frustration, grit, and presence variables do not need to meet assumptions of normality to be included in the ANOVA or rANOVA analyses (Verma, 2016).

Section 2. Data Analysis Plan and Descriptives

Analyses Plan

Depending upon the hypothesis being tested, either a one-way ANOVA, a one-way repeated measures ANOVA, or a mixed design repeated measures ANOVA was run. Analyses were run to test that all assumptions were met to run the appropriate tests.

Participant Data

The log files extracted from GIFT of all 141 participants were examined to ensure all files were complete data sets. Of the 141 participants 17 participants' log files had a gap in the output where the participant either did not have a pre-test or post-test – a result of the system failures of the laptop and/or GIFT. Subsequently, these 17 participants were dropped from the data analysis, as it was impossible to calculate learning gains from incomplete sets of data.

In total, the final data analysis was run on 124 participants: 26 participants were in the first condition, 26 participants in the second condition, 24 participants in the third condition, 25 participants were in the first control condition, and 23 participants in the second control condition (see Table 6).

Table 6

N participants per condition

Participants	Condition
26	(1) Motivational: Control-value theory feedback
26	(2) Motivational: Social identity theory feedback
24	(3) Motivational: Self-efficacy theory feedback
25	(1c) Non motivational: Non motivational feedback
23	(2c) Non motivational: No messages or feedback

An rANOVA was conducted to determine if there was a statistically significant difference in learning gains between gender (N=123)[7], and whether there was a significant interaction between gender and condition. There was not a statistically significant difference in learning gains by gender $F(1,121) = .944$, $p = .333$, $d= .59$, $\eta p2 = .008$, power = .161 nor any significant interaction between gender and conditions between conditions, $F(9,113) = .358$, $p = .952$, $d=0.34$, $\eta p2 = .028$, power = .173, (see Table 7).

Table 7

rANOVA Gender differences and means of learning gains (N=123)

Gender	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
Female	13.357	1.4991	13.357	2.4054	0
Male	13.55	1.9602	14.248	1.9253	0.698

[7] One participant did not identify gender and was excluded from analysis.

Table 8

rANOVA learning gains by gender

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	3.016	1	3.016	.944	.333	.008	.944	.161
	Greenhouse-Geisser	3.016	1.000	3.016	.944	.333	.008	.944	.161
	Huynh-Feldt	3.016	1.000	3.016	.944	.333	.008	.944	.161
	Lower-bound	3.016	1.000	3.016	.944	.333	.008	.944	.161
Tests * Gender	Sphericity Assumed	3.016	1	3.016	.944	.333	.008	.944	.161
	Greenhouse-Geisser	3.016	1.000	3.016	.944	.333	.008	.944	.161
	Huynh-Feldt	3.016	1.000	3.016	.944	.333	.008	.944	.161
	Lower-bound	3.016	1.000	3.016	.944	.333	.008	.944	.161
Error(Tests)	Sphericity Assumed	386.505	121	3.194					
	Greenhouse-Geisser	386.505	121.000	3.194					
	Huynh-Feldt	386.505	121.000	3.194					
	Lower-bound	386.505	121.000	3.194					

Table 9

rANOVA interaction effect of gender and condition and means of learning gains

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	2.780	1	2.780	.830	.364	.007	.830	.147
	Greenhouse-Geisser	2.780	1.000	2.780	.830	.364	.007	.830	.147
	Huynh-Feldt	2.780	1.000	2.780	.830	.364	.007	.830	.147
	Lower-bound	2.780	1.000	2.780	.830	.364	.007	.830	.147
Tests * Gender * Condition	Sphericity Assumed	10.793	9	1.199	.358	.952	.028	3.220	.173
	Greenhouse-Geisser	10.793	9.000	1.199	.358	.952	.028	3.220	.173
	Huynh-Feldt	10.793	9.000	1.199	.358	.952	.028	3.220	.173
	Lower-bound	10.793	9.000	1.199	.358	.952	.028	3.220	.173
Error(Tests)	Sphericity Assumed	378.727	113	3.352					
	Greenhouse-Geisser	378.727	113.000	3.352					
	Huynh-Feldt	378.727	113.000	3.352					
	Lower-bound	378.727	113.000	3.352					

An rANOVA was conducted to determine if there was a statistically significant difference in learning gains between participants who had prior experience as an active duty soldier prior to enrolling at West Point and those who had no prior active duty experience, and whether there was a significant interaction between prior active duty and condition. There was not a statistically significant difference in learning gains by prior military service $F(1,122) = 1.349$, $p = .0.248$, $d= 0.21$, $\eta p2 = .011$, power = .211 (see Table 11), nor any significant interaction between prior military service by conditions, $F(9,114) = .999$, $p = .445$, d= .56, ηp2 = .073 power = .475, (see Table 12).

Table 10

rANOVA Prior active duty service differences and means of learning gains (N=123)

Active duty	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$	F(1,122)	Sig.	ηp2
Prior	14.733	1.7915	14.667	2.4976	-0.063	1.349	0.248	0.011
None	13.349	1.8677	14.092	1.9175	0.742			

Table 11

rANOVA learning gains by prior military service

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	3.017	1	3.017	.942	.334	.008	.942	.161
	Greenhouse-Geisser	3.017	1.000	3.017	.942	.334	.008	.942	.161
	Huynh-Feldt	3.017	1.000	3.017	.942	.334	.008	.942	.161
	Lower-bound	3.017	1.000	3.017	.942	.334	.008	.942	.161
Tests * Militaryservice	Sphericity Assumed	4.323	1	4.323	1.349	.248	.011	1.349	.211
	Greenhouse-Geisser	4.323	1.000	4.323	1.349	.248	.011	1.349	.211
	Huynh-Feldt	4.323	1.000	4.323	1.349	.248	.011	1.349	.211
	Lower-bound	4.323	1.000	4.323	1.349	.248	.011	1.349	.211
Error(Tests)	Sphericity Assumed	390.870	122	3.204					
	Greenhouse-Geisser	390.870	122.000	3.204					
	Huynh-Feldt	390.870	122.000	3.204					
	Lower-bound	390.870	122.000	3.204					

Table 12

rANOVA interaction effect of prior military service by condition on learning gains

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	.160	1	.160	.050	.824	.000	.050	.056
	Greenhouse-Geisser	.160	1.000	.160	.050	.824	.000	.050	.056
	Huynh-Feldt	.160	1.000	.160	.050	.824	.000	.050	.056
	Lower-bound	.160	1.000	.160	.050	.824	.000	.050	.056
Tests * Militaryservice * Condition	Sphericity Assumed	28.897	9	3.211	.999	.445	.073	8.994	.475
	Greenhouse-Geisser	28.897	9.000	3.211	.999	.445	.073	8.994	.475
	Huynh-Feldt	28.897	9.000	3.211	.999	.445	.073	8.994	.475
	Lower-bound	28.897	9.000	3.211	.999	.445	.073	8.994	.475
Error(Tests)	Sphericity Assumed	366.296	114	3.213					
	Greenhouse-Geisser	366.296	114.000	3.213					
	Huynh-Feldt	366.296	114.000	3.213					
	Lower-bound	366.296	114.000	3.213					

a. Computed using alpha =

System detected high frustration measures

System detected high frustration data was analyzed and used to determine whether frustration had a mediating effect on the pre-post test outcomes by condition.

To review, the affect detectors distilled individual actions into proxy data that was then imported into to RapidMiner 5.3, a predictive analytics software platform that was previously built into GIFT. In this process, the detectors are able to evaluate whether the participant reached a high or a low level of frustration based on a threshold probability level of 0.5. The detectors would evaluate participant's actions in vMedic in 20 second intervals upon the beginning of the first scenario of vMedic and through the subsequent four scenarios (five scenarios in all). The sensor-free detectors can be seen as a proxy for system interventions, as the interventions were triggered based on these detectors. Upon the first detection of high frustration, an intervention message would be delivered through GIFT into the vMedic scenario.. It is important to note that the system was configured so that only one message would be delivered per scenario to the participant upon the first system detection of high frustration, irrespective of the amount of times the detectors detected a participant's frustration.

The grand mean frequency of detected frustration for all conditions was 6.43 instances of detected frustration while participants engaged with vMedic. The condition with the greatest frequency of system-detected frustration was the no message condition, (the full control condition 2), with a mean frequency of 6.70 times of detected high frustration. The two conditions with the lowest apparent frequencies detected for high frustration were the control-value condition (condition 1), with a mean of 6.19 detected high frustration events, and the self-efficacy condition (condition 3), with a mean of 6.33 detected high frustration events (see Table 13, Figure 16).

Table 13

Frequency mean, minimum, and maximum scores of system detected frustration

Condition	Mean	Std. Deviation	N	Min	Max
1_CValue	6.19	1.7893	26	0	8.0
2_Socl_ID	6.65	0.9356	26	4	8.0
3_Self_Ef	6.33	1.3726	24	3	8.0
4_1c_NonMotv	6.28	1.9044	25	0	9.0
5_2c_NoMessages	6.70	1.2223	23	4	9.0

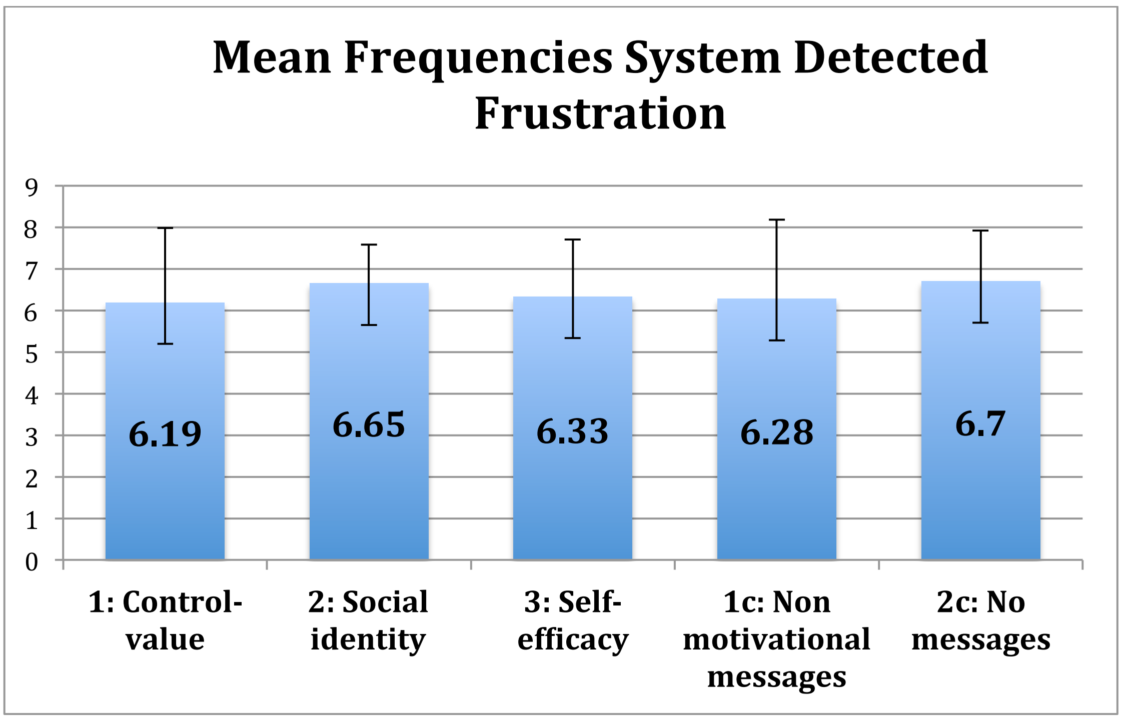


Figure 16. Mean frequency of system detected frustration by condition

In order to use system detected frustration as a covariate in hypothesis testing, there are two important considerations (1) independence of covariate and treatment effect, (2) homogeneity of regression slopes. In the first instance, the covariate should not be different across groups, meaning, in running an ANOVA using the groups as an independent variable and the covariate as an outcome, the analysis should be non-significant. A one-way ANOVA showed there was not a statistically significant difference in system detected frustration between conditions, $F(4,119) = .581$, $p = .677$, d= 0. 278, $\eta p2 = .019$, power = .188 (see Table 14).

Table 14

One-way ANOVA system high frustration by condition

Tests of Between-Subjects Effects

Dependent Variable: SystemHIGHFrustration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	5.181[a]	4	1.295	.581	.677	.019	2.325	.188
Intercept	5116.857	1	5116.857	2296.320	.000	.951	2296.320	1.000
Condition	5.181	4	1.295	.581	.677	.019	2.325	.188
Error	265.166	119	2.228					
Total	5393.000	124						
Corrected Total	270.347	123						

a. R Squared = .019 (Adjusted R Squared = -.014)

b. Computed using alpha =

To test for the second consideration, homogeneity of the slopes, Levene's Test was not significant (p=.533) (see Table 7), indicating that the assumption of homogeneity of variances had been met (see Table 15).

Table 15

Levene's test for homogeneity of the variances of system high frustration by condition

Levene's Test of Equality of Error Variances[a]

Dependent Variable: SystemHIGHFrustration

F	df1	df2	Sig.
.791	4	119	.533

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Condition

Presence Measures

The Presence survey from the Intrinsic Motivation Inventory (IMI) (Witmer & Singer, 1994) was administered to assess a participants' subjective experience in vMedic. There was an overall (n=124) mean presence score of 109.573 with a standard deviation of 16.52 (see Table 16 and Figure 17).

Table 16

Presence mean scores and standard deviation by condition

Condition	N	Mean	Std. Deviation
1_Control-value	26	106.30	17.36
2_Social identity	26	110.27	17.68
3_Self efficacy	24	107.62	17.12
1c_ Non motivational messages	25	114.28	16.70
2c_No messages	23	109.40	13.26

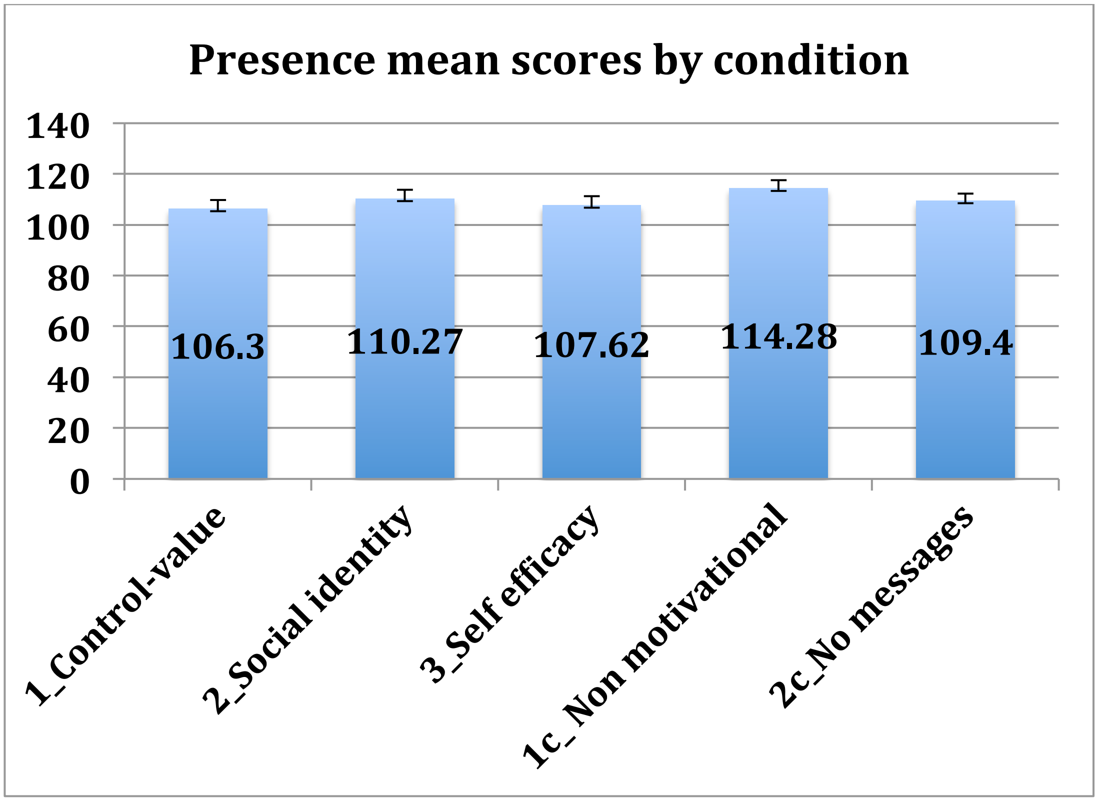


Figure 17. Presence mean scores and standard error.

A comparable benchmark for presence scores was identified in a study by Johnson and Stewart (1999). In this study, the authors reported the mean presence scores under three conditions differing in immersiveness: (1) a wide field of view (FOV) 3-D helmet-mounted display (HMD); (2) a 3-D HMD with a narrow FOV; and (3) a stationary, rear-projection, wide screen display (WSD). The mean presence scores were as follows: (1) M = 109.90; (2) M = 111.90; (3) M = 114.20. This study's mean presence score of 109.573 falls at the low range of medium level of presence, 100 <120 = medium (Johns et al., 2000). However, it is important to note that a participant's level of presence likely changed during the course of the vMedic simulation due to the transitions that occurred in-between game scenarios; blank screens in-between scenarios probably disrupted sense of immersion in the game.

In examining the scores on the post-test measures of spatial knowledge, there were no significant correlations between any of the three scores. While this was in contrast to Witmar and Singer's (1995) results that reported a significant correlation between presence scores and configuration knowledge, it was aligned with Witmer and Singer's (1995) finding that there was no correlation between presence and spatial knowledge.

In order to use system presence as a covariate in hypothesis testing, the presence measures met the following two criteria: (1) independence of covariate and treatment effect, (2) homogeneity of regression slopes.

Running an ANOVA using the groups as an independent variable and the covariate as an outcome, the analysis should be non-significant in order to use the variable as a covariate in subsequent ANOVA analyses. Accordingly, a one-way ANOVA analysis was conducted examining the significance of presence by condition. There was not a statistically significant difference in presence measures between conditions, $F(4,119) = .853$, $p = .495$, $d = 0.34$, $\eta p2 = .028$, power = .265 (see Table 17). Also, Levene's Test was not significant ($p=.747$) indicating that the assumption of homogeneity of variances had been met (see Table 18).

Table 17

One-way ANOVA, presence by condition

Tests of Between-Subjects Effects

Dependent Variable: PresenceScore

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	935.550[a]	4	233.887	.853	.495	.028	3.410	.265
Intercept	1485465.62	1	1485465.62	5414.964	.000	.978	5414.964	1.000
Condition	935.550	4	233.887	.853	.495	.028	3.410	.265
Error	32644.797	119	274.326					
Total	1522343.00	124						
Corrected Total	33580.347	123						

a. R Squared = .028 (Adjusted R Squared = -.005)

b. Computed using alpha =

Table 18

Levene's test for homogeneity of the variances of presence by condition

Levene's Test of Equality of Error Variances[a]

Dependent Variable: PresenceScore

F	df1	df2	Sig.
.485	4	119	.747

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Condition

Grit measures

The Short Grit Scale (Duckworth and Quinn, 2009) was administered to participants during the experiment. The overall (N=124) grand mean was 3.80 with a standard deviation of 0.56. The range of mean scores were 3.67 (self-efficacy condition 3) to 3.89 (control-value theory condition 1) (see Table 9 and Figure 18).

Table 19

Grit mean scores and standard deviations by condition

Condition	N	Mean	Std. Deviation
1. Control-value	26	3.89	.54
2. Social identity	26	3.84	.49
3. Self efficacy	24	3.67	.70
1c. Non motivational	25	3.71	.62
2c. No messages	23	3.88	.42

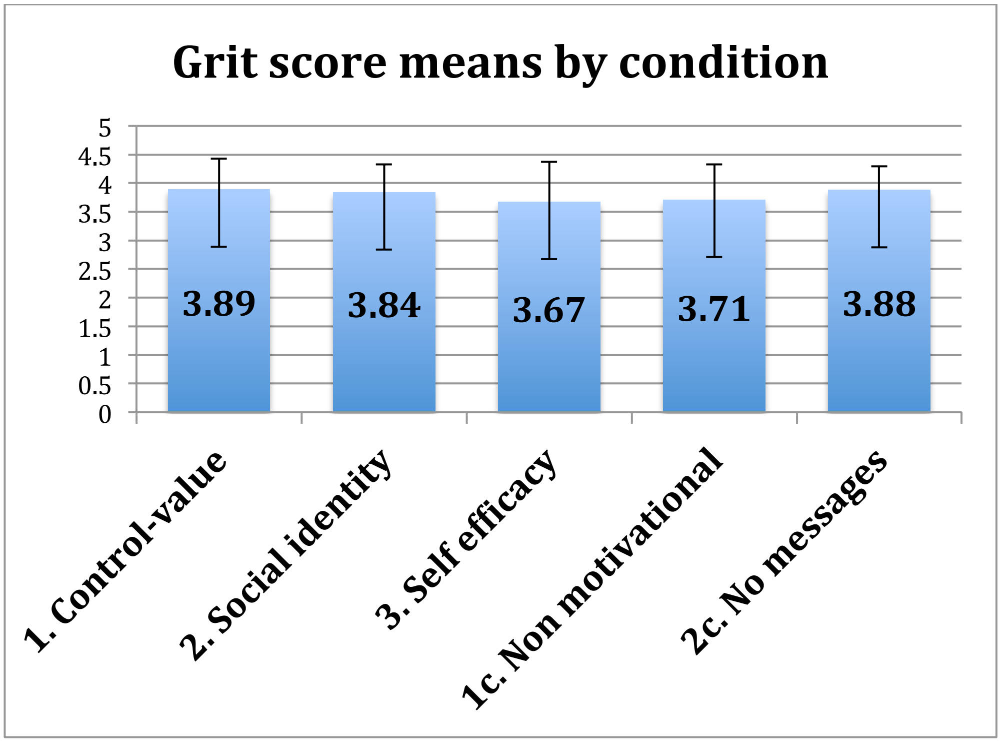


Figure 18. Grit mean scores by condition.

To examine if there was a statistical differences in grit scores between conditions, a one-way ANOVA analysis was conducted. There was not a statistically significant difference in grit between conditions, $F(4,119) = .792$, $p = .533$, $d = 0.33$, $\eta p2 = .026$, power = .248 (see Table 20). Also, Levene's Test was not significant ($p=.273$) indicating that the assumption of homogeneity of variances had been met (see Table 21).

Table 20

One-way ANOVA, grit by condition

Tests of Between-Subjects Effects

Dependent Variable: Grit_Score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	1.009[a]	4	.252	.792	.533	.026	3.168	.248
Intercept	1787.975	1	1787.975	5613.907	.000	.979	5613.907	1.000
Condition	1.009	4	.252	.792	.533	.026	3.168	.248
Error	37.900	119	.318					
Total	1831.750	124						
Corrected Total	38.909	123						

a. R Squared = .026 (Adjusted R Squared = -.007)

b. Computed using alpha =

Table 21

Levene's Test for homogeneity of the variances of grit by condition

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Grit_Score

F	df1	df2	Sig.
2.840	4	119	.273

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

A comparable benchmark for grit scores was identified in a study by Kelly, Matthews, and Bartone (2014) that investigated grit as a predictor of performance among West Point cadets for the class of 2010. In this study, the grit scores for cadets that continued at USMA after cadet basic training was $M = 3.77$. As such, the grand mean grit score of 3.80 (N=124) obtained for this dissertation study is very similar to the Kelly, Matthews, and Bartone (2014) benchmark. The 0.03 mean difference from the benchmark 2010 freshman class to the mixed-class participant sample of this September 2016 study is only a 0.006% deviation from the grand mean grit scored obtained in this study (*M*=3.80).

Pre-Post Test Data – Dependent Measures

Pre-test. Pre test data was examined to determine if assumptions of normality were met (see Figures 19, 20, 21).

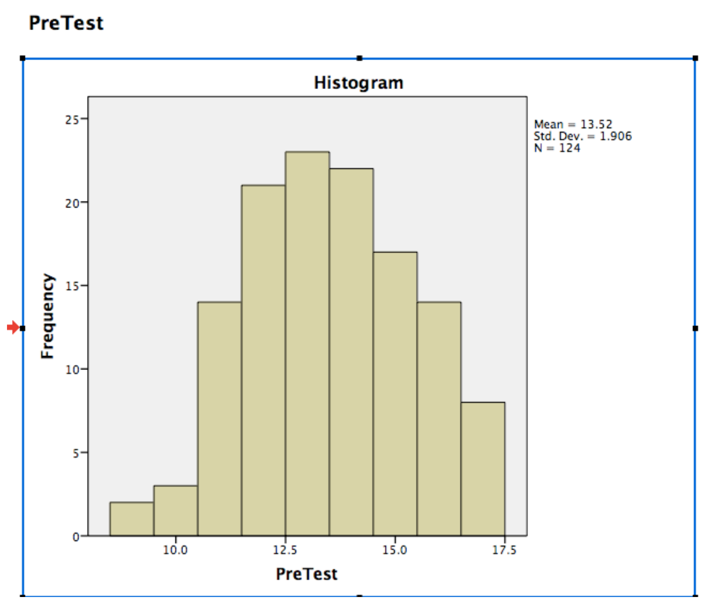


Figure 19. Pre-test histogram

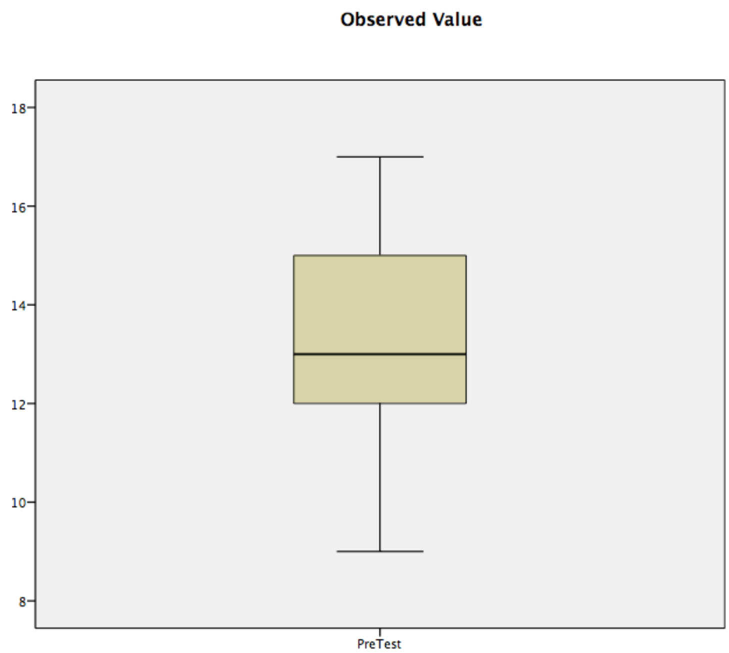


Figure 20. Pre-test boxplot

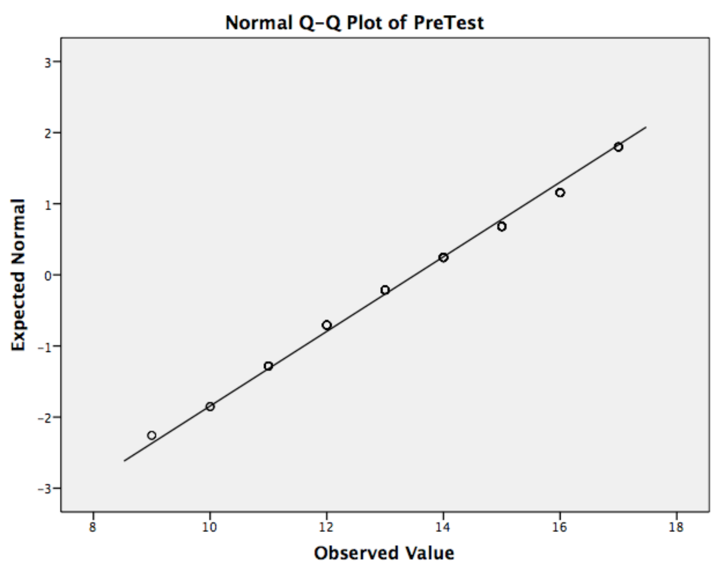


Figure 21. Q-Q plot for pre-test

The data appeared approximately normal. Further, the pre-test measures satisfied the criteria for a normal distribution: skewness of the distribution = -.004, between -1.0 and +1.0; kurtosis of the distribution = -.644, between -1.0 and +1.0 (see Table 22).

Table 22

Pre-Post test data descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
PreTest	124	9.0	17.0	13.516	1.9063	-.004	.217	-.644	.431
PostTest	124	8.0	18.0	14.161	1.9934	-.452	.217	.246	.431
Valid N (listwise)	124								

Post-test. The post test data satisfied the criteria for a normal distribution: skewness of the distribution = -.452, was between -1.0 and +1.0; kurtosis of the distribution = -.246 was between -1.0 and +1.0 (see Table 11 above, Figures 22, 23, and 24). There was one outlier in the post test (see Figure 23): participant #34 had a post test of 8 which was less than the pre-test of 11. However, this data was not omitted from subsequent analyses as it appears to be a genuine score.

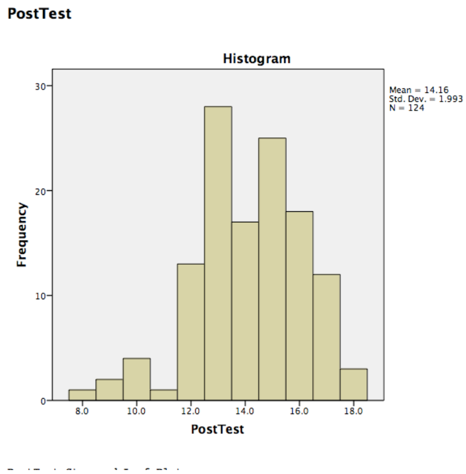


Figure 22. Post-test histogram

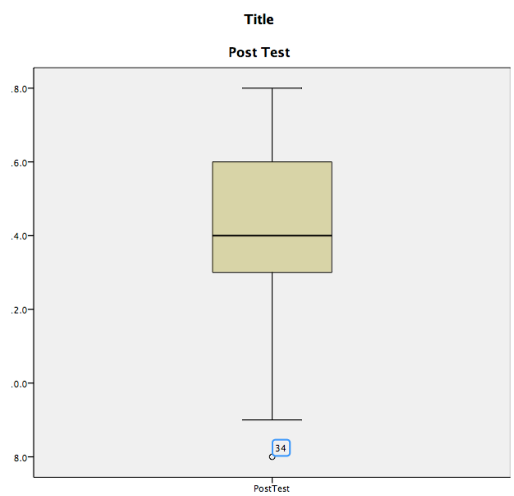


Figure 23. Post-test boxplot

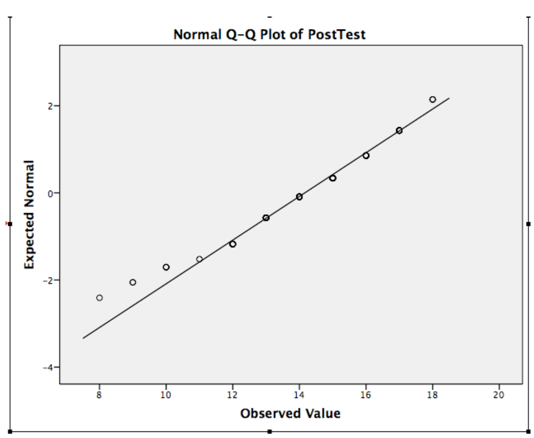


Figure 24. Q-Q plot for post-test

Homogeneity of variance. The assumption for homogeneity of the variance of the dependent variable was equal across the levels of between-subjects factor for each level, as confirmed by examining the Levene's test in SPSS, Pre-test, p=.878; Post-test, p=.692 (see Table 23).

Table 23

Levene's test of homogeneity of variance on pre-post test data

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
PreTest	.300	4	119	.878
PostTest	.561	4	119	.692

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Condition
Within Subjects Design: Tests

Sphericity. The sphericity assumption is not violated for rANOVA analyses, because there are only two repeated measures taken (pre and post test) (see Table 24).

Table 24

Mauchly's test of sphericity on pre-post test data

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

					Epsilon[b]		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Tests	1.000	.000	0	.	1.000	1.000	1.000

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + Condition
Within Subjects Design: Tests

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Homogeneity of variances-covariance matrices. Examining the variance-covariance matrices of the dependent variable was equal across the cells of all levels of the between-subjects factor (condition) as indicated by a non-significant Box's M test in SPSS, p=.977 (see Table 14).

Table 25

Box's M test for homogeneity of variance-covariance matrices

Box's Test of Equality of Covariance Matrices[a]

Box's M	4.493
F	.361
df1	12
df2	101802.553
Sig.	.977

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Condition
Within Subjects Design: Tests

Correlation Analysis. Lastly, a bivariate correlation analysis was run on all the following measures: system high frustration, presence, and grit. There were no significant correlations between any the measures, which indicated that frustration, presence, and grit could be used as independent covariates in analyses.

Table 26

Bivariate correlation analysis on frustration, grit, and presence

Correlations

		SystemHIGHFrustration	Grit_Score	PresenceScore
SystemHIGHFrustration	Pearson Correlation	1	-.019	.007
	Sig. (2-tailed)		.838	.943
	N	124	124	124
Grit_Score	Pearson Correlation	-.019	1	.035
	Sig. (2-tailed)	.838		.701
	N	124	124	124
PresenceScore	Pearson Correlation	.007	.035	1
	Sig. (2-tailed)	.943	.701	
	N	124	124	124

Positive learning gains. There were positive learning gains[8] in all conditions (see Table 27 and Figure 25).

Table 27

Mean and standard deviations for pre- and post-tests by condition

Condition	Pre-test mean	Post-test mean	Std. Dev. Pre-test	Std. Dev. Post-test
1_Control-value	13.115	13.731	1.97	2.21
2_Social identity	13.885	14.231	1.90	2.29
3_Self efficacy	13.167	14.333	2.06	1.93
1c_Non motivational messages	13.56	14.24	1.93	1.71
2c_No messages	13.87	14.304	1.63	1.82

[8] The pre-test and post-test score ranged from 0 (minimum) to 20 (maximum).

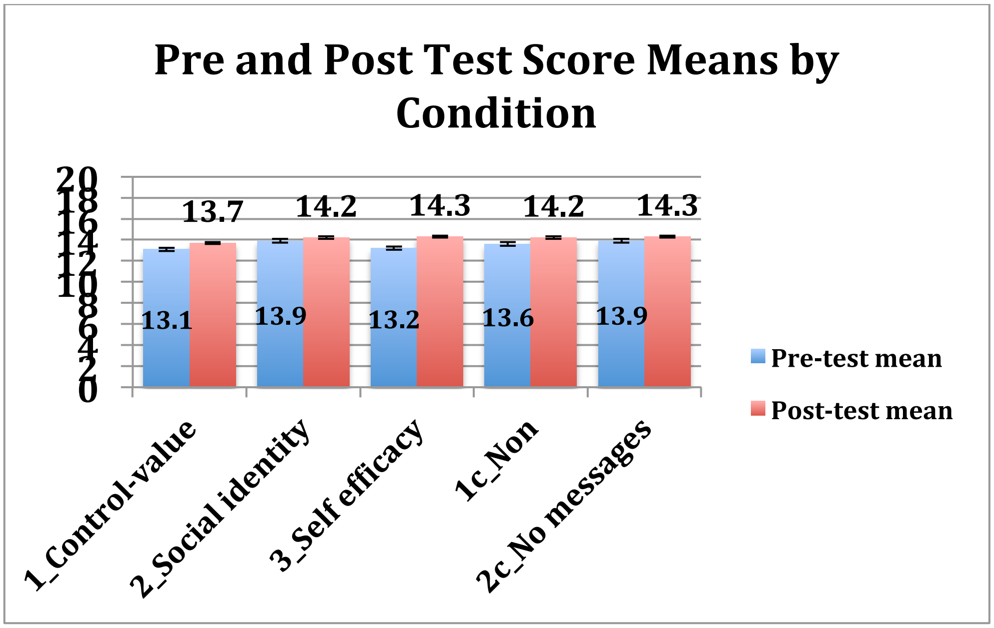


Figure 25. Pre and post-test mean scores by condition

Section 3. Assumptions Summary

To proceed with testing the hypotheses using one way repeated measures ANOVA and a mixed-design repeated measures ANOVA (rANOVA), the following assumptions were met: data type, normality, independence of observations, homogeneity of the variance, sample size, sphericity, and homogeneity of variance-covariance matrices (Verma, 2016).

I. **Data type:** For this experiment, the independent variable was categorical (motivational conditions) and had five levels: (1) control-value, (2) social identity, (3) self-efficacy, (4) non-motivational messages, (5) no messages. These five levels exceeded the required three levels needed to run an rANOVA. Also, the dependent variable (tests) was measured on an interval scale (Verma, 2016). Two-way mixed designs are used to analyze the effect of two independent factors on the dependent variable where one of the factors is a between-subjects and the other is a within-subjects (Verma, 2016). For the two-way mixed rANOVA, the between-subjects variable was condition and the other independent factors were within-subjects factors (system detected frustration, presence, and grit) and treated as covariates in the model. In this way, the rANOVA could test those hypotheses to investigate whether there were any significant interactions between tests, conditions, and the independent factors of system detected frustration, presence, and grit.

II. **Normality**: The distribution of the dependent variable (tests) in the groups were approximately normally distributed (as noted in the Section 3, Pre-Post-Test Data). There was only one outlier in the post-test data (participant 34), but upon inspection this score seemed genuine and did not warrant exclusion from subsequent analyses (Verma, 2016).

III. **Independence of observations**: This assumption was met as the observations between groups were independent and were made up with different participants in each group. Also, within each group the observations were independent (Verma, 2016).

IV. **Homogeneity of variance**: Using Levene's test, (Verma, 2016) the assumption for homogeneity of the variance of the dependent variable was met: Pre-test, p=.878; Post-test, p=.692.

V. **Sample size**: For each cell, the subjects per cell were greater than the 20 subjects per cell recommended (Verma, 2016).

VI. **Sphericity**: To run an rANOVA, the variances of the differences between all combinations of related groups must be equal. In this experiment, the assumption of sphericity was met as there were only two levels of the repeated measures variable (pre and post tests).

VII. **Homogeneity of variance-covariance matrices**: The variance-covariance matrices of the dependent variable was equal across the cells of all levels of the between-subjects factor (condition) as indicated by a non-significant Box's M test in SPSS, p=.977.

Section 4. Hypotheses Testing

Analyses were run to test the hypotheses identified in Chapter IV, Section 4.

Hypothesis 1

The first hypothesis states that there will be a statistically significant difference between motivational feedback vs. non-motivational conditions when controlling for frustration in a game-based learning environment. A two-way mixed design rANOVA design (Verma, 2016) was used to analyze the effect of two independent factors on the dependent variable (tests), where one of the factors was the between subjects (condition) and the other was a within-subjects factor (system detected frustration).

Test. Two-way mixed design rANOVA analysis was run to determine if there is a main effect when controlling for frustration and three-way interaction with frustration[9]. There was a significant main effect: (rANOVA): $F(4, 114) = 3.68$, $p = .007$, $\eta p2 = .114$ (see Table 28).

[9] Frustration here is functioning as a mediator to the main effects of condition and test scores, as mediators address when a variable alters the direction or strength of the relation between a predictor and an outcome -- in this case, frustration alters the relationship between differences in conditions and test scores (Frazier, Tix, & Barron, 2004).

Table 28

Test for main effect controlling for frustration and three-way interaction with frustration

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Tests	Pillai's Trace	.006	.699[b]	1.000	114.000	.405	.006
	Wilks' Lambda	.994	.699[b]	1.000	114.000	.405	.006
	Hotelling's Trace	.006	.699[b]	1.000	114.000	.405	.006
	Roy's Largest Root	.006	.699[b]	1.000	114.000	.405	.006
Tests * Condition	Pillai's Trace	.114	3.680[b]	4.000	114.000	.007	.114
	Wilks' Lambda	.886	3.680[b]	4.000	114.000	.007	.114
	Hotelling's Trace	.129	3.680[b]	4.000	114.000	.007	.114
	Roy's Largest Root	.129	3.680[b]	4.000	114.000	.007	.114
Tests * SystemHIGHFrustration	Pillai's Trace	.016	1.883[b]	1.000	114.000	.173	.016
	Wilks' Lambda	.984	1.883[b]	1.000	114.000	.173	.016
	Hotelling's Trace	.017	1.883[b]	1.000	114.000	.173	.016
	Roy's Largest Root	.017	1.883[b]	1.000	114.000	.173	.016
Tests * Condition * SystemHIGHFrustration	Pillai's Trace	.108	3.443[b]	4.000	114.000	.011	.108
	Wilks' Lambda	.892	3.443[b]	4.000	114.000	.011	.108
	Hotelling's Trace	.121	3.443[b]	4.000	114.000	.011	.108
	Roy's Largest Root	.121	3.443[b]	4.000	114.000	.011	.108

a. Design: Intercept + Condition + SystemHIGHFrustration + Condition * SystemHIGHFrustration
Within Subjects Design: Tests

b. Exact statistic

Then, a two-way mixed design rANOVA analysis was run comparing the motivational conditions (conditions 1, 2, & 3) to the control conditions (conditions 4, & 5) (see Table 29). This analysis indicated that motivational conditions had higher positive learning outcomes than the control conditions: (rANOVA): $F(1, 120) = 5.627$, $p = .019$. $\eta p2 = .045$, power=.653. Also, there was a statistically significant interaction between conditions, frustration, and learning outcomes (rANOVA): $F(1, 120) = 5.578$, $p = .020$. $\eta p2 = .044$, power=.649 (see Table 30).

Table 29

Means chart for Motivational vs. Non-motivational conditions rANOVA test

Groups	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
Motivational	13.40	1.9804	14.09	2.14	0.70
Non-motivational	13.71	1.786	14.27	1.75	0.56

Table 30

Two-way mixed design rANOVA comparing motivational feedback vs. no motivational feedback

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	.751	1	.751	.240	.625	.002	.240	.077
	Greenhouse-Geisser	.751	1.000	.751	.240	.625	.002	.240	.077
	Huynh-Feldt	.751	1.000	.751	.240	.625	.002	.240	.077
	Lower-bound	.751	1.000	.751	.240	.625	.002	.240	.077
Tests * FeedbackVsControl	Sphericity Assumed	17.581	1	17.581	5.627	.019	.045	5.627	.653
	Greenhouse-Geisser	17.581	1.000	17.581	5.627	.019	.045	5.627	.653
	Huynh-Feldt	17.581	1.000	17.581	5.627	.019	.045	5.627	.653
	Lower-bound	17.581	1.000	17.581	5.627	.019	.045	5.627	.653
Tests * SystemHIGHFrustration	Sphericity Assumed	3.980	1	3.980	1.274	.261	.011	1.274	.201
	Greenhouse-Geisser	3.980	1.000	3.980	1.274	.261	.011	1.274	.201
	Huynh-Feldt	3.980	1.000	3.980	1.274	.261	.011	1.274	.201
	Lower-bound	3.980	1.000	3.980	1.274	.261	.011	1.274	.201
Tests * FeedbackVsControl * SystemHIGHFrustration	Sphericity Assumed	17.429	1	17.429	5.578	.020	.044	5.578	.649
	Greenhouse-Geisser	17.429	1.000	17.429	5.578	.020	.044	5.578	.649
	Huynh-Feldt	17.429	1.000	17.429	5.578	.020	.044	5.578	.649
	Lower-bound	17.429	1.000	17.429	5.578	.020	.044	5.578	.649
Error(Tests)	Sphericity Assumed	374.943	120	3.125					
	Greenhouse-Geisser	374.943	120.000	3.125					
	Huynh-Feldt	374.943	120.000	3.125					
	Lower-bound	374.943	120.000	3.125					

Conducting a post-hoc, simple main analysis for a two-way rANOVA with a significant interaction requires running separate rANOVA's for factors under investigation (Keselman, 1998; Verma, 2016; Weinberg & Abramowitz, 2002). Therefore, subsequent independent pairwise analyses rANOVA's were run comparing each motivational condition separately to each control condition as follows (see pre-post test means in Table 31, and summary of findings in Table 32).

This analysis indicated that the self-efficacy condition (N=24) had higher positive learning outcomes than the non-motivational feedback control group (*N* = 25), (rANOVA): F(1, 45) = 10.483, p = .002, ηp2 = .189, power = .886, (see Table 33). Using the Benjamini-Hochberg[10] adjusted alpha, these results are still significant: p = .002 < B-H α = .008.

10 The Benjamini-Hochberg procedure is an approach to controlling the false discovery rate in multiple comparisons (Benjamini & Hochberg, 1995; Thissen, Steinberg, & Kuang, 2002) which is thought to balance between Type I and Type II error better than more traditional family-wise error rate tests such as Bonferroni.

This analysis indicated that the self-efficacy condition (*N*=24) had higher positive learning outcomes than the no message control group (*N* = 23), (rANOVA): $F(1, 43) = 7.355$, $p = .007$, $\eta p2 = .159$, power = .796 (see Table 34). Again, using the Benjamini-Hochberg adjusted alpha, these results are still significant: $p = .007 <$ B-H $\alpha = .016$.

No other comparisons were significant when using the Benjamini-Hochberg procedure. There was no statistically significant difference between the control-value condition (*N* = 26) and the non-motivational feedback control group (*N* = 25), (rANOVA): $F(1, 48) = .004$, $p = .948$. $\eta p2 = .000$, power = .050; $p = .948 >$ B-H $\alpha = .033$, (see Table 35).

There was no statistically significant difference between the control-value condition and the no messages control group, (rANOVA): $F(1, 45) = 2.290$, $p = .137$. $\eta p2 = .048$, power = .326; $p = .137 >$ B-H $\alpha = 0.025$, (see Table 36).

There was no statistically significant difference between the social identity condition (*N* = 26) and the non-motivational feedback control group (*N* =25), (rANOVA): $F(1, 47) = .877$, $p = .354$, $\eta p2 = .018$, power = .152; $p = .354 >$ B-H $\alpha = 0.041$, (see Table 37).

There was no statistically significant difference between the social identity condition and the no messages control group, (rANOVA): $F(1, 45) = .352$, $p = .556$. $\eta p2 = .008$, power = .089; $p = .556 >$ B-H $\alpha = 0.500$, (see Table 38).

Table 31

Means chart for rANOVA test by condition w/interaction effect of frustration

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
1_Control Value	13.115	1.966	13.731	2.219	0.61
2_Social Identity	13.885	1.904	14.231	2.286	0.35
3_Self-Effficacy	13.167	2.057	14.333	1.926	1.166
4_1c_NonMotivational	13.56	1.938	14.240	1.715	0.68
5_2c_NoMessages	13.87	1.632	14.304	1.820	0.434

Table 32

Summary of pairwise analyses (rANOVA's) between intervention conditions vs. control groups

Intervention	*Control group*	*df*	*F*	*Sig*	*Adjusted* α	ηp2	*Power*
Control-value	Non motivational messages	1	.004	.948	0.033	.000	.050
Social identity	Non motivational messages	1	.877	.354	0.041	.018	.151
**Self efficacy*	*Non motivational messages*	1	9.945	.002	0.008	.181	.870
Control-value	No messages	1	2.290	.137	0.025	.048	.316
Social identity	No messages	1	.352	.556	0.500	.008	.089
**Self-efficacy*	*No messages*	1	7.355	.007	0.016	.146	.755

* statistically significant after Benjamini-Hochberg correction

Table 33

Comparing self-efficacy (condition 3) vs. non-motivational feedback control group (control group 1)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	4.623	1	4.623	1.753	.192	.037	1.753	.254
	Greenhouse-Geisser	4.623	1.000	4.623	1.753	.192	.037	1.753	.254
	Huynh-Feldt	4.623	1.000	4.623	1.753	.192	.037	1.753	.254
	Lower-bound	4.623	1.000	4.623	1.753	.192	.037	1.753	.254
Tests * SelfEfficacy_vs_NonMotivationalMessages	Sphericity Assumed	27.652	1	27.652	10.483	.002	.189	10.483	.886
	Greenhouse-Geisser	27.652	1.000	27.652	10.483	.002	.189	10.483	.886
	Huynh-Feldt	27.652	1.000	27.652	10.483	.002	.189	10.483	.886
	Lower-bound	27.652	1.000	27.652	10.483	.002	.189	10.483	.886
Tests * SystemHIGHFrustration	Sphericity Assumed	1.168	1	1.168	.443	.509	.010	.443	.100
	Greenhouse-Geisser	1.168	1.000	1.168	.443	.509	.010	.443	.100
	Huynh-Feldt	1.168	1.000	1.168	.443	.509	.010	.443	.100
	Lower-bound	1.168	1.000	1.168	.443	.509	.010	.443	.100
Tests * SelfEfficacy_vs_NonMotivationalMessages * SystemHIGHFrustration	Sphericity Assumed	26.231	1	26.231	9.945	.003	.181	9.945	.870
	Greenhouse-Geisser	26.231	1.000	26.231	9.945	.003	.181	9.945	.870
	Huynh-Feldt	26.231	1.000	26.231	9.945	.003	.181	9.945	.870
	Lower-bound	26.231	1.000	26.231	9.945	.003	.181	9.945	.870
Error(Tests)	Sphericity Assumed	118.699	45	2.638					
	Greenhouse-Geisser	118.699	45.000	2.638					
	Huynh-Feldt	118.699	45.000	2.638					
	Lower-bound	118.699	45.000	2.638					

a. Computed using alpha =

Table 34

Comparing self-efficacy (condition 3) vs. no feedback control group (control group 2)

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
Tests	Pillai's Trace	.013	.587[b]	1.000	43.000	.448	.013	.587	.116
	Wilks' Lambda	.987	.587[b]	1.000	43.000	.448	.013	.587	.116
	Hotelling's Trace	.014	.587[b]	1.000	43.000	.448	.013	.587	.116
	Roy's Largest Root	.014	.587[b]	1.000	43.000	.448	.013	.587	.116
Tests * SelfEfficacy_vs_NoMessages	Pillai's Trace	.159	8.124[b]	1.000	43.000	.007	.159	8.124	.796
	Wilks' Lambda	.841	8.124[b]	1.000	43.000	.007	.159	8.124	.796
	Hotelling's Trace	.189	8.124[b]	1.000	43.000	.007	.159	8.124	.796
	Roy's Largest Root	.189	8.124[b]	1.000	43.000	.007	.159	8.124	.796
Tests * SystemHIGHFrustration	Pillai's Trace	.004	.153[b]	1.000	43.000	.698	.004	.153	.067
	Wilks' Lambda	.996	.153[b]	1.000	43.000	.698	.004	.153	.067
	Hotelling's Trace	.004	.153[b]	1.000	43.000	.698	.004	.153	.067
	Roy's Largest Root	.004	.153[b]	1.000	43.000	.698	.004	.153	.067
Tests * SelfEfficacy_vs_NoMessages * SystemHIGHFrustration	Pillai's Trace	.146	7.355[b]	1.000	43.000	.010	.146	7.355	.755
	Wilks' Lambda	.854	7.355[b]	1.000	43.000	.010	.146	7.355	.755
	Hotelling's Trace	.171	7.355[b]	1.000	43.000	.010	.146	7.355	.755
	Roy's Largest Root	.171	7.355[b]	1.000	43.000	.010	.146	7.355	.755

a. Design: Intercept + SelfEfficacy_vs_NoMessages + SystemHIGHFrustration + SelfEfficacy_vs_NoMessages * SystemHIGHFrustration
Within Subjects Design: Tests

b. Exact statistic

c. Computed using alpha =

Table 35

Comparing control value (condition 1) vs. non-motivational feedback (control group 1)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	.780	1	.780	.241	.626	.005	.241	.077
	Greenhouse-Geisser	.780	1.000	.780	.241	.626	.005	.241	.077
	Huynh-Feldt	.780	1.000	.780	.241	.626	.005	.241	.077
	Lower-bound	.780	1.000	.780	.241	.626	.005	.241	.077
Tests * ControlValue_vs_NonMotivationalMessages	Sphericity Assumed	.014	1	.014	.004	.948	.000	.004	.050
	Greenhouse-Geisser	.014	1.000	.014	.004	.948	.000	.004	.050
	Huynh-Feldt	.014	1.000	.014	.004	.948	.000	.004	.050
	Lower-bound	.014	1.000	.014	.004	.948	.000	.004	.050
Tests * SystemHIGHFrustration	Sphericity Assumed	3.493	1	3.493	1.079	.304	.022	1.079	.175
	Greenhouse-Geisser	3.493	1.000	3.493	1.079	.304	.022	1.079	.175
	Huynh-Feldt	3.493	1.000	3.493	1.079	.304	.022	1.079	.175
	Lower-bound	3.493	1.000	3.493	1.079	.304	.022	1.079	.175
Error(Tests)	Sphericity Assumed	155.304	48	3.236					
	Greenhouse-Geisser	155.304	48.000	3.236					
	Huynh-Feldt	155.304	48.000	3.236					
	Lower-bound	155.304	48.000	3.236					

a. Computed using alpha =

Table 36

Comparing control-value (condition 1) vs. no messages (control group 2)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	1.144	1	1.144	.373	.544	.008	.373	.092
	Greenhouse-Geisser	1.144	1.000	1.144	.373	.544	.008	.373	.092
	Huynh-Feldt	1.144	1.000	1.144	.373	.544	.008	.373	.092
	Lower-bound	1.144	1.000	1.144	.373	.544	.008	.373	.092
Tests * ControValue_VS_NoMessages	Sphericity Assumed	7.018	1	7.018	2.290	.137	.048	2.290	.316
	Greenhouse-Geisser	7.018	1.000	7.018	2.290	.137	.048	2.290	.316
	Huynh-Feldt	7.018	1.000	7.018	2.290	.137	.048	2.290	.316
	Lower-bound	7.018	1.000	7.018	2.290	.137	.048	2.290	.316
Tests * SystemHIGHFrustration	Sphericity Assumed	2.453	1	2.453	.801	.376	.017	.801	.141
	Greenhouse-Geisser	2.453	1.000	2.453	.801	.376	.017	.801	.141
	Huynh-Feldt	2.453	1.000	2.453	.801	.376	.017	.801	.141
	Lower-bound	2.453	1.000	2.453	.801	.376	.017	.801	.141
Tests * ControValue_VS_NoMessages * SystemHIGHFrustration	Sphericity Assumed	6.741	1	6.741	2.200	.145	.047	2.200	.306
	Greenhouse-Geisser	6.741	1.000	6.741	2.200	.145	.047	2.200	.306
	Huynh-Feldt	6.741	1.000	6.741	2.200	.145	.047	2.200	.306
	Lower-bound	6.741	1.000	6.741	2.200	.145	.047	2.200	.306
Error(Tests)	Sphericity Assumed	137.881	45	3.064					
	Greenhouse-Geisser	137.881	45.000	3.064					
	Huynh-Feldt	137.881	45.000	3.064					
	Lower-bound	137.881	45.000	3.064					

a. Computed using alpha =

Table 37

Comparing social identity (condition 2) vs. non-motivational messages (control group 1)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	15.974	1	15.974	5.170	.028	.099	5.170	.605
	Greenhouse-Geisser	15.974	1.000	15.974	5.170	.028	.099	5.170	.605
	Huynh-Feldt	15.974	1.000	15.974	5.170	.028	.099	5.170	.605
	Lower-bound	15.974	1.000	15.974	5.170	.028	.099	5.170	.605
Tests * Social_Identity_vs_NonMotivationalMessages	Sphericity Assumed	2.711	1	2.711	.877	.354	.018	.877	.151
	Greenhouse-Geisser	2.711	1.000	2.711	.877	.354	.018	.877	.151
	Huynh-Feldt	2.711	1.000	2.711	.877	.354	.018	.877	.151
	Lower-bound	2.711	1.000	2.711	.877	.354	.018	.877	.151
Tests * SystemHIGHFrustration	Sphericity Assumed	20.785	1	20.785	6.728	.013	.125	6.728	.719
	Greenhouse-Geisser	20.785	1.000	20.785	6.728	.013	.125	6.728	.719
	Huynh-Feldt	20.785	1.000	20.785	6.728	.013	.125	6.728	.719
	Lower-bound	20.785	1.000	20.785	6.728	.013	.125	6.728	.719
Tests * Social_Identity_vs_NonMotivationalMessages * SystemHIGHFrustration	Sphericity Assumed	2.009	1	2.009	.650	.424	.014	.650	.124
	Greenhouse-Geisser	2.009	1.000	2.009	.650	.424	.014	.650	.124
	Huynh-Feldt	2.009	1.000	2.009	.650	.424	.014	.650	.124
	Lower-bound	2.009	1.000	2.009	.650	.424	.014	.650	.124
Error(Tests)	Sphericity Assumed	145.206	47	3.089					
	Greenhouse-Geisser	145.206	47.000	3.089					
	Huynh-Feldt	145.206	47.000	3.089					
	Lower-bound	145.206	47.000	3.089					

a. Computed using alpha =

Table 38

Comparing social identity (condition 2) vs. no messages (control group 2)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	14.489	1	14.489	4.742	.035	.095	4.742	.568
	Greenhouse-Geisser	14.489	1.000	14.489	4.742	.035	.095	4.742	.568
	Huynh-Feldt	14.489	1.000	14.489	4.742	.035	.095	4.742	.568
	Lower-bound	14.489	1.000	14.489	4.742	.035	.095	4.742	.568
Tests * Social_Identity_vs_NoMessages	Sphericity Assumed	1.074	1	1.074	.352	.556	.008	.352	.089
	Greenhouse-Geisser	1.074	1.000	1.074	.352	.556	.008	.352	.089
	Huynh-Feldt	1.074	1.000	1.074	.352	.556	.008	.352	.089
	Lower-bound	1.074	1.000	1.074	.352	.556	.008	.352	.089
Tests * SystemHIGHFrustration	Sphericity Assumed	17.246	1	17.246	5.645	.022	.111	5.645	.642
	Greenhouse-Geisser	17.246	1.000	17.246	5.645	.022	.111	5.645	.642
	Huynh-Feldt	17.246	1.000	17.246	5.645	.022	.111	5.645	.642
	Lower-bound	17.246	1.000	17.246	5.645	.022	.111	5.645	.642
Tests * Social_Identity_vs_NoMessages * SystemHIGHFrustration	Sphericity Assumed	1.055	1	1.055	.345	.560	.008	.345	.089
	Greenhouse-Geisser	1.055	1.000	1.055	.345	.560	.008	.345	.089
	Huynh-Feldt	1.055	1.000	1.055	.345	.560	.008	.345	.089
	Lower-bound	1.055	1.000	1.055	.345	.560	.008	.345	.089
Error(Tests)	Sphericity Assumed	137.483	45	3.055					
	Greenhouse-Geisser	137.483	45.000	3.055					
	Huynh-Feldt	137.483	45.000	3.055					
	Lower-bound	137.483	45.000	3.055					

a. Computed using alpha =

Hypothesis 2

The second hypothesis that there will be a statistically significant difference in learning outcomes between conditions that provide intervention messages vs. the condition where no messages are provided (see Table 39).

Test. To test the second hypothesis, two-way mixed design rANOVA analysis was run comparing the message conditions (conditions 1, 2, 3, & 4) to the no-message condition (condition 5). There was not a statistically significant difference in learning outcomes between messages that gave feedback messages and the condition that did not (rANOVA): $F(1, 120) = 1.439$, $p = .233$, $\eta p2 = .012$, power = .222 (see Table 40).

Table 39

Means chart for messages vs. no message conditions rANOVA test

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
Messages	13.44	1.96	14.13	2.04	1.05
No Messages	13.87	1.63	14.30	1.82	0.43

Table 40

Two-way mixed design rANOVA for frustration interaction messages vs. no message conditions

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
Tests	Pillai's Trace	.010	1.160[b]	1.000	120.000	.284	.010	1.160	.188
	Wilks' Lambda	.990	1.160[b]	1.000	120.000	.284	.010	1.160	.188
	Hotelling's Trace	.010	1.160[b]	1.000	120.000	.284	.010	1.160	.188
	Roy's Largest Root	.010	1.160[b]	1.000	120.000	.284	.010	1.160	.188
Tests * FeedbackVSNoFeedback	Pillai's Trace	.012	1.439[b]	1.000	120.000	.233	.012	1.439	.222
	Wilks' Lambda	.988	1.439[b]	1.000	120.000	.233	.012	1.439	.222
	Hotelling's Trace	.012	1.439[b]	1.000	120.000	.233	.012	1.439	.222
	Roy's Largest Root	.012	1.439[b]	1.000	120.000	.233	.012	1.439	.222
Tests * SystemHIGHFrustration	Pillai's Trace	.017	2.111[b]	1.000	120.000	.149	.017	2.111	.302
	Wilks' Lambda	.983	2.111[b]	1.000	120.000	.149	.017	2.111	.302
	Hotelling's Trace	.018	2.111[b]	1.000	120.000	.149	.017	2.111	.302
	Roy's Largest Root	.018	2.111[b]	1.000	120.000	.149	.017	2.111	.302
Tests * FeedbackVSNoFeedback * SystemHIGHFrustration	Pillai's Trace	.010	1.262[b]	1.000	120.000	.263	.010	1.262	.200
	Wilks' Lambda	.990	1.262[b]	1.000	120.000	.263	.010	1.262	.200
	Hotelling's Trace	.011	1.262[b]	1.000	120.000	.263	.010	1.262	.200
	Roy's Largest Root	.011	1.262[b]	1.000	120.000	.263	.010	1.262	.200

a. Design: Intercept + FeedbackVSNoFeedback + SystemHIGHFrustration + FeedbackVSNoFeedback * SystemHIGHFrustration
Within Subjects Design: Tests

b. Exact statistic

c. Computed using alpha =

Hypothesis 3

It is hypothesized that there will be statistically significant differences between motivational feedback conditions when addressing frustration in game-based learning environment.

Test. To test the difference in pre-post test scores by motivational condition, a two-way mixed design rANOVA analysis, was conducted. There was a statistically significant difference in positive learning outcomes between motivational feedback conditions (rANOVA), $F(2, 70) = 3.917$, $p = .024$, $\eta p2 = .101$, power = .688, and the 3-way tests-frustration-condition interaction was statistically significant (rANOVA), $F(2, 70) = 3.539$, $p = .034$, $\eta p2 = .092$, power = .641 (see Tables 41 and 42).

Table 41

Means chart for rANOVA testing between motivational conditions

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
1 Control Value	13.115	1.966	13.731	2.219	0.61
2 Social Identity	13.885	1.904	14.231	2.286	0.35
3 Self-Efficacy	13.167	2.057	14.333	1.926	1.166

Table 42

Two-way mixed design rANOVA for motivational feedback conditions

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
Tests	Pillai's Trace	.001	.093[b]	1.000	70.000	.762	.001	.093	.060
	Wilks' Lambda	.999	.093[b]	1.000	70.000	.762	.001	.093	.060
	Hotelling's Trace	.001	.093[b]	1.000	70.000	.762	.001	.093	.060
	Roy's Largest Root	.001	.093[b]	1.000	70.000	.762	.001	.093	.060
Tests * Motivational_Feedback_Messages	Pillai's Trace	.101	3.917[b]	2.000	70.000	.024	.101	7.835	.688
	Wilks' Lambda	.899	3.917[b]	2.000	70.000	.024	.101	7.835	.688
	Hotelling's Trace	.112	3.917[b]	2.000	70.000	.024	.101	7.835	.688
	Roy's Largest Root	.112	3.917[b]	2.000	70.000	.024	.101	7.835	.688
Tests * SystemHIGHFrustration	Pillai's Trace	.000	.004[b]	1.000	70.000	.953	.000	.004	.050
	Wilks' Lambda	1.000	.004[b]	1.000	70.000	.953	.000	.004	.050
	Hotelling's Trace	.000	.004[b]	1.000	70.000	.953	.000	.004	.050
	Roy's Largest Root	.000	.004[b]	1.000	70.000	.953	.000	.004	.050
Tests * Motivational_Feedback_Messages * SystemHIGHFrustration	Pillai's Trace	.092	3.539[b]	2.000	70.000	.034	.092	7.078	.641
	Wilks' Lambda	.908	3.539[b]	2.000	70.000	.034	.092	7.078	.641
	Hotelling's Trace	.101	3.539[b]	2.000	70.000	.034	.092	7.078	.641
	Roy's Largest Root	.101	3.539[b]	2.000	70.000	.034	.092	7.078	.641

a. Design: Intercept + Motivational_Feedback_Messages + SystemHIGHFrustration + Motivational_Feedback_Messages * SystemHIGHFrustration
Within Subjects Design: Tests

b. Exact statistic

c. Computed using alpha =

Testing simple main effects by condition. As previously mentioned, post-hoc analyses for rANOVA's include running simple main effect analyses, separately at each level of the factor under investigation (Keselman, 1998; Verma, 2016; Weinberg & Abramowitz, 2002) (see Table 43). As such, the simple main effects of each motivational condition was analyzed separately at each level of condition. The results of a simple main effects analysis yielded a statistically significant difference in higher positive learning outcomes in the self-efficacy condition (condition 3) than the control-value (condition 1) and social identity conditions (condition 2) (rANOVA), $F(1, 22) = 5.09$, $p = .034$, $\eta p2 = .188$, power = .578; still significant after Benjamini-Hochberg alpha adjustment: $p = .034 <$ B-H $\alpha = 0.016$ (see Table 44).

Table 43

Simple main effects rANOVA analysis by condition

Multivariate Tests[a]

Condition	Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
1_CValue	Tests	Pillai's Trace	.025	.625[b]	1.000	24.000	.437	.025	.625	.118
		Wilks' Lambda	.975	.625[b]	1.000	24.000	.437	.025	.625	.118
		Hotelling's Trace	.026	.625[b]	1.000	24.000	.437	.025	.625	.118
		Roy's Largest Root	.026	.625[b]	1.000	24.000	.437	.025	.625	.118
	Tests * SystemHIGHFrustration	Pillai's Trace	.010	.245[b]	1.000	24.000	.625	.010	.245	.076
		Wilks' Lambda	.990	.245[b]	1.000	24.000	.625	.010	.245	.076
		Hotelling's Trace	.010	.245[b]	1.000	24.000	.625	.010	.245	.076
		Roy's Largest Root	.010	.245[b]	1.000	24.000	.625	.010	.245	.076
2_Socl_ID	Tests	Pillai's Trace	.100	2.667[b]	1.000	24.000	.115	.100	2.667	.348
		Wilks' Lambda	.900	2.667[b]	1.000	24.000	.115	.100	2.667	.348
		Hotelling's Trace	.111	2.667[b]	1.000	24.000	.115	.100	2.667	.348
		Roy's Largest Root	.111	2.667[b]	1.000	24.000	.115	.100	2.667	.348
	Tests * SystemHIGHFrustration	Pillai's Trace	.112	3.021[b]	1.000	24.000	.095	.112	3.021	.386
		Wilks' Lambda	.888	3.021[b]	1.000	24.000	.095	.112	3.021	.386
		Hotelling's Trace	.126	3.021[b]	1.000	24.000	.095	.112	3.021	.386
		Roy's Largest Root	.126	3.021[b]	1.000	24.000	.095	.112	3.021	.386
3_Self_Ef	Tests	Pillai's Trace	.250	7.320[b]	1.000	22.000	.013	.250	7.320	.734
		Wilks' Lambda	.750	7.320[b]	1.000	22.000	.013	.250	7.320	.734
		Hotelling's Trace	.333	7.320[b]	1.000	22.000	.013	.250	7.320	.734
		Roy's Largest Root	.333	7.320[b]	1.000	22.000	.013	.250	7.320	.734
	Tests * SystemHIGHFrustration	Pillai's Trace	.188	5.090[b]	1.000	22.000	.034	.188	5.090	.578
		Wilks' Lambda	.812	5.090[b]	1.000	22.000	.034	.188	5.090	.578
		Hotelling's Trace	.231	5.090[b]	1.000	22.000	.034	.188	5.090	.578
		Roy's Largest Root	.231	5.090[b]	1.000	22.000	.034	.188	5.090	.578

Table 44

Benjamini-Hochberg alpha post-hoc adjustments for motivational conditions

Intervention	*df*	*F*	*Sig*	*Adjusted* α	ηp2	*Power*
**Self efficacy*	1	5.09	0.034	0.016	0.188	0.578
Social identity	1	3.021	0.095	0.033	0.112	0.386
Control-value	1	0.245	0.625	0.05	0.01	0.076

* statistically significant after Benjamini-Hochberg correction

Follow-up tests were run conducting pairwise rANOVA's comparing each motivational condition to each other (see summary of findings in Table 45). This analysis indicated that the self-efficacy condition (N=24) had higher positive learning outcomes than the social identity condition (N = 26); (rANOVA): $F(1, 46) = 8.121$, $p = .007$, ηp2 = .150, power = .797 (see Table

45) ; still significant after making the Benjamini-Hochberg alpha adjustment: p =.007 < B-H α = 0.016 (see Table 46).

Table 45

Comparing self-efficacy condition (condition 3) vs. social identity condition (condition 2)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	.000	1	.000	.000	.992	.000	.000	.050
	Greenhouse-Geisser	.000	1.000	.000	.000	.992	.000	.000	.050
	Huynh-Feldt	.000	1.000	.000	.000	.992	.000	.000	.050
	Lower-bound	.000	1.000	.000	.000	.992	.000	.000	.050
Tests * SystemHIGHFrustration	Sphericity Assumed	.277	1	.277	.084	.773	.002	.084	.059
	Greenhouse-Geisser	.277	1.000	.277	.084	.773	.002	.084	.059
	Huynh-Feldt	.277	1.000	.277	.084	.773	.002	.084	.059
	Lower-bound	.277	1.000	.277	.084	.773	.002	.084	.059
Tests * SocialIdentity_vs_SelfEfficacy	Sphericity Assumed	26.662	1	26.662	8.121	.007	.150	8.121	.797
	Greenhouse-Geisser	26.662	1.000	26.662	8.121	.007	.150	8.121	.797
	Huynh-Feldt	26.662	1.000	26.662	8.121	.007	.150	8.121	.797
	Lower-bound	26.662	1.000	26.662	8.121	.007	.150	8.121	.797
Tests * SocialIdentity_vs_SelfEfficacy * SystemHIGHFrustration	Sphericity Assumed	24.240	1	24.240	7.383	.009	.138	7.383	.758
	Greenhouse-Geisser	24.240	1.000	24.240	7.383	.009	.138	7.383	.758
	Huynh-Feldt	24.240	1.000	24.240	7.383	.009	.138	7.383	.758
	Lower-bound	24.240	1.000	24.240	7.383	.009	.138	7.383	.758
Error(Tests)	Sphericity Assumed	151.030	46	3.283					
	Greenhouse-Geisser	151.030	46.000	3.283					
	Huynh-Feldt	151.030	46.000	3.283					
	Lower-bound	151.030	46.000	3.283					

a. Computed using alpha =

Table 46

Benjamini-Hochberg alpha post-hoc adjustments comparing motivational conditions

Motivational condition (A)	*Motivational condition (B)*	*df*	*F*	*Sig*	*Adjusted* α	ηp2	*Power*
**Self*-efficacy	Social identity	1	8.121	.007	0.016	.150	.797
Control-value	Social identity	1	3.127	.076	0.033	.064	.427
Control-value	Self-efficacy	1	2.338	.133	0.05	.048	.322

* statistically significant after Benjamini-Hochberg correction

There was no significant statistical difference between the control value (*N = 26) vs.* social identity (*N* = 26) (rANOVA): F(1, 48) = 3.127, p = .076, ηp2 = .064, power = .427; p =.076 > B-H α = 0.033 (see Table 47). There was also no significant statistical difference between control value (*N* = 26) and self-efficacy (*N* = 26); $F(1, 46) = 2.338$, $p = .133$, $\eta p2 = .048$, power = .322; $p = .133 >$ B-H $\alpha = 0.05$ (see Table 48).

Table 47

Comparing control value (condition 1) vs. social identity (condition 2)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	4.519	1	4.519	1.219	.275	.025	1.219	.191
	Greenhouse-Geisser	4.519	1.000	4.519	1.219	.275	.025	1.219	.191
	Huynh-Feldt	4.519	1.000	4.519	1.219	.275	.025	1.219	.191
	Lower-bound	4.519	1.000	4.519	1.219	.275	.025	1.219	.191
Tests * SystemHIGHFrustration	Sphericity Assumed	6.350	1	6.350	1.713	.197	.034	1.713	.250
	Greenhouse-Geisser	6.350	1.000	6.350	1.713	.197	.034	1.713	.250
	Huynh-Feldt	6.350	1.000	6.350	1.713	.197	.034	1.713	.250
	Lower-bound	6.350	1.000	6.350	1.713	.197	.034	1.713	.250
Tests * ControlValue_vs_SocialIdentity	Sphericity Assumed	12.184	1	12.184	3.287	.076	.064	3.287	.427
	Greenhouse-Geisser	12.184	1.000	12.184	3.287	.076	.064	3.287	.427
	Huynh-Feldt	12.184	1.000	12.184	3.287	.076	.064	3.287	.427
	Lower-bound	12.184	1.000	12.184	3.287	.076	.064	3.287	.427
Tests * ControlValue_vs_SocialIdentity * SystemHIGHFrustration	Sphericity Assumed	11.591	1	11.591	3.127	.083	.061	3.127	.410
	Greenhouse-Geisser	11.591	1.000	11.591	3.127	.083	.061	3.127	.410
	Huynh-Feldt	11.591	1.000	11.591	3.127	.083	.061	3.127	.410
	Lower-bound	11.591	1.000	11.591	3.127	.083	.061	3.127	.410
Error(Tests)	Sphericity Assumed	177.934	48	3.707					
	Greenhouse-Geisser	177.934	48.000	3.707					
	Huynh-Feldt	177.934	48.000	3.707					
	Lower-bound	177.934	48.000	3.707					

a. Computed using alpha =

Table 48

Comparing control value (condition 1) vs. self-efficacy condition (condition 3)

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	20.912	1	20.912	6.353	.015	.121	6.353	.694
	Greenhouse-Geisser	20.912	1.000	20.912	6.353	.015	.121	6.353	.694
	Huynh-Feldt	20.912	1.000	20.912	6.353	.015	.121	6.353	.694
	Lower-bound	20.912	1.000	20.912	6.353	.015	.121	6.353	.694
Tests * ControlValue_vs_SelfEfficacy	Sphericity Assumed	7.697	1	7.697	2.338	.133	.048	2.338	.322
	Greenhouse-Geisser	7.697	1.000	7.697	2.338	.133	.048	2.338	.322
	Huynh-Feldt	7.697	1.000	7.697	2.338	.133	.048	2.338	.322
	Lower-bound	7.697	1.000	7.697	2.338	.133	.048	2.338	.322
Tests * SystemHIGHFrustration	Sphericity Assumed	13.124	1	13.124	3.987	.052	.080	3.987	.498
	Greenhouse-Geisser	13.124	1.000	13.124	3.987	.052	.080	3.987	.498
	Huynh-Feldt	13.124	1.000	13.124	3.987	.052	.080	3.987	.498
	Lower-bound	13.124	1.000	13.124	3.987	.052	.080	3.987	.498
Tests * ControlValue_vs_SelfEfficacy * SystemHIGHFrustration	Sphericity Assumed	6.207	1	6.207	1.885	.176	.039	1.885	.270
	Greenhouse-Geisser	6.207	1.000	6.207	1.885	.176	.039	1.885	.270
	Huynh-Feldt	6.207	1.000	6.207	1.885	.176	.039	1.885	.270
	Lower-bound	6.207	1.000	6.207	1.885	.176	.039	1.885	.270
Error(Tests)	Sphericity Assumed	151.427	46	3.292					
	Greenhouse-Geisser	151.427	46.000	3.292					
	Huynh-Feldt	151.427	46.000	3.292					
	Lower-bound	151.427	46.000	3.292					

a. Computed using alpha =

Hypothesis 4

The fourth hypothesis stated that a participant's perceived presence in a game-based learning environment will mediate the differences in learning outcome between motivational feedback conditions (see Table 49).

Test. To test whether presence had a mediating effect between condition, a two-way mixed design rANOVA analysis was conducted. Presence did not have a statistically significantly effect associated with learning outcomes, (rANOVA): $F(1,114)= 1.639$, $p = .203$, $\eta p2 = .014$, power = .246, and there was not a statistically significant interaction between presence and condition on pre-post test scores, (rANOVA): $F(4,114)=0.162$, $p=0.957$, $\eta p2 = .006$, power = .083 (see Table 50).

Table 49

Means chart rANOVA test with interaction of presence by condition

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
1 Control Value	13.115	1.966	13.731	2.219	0.61
2 Social Identity	13.885	1.904	14.231	2.286	0.35
3 Self-Efficacy	13.167	2.057	14.333	1.926	1.166
4 1c - NonMotivational	13.56	1.938	14.240	1.715	0.68
5 2c - NoMessages	13.87	1.632	14.304	1.820	0.434

Table 50

Two-way mixed design rANOVA, tests by condition and presence

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	2.528	1	2.528	.754	.387	.007	.754	.138
	Greenhouse-Geisser	2.528	1.000	2.528	.754	.387	.007	.754	.138
	Huynh-Feldt	2.528	1.000	2.528	.754	.387	.007	.754	.138
	Lower-bound	2.528	1.000	2.528	.754	.387	.007	.754	.138
Tests * Condition	Sphericity Assumed	1.718	4	.430	.128	.972	.004	.512	.076
	Greenhouse-Geisser	1.718	4.000	.430	.128	.972	.004	.512	.076
	Huynh-Feldt	1.718	4.000	.430	.128	.972	.004	.512	.076
	Lower-bound	1.718	4.000	.430	.128	.972	.004	.512	.076
Tests * PresenceScore	Sphericity Assumed	5.499	1	5.499	1.639	.203	.014	1.639	.246
	Greenhouse-Geisser	5.499	1.000	5.499	1.639	.203	.014	1.639	.246
	Huynh-Feldt	5.499	1.000	5.499	1.639	.203	.014	1.639	.246
	Lower-bound	5.499	1.000	5.499	1.639	.203	.014	1.639	.246
Tests * Condition * PresenceScore	Sphericity Assumed	2.169	4	.542	.162	.957	.006	.647	.083
	Greenhouse-Geisser	2.169	4.000	.542	.162	.957	.006	.647	.083
	Huynh-Feldt	2.169	4.000	.542	.162	.957	.006	.647	.083
	Lower-bound	2.169	4.000	.542	.162	.957	.006	.647	.083
Error(Tests)	Sphericity Assumed	382.521	114	3.355					
	Greenhouse-Geisser	382.521	114.000	3.355					
	Huynh-Feldt	382.521	114.000	3.355					
	Lower-bound	382.521	114.000	3.355					

a. Computed using alpha =

Hypothesis 5

The fifth hypothesis claims that there will be a difference on learning outcomes between motivational feedback conditions based on a person's level of grit.

Test. To test whether grit had a moderating effect on outcomes across conditions, a two-way mixed design rANOVA analysis was conducted (see Table 51). There was a statistically significant difference between conditions and positive learning outcomes controlling for grit and interaction of grit by condition and learning outcomes: (rANOVA): $F(4,114) = 2.631$, $p = .038$, $\eta p2 = .085$, power = .721. There was also a statistically significant interaction effect of grit by condition and learning outcomes (rANOVA): $F(4,114) = 2.903$, $p = .025$, $\eta p2 = .092$, power = .768.

Table 51

Two-way mixed design rANOVA, tests by condition and grit

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Tests	Sphericity Assumed	10.597	1	10.597	3.422	.067	.029	3.422	.450
	Greenhouse-Geisser	10.597	1.000	10.597	3.422	.067	.029	3.422	.450
	Huynh-Feldt	10.597	1.000	10.597	3.422	.067	.029	3.422	.450
	Lower-bound	10.597	1.000	10.597	3.422	.067	.029	3.422	.450
Tests * Condition	Sphericity Assumed	32.586	4	8.146	2.631	.038	.085	10.524	.721
	Greenhouse-Geisser	32.586	4.000	8.146	2.631	.038	.085	10.524	.721
	Huynh-Feldt	32.586	4.000	8.146	2.631	.038	.085	10.524	.721
	Lower-bound	32.586	4.000	8.146	2.631	.038	.085	10.524	.721
Tests * Grit_Score	Sphericity Assumed	6.206	1	6.206	2.004	.160	.017	2.004	.289
	Greenhouse-Geisser	6.206	1.000	6.206	2.004	.160	.017	2.004	.289
	Huynh-Feldt	6.206	1.000	6.206	2.004	.160	.017	2.004	.289
	Lower-bound	6.206	1.000	6.206	2.004	.160	.017	2.004	.289
Tests * Condition * Grit_Score	Sphericity Assumed	35.956	4	8.989	2.903	.025	.092	11.613	.768
	Greenhouse-Geisser	35.956	4.000	8.989	2.903	.025	.092	11.613	.768
	Huynh-Feldt	35.956	4.000	8.989	2.903	.025	.092	11.613	.768
	Lower-bound	35.956	4.000	8.989	2.903	.025	.092	11.613	.768
Error(Tests)	Sphericity Assumed	352.971	114	3.096					
	Greenhouse-Geisser	352.971	114.000	3.096					
	Huynh-Feldt	352.971	114.000	3.096					
	Lower-bound	352.971	114.000	3.096					

a. Computed using alpha =

With this significant interaction, an analysis on the simple effects of grit by condition were conducted, running simple main effect analyses separately at each level of condition

(Keselman, 1998; Verma, 2016; Weinberg & Abramowitz, 2002). The analysis of the simple means showed that grit had a statistically significant effect in learning outcomes only within the control-value condition (condition 1) (see Table 51), $F(1, 24)=7.304$, $p = .012$, $\eta p2 = .233$, power = .737 (see tables 52 & 53).

Table 52

Means chart rANOVA test with interaction of grit by condition

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$
1 Control Value	13.115	1.966	13.731	2.219	0.61
2 Social Identity	13.885	1.904	14.231	2.286	0.35
3 Self-Efficacy	13.167	2.057	14.333	1.926	1.166
4 1c - NonMotivational	13.56	1.938	14.240	1.715	0.68
5 2c - NoMessages	13.87	1.632	14.304	1.820	0.434

Table 53

Test for simple of effects of grit by condition

Multivariate Tests[a]

Condition	Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[c]
1_CValue	Tests	Pillai's Trace	.253	8.148[b]	1.000	24.000	.009	.253	8.148	.782
		Wilks' Lambda	.747	8.148[b]	1.000	24.000	.009	.253	8.148	.782
		Hotelling's Trace	.340	8.148[b]	1.000	24.000	.009	.253	8.148	.782
		Roy's Largest Root	.340	8.148[b]	1.000	24.000	.009	.253	8.148	.782
	Tests * Grit_Score	Pillai's Trace	.233	7.304[b]	1.000	24.000	.012	.233	7.304	.737
		Wilks' Lambda	.767	7.304[b]	1.000	24.000	.012	.233	7.304	.737
		Hotelling's Trace	.304	7.304[b]	1.000	24.000	.012	.233	7.304	.737
		Roy's Largest Root	.304	7.304[b]	1.000	24.000	.012	.233	7.304	.737
2_Socl_ID	Tests	Pillai's Trace	.010	.232[b]	1.000	24.000	.635	.010	.232	.075
		Wilks' Lambda	.990	.232[b]	1.000	24.000	.635	.010	.232	.075
		Hotelling's Trace	.010	.232[b]	1.000	24.000	.635	.010	.232	.075
		Roy's Largest Root	.010	.232[b]	1.000	24.000	.635	.010	.232	.075
	Tests * Grit_Score	Pillai's Trace	.007	.167[b]	1.000	24.000	.686	.007	.167	.068
		Wilks' Lambda	.993	.167[b]	1.000	24.000	.686	.007	.167	.068
		Hotelling's Trace	.007	.167[b]	1.000	24.000	.686	.007	.167	.068
		Roy's Largest Root	.007	.167[b]	1.000	24.000	.686	.007	.167	.068
3_Self_Ef	Tests	Pillai's Trace	.066	1.550[b]	1.000	22.000	.226	.066	1.550	.222
		Wilks' Lambda	.934	1.550[b]	1.000	22.000	.226	.066	1.550	.222
		Hotelling's Trace	.070	1.550[b]	1.000	22.000	.226	.066	1.550	.222
		Roy's Largest Root	.070	1.550[b]	1.000	22.000	.226	.066	1.550	.222
	Tests * Grit_Score	Pillai's Trace	.116	2.893[b]	1.000	22.000	.103	.116	2.893	.370
		Wilks' Lambda	.884	2.893[b]	1.000	22.000	.103	.116	2.893	.370
		Hotelling's Trace	.132	2.893[b]	1.000	22.000	.103	.116	2.893	.370
		Roy's Largest Root	.132	2.893[b]	1.000	22.000	.103	.116	2.893	.370

4_1c_NonMotv	Tests	Pillai's Trace	.002	.035[b]	1.000	23.000	.852	.002	.035	.054
		Wilks' Lambda	.998	.035[b]	1.000	23.000	.852	.002	.035	.054
		Hotelling's Trace	.002	.035[b]	1.000	23.000	.852	.002	.035	.054
		Roy's Largest Root	.002	.035[b]	1.000	23.000	.852	.002	.035	.054
	Tests * Grit_Score	Pillai's Trace	.008	.175[b]	1.000	23.000	.679	.008	.175	.069
		Wilks' Lambda	.992	.175[b]	1.000	23.000	.679	.008	.175	.069
		Hotelling's Trace	.008	.175[b]	1.000	23.000	.679	.008	.175	.069
		Roy's Largest Root	.008	.175[b]	1.000	23.000	.679	.008	.175	.069
5_2c_NoMessages	Tests	Pillai's Trace	.123	2.952[b]	1.000	21.000	.100	.123	2.952	.375
		Wilks' Lambda	.877	2.952[b]	1.000	21.000	.100	.123	2.952	.375
		Hotelling's Trace	.141	2.952[b]	1.000	21.000	.100	.123	2.952	.375
		Roy's Largest Root	.141	2.952[b]	1.000	21.000	.100	.123	2.952	.375
	Tests * Grit_Score	Pillai's Trace	.112	2.640[b]	1.000	21.000	.119	.112	2.640	.341
		Wilks' Lambda	.888	2.640[b]	1.000	21.000	.119	.112	2.640	.341
		Hotelling's Trace	.126	2.640[b]	1.000	21.000	.119	.112	2.640	.341
		Roy's Largest Root	.126	2.640[b]	1.000	21.000	.119	.112	2.640	.341

a. Design: Intercept + Grit_Score
Within Subjects Design: Tests

However, in examining the Benjamini-Hochberg alpha adjustments, the control-value condition marginally misses significance $p = .012 > \text{B-H } \alpha = 0.01$ (see Table 54).

Table 54

Benjamini-Hochberg alpha adjustments pre-post-test scores by condition by grit

*Condition	p-value	B-H α
Control value	0.012	0.01
Social identity	0.686	0.05
Self-efficacy	0.103	0.02
Non-motivational messages	0.679	0.04
No messages	0.119	0.03

** No condition reached significance after adjusting with Benjamini-Hochberg alpha*

Yet, in conducting a follow up analysis, splitting the data further into high and low grit groups, using the mean grit value of 3.80, the analysis indicated that low grit participants in the control-value theory condition (condition 1) had higher positive learning outcomes than high grit participants in the control-value theory condition, (rANOVA): $F(1, 25)= 35.000$, $p=0.001$, $\eta p2 =$

.883, power = .999 (see Table 53). After making Benjamini-Hochberg alpha adjustments, the low-grit condition remained significant: p = .001 < B-H α = 0.005 (see Table 56).

Table 55

Testing simple effects between conditions and between low (1) and high (2) grit participants

Tests of Within-Subjects Effects

Measure: MEASURE_1

Condition	HighLowGrit	Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
1_CValue	1.00	Tests	Sphericity Assumed	25.000	1	25.000	35.000	.001	.833	35.000	.999
			Greenhouse-Geisser	25.000	1.000	25.000	35.000	.001	.833	35.000	.999
			Huynh-Feldt	25.000	1.000	25.000	35.000	.001	.833	35.000	.999
			Lower-bound	25.000	1.000	25.000	35.000	.001	.833	35.000	.999
		Error(Tests)	Sphericity Assumed	5.000	7	.714					
			Greenhouse-Geisser	5.000	7.000	.714					
			Huynh-Feldt	5.000	7.000	.714					
			Lower-bound	5.000	7.000	.714					
	2.00	Tests	Sphericity Assumed	.444	1	.444	.117	.736	.007	.117	.062
			Greenhouse-Geisser	.444	1.000	.444	.117	.736	.007	.117	.062
			Huynh-Feldt	.444	1.000	.444	.117	.736	.007	.117	.062
			Lower-bound	.444	1.000	.444	.117	.736	.007	.117	.062
		Error(Tests)	Sphericity Assumed	64.556	17	3.797					
			Greenhouse-Geisser	64.556	17.000	3.797					
			Huynh-Feldt	64.556	17.000	3.797					
			Lower-bound	64.556	17.000	3.797					
2_Socl_ID	1.00	Tests	Sphericity Assumed	5.500	1	5.500	1.618	.232	.139	1.618	.211
			Greenhouse-Geisser	5.500	1.000	5.500	1.618	.232	.139	1.618	.211
			Huynh-Feldt	5.500	1.000	5.500	1.618	.232	.139	1.618	.211
			Lower-bound	5.500	1.000	5.500	1.618	.232	.139	1.618	.211
		Error(Tests)	Sphericity Assumed	34.000	10	3.400					
			Greenhouse-Geisser	34.000	10.000	3.400					
			Huynh-Feldt	34.000	10.000	3.400					
			Lower-bound	34.000	10.000	3.400					
	2.00	Tests	Sphericity Assumed	.133	1	.133	.030	.865	.002	.030	.053
			Greenhouse-Geisser	.133	1.000	.133	.030	.865	.002	.030	.053
			Huynh-Feldt	.133	1.000	.133	.030	.865	.002	.030	.053
			Lower-bound	.133	1.000	.133	.030	.865	.002	.030	.053
		Error(Tests)	Sphericity Assumed	61.867	14	4.419					
			Greenhouse-Geisser	61.867	14.000	4.419					
			Huynh-Feldt	61.867	14.000	4.419					
			Lower-bound	61.867	14.000	4.419					

3_Self_Ef	1.00	Tests	Sphericity Assumed	1.786	1	1.786	.915	.376	.132	.915	.128
			Greenhouse-Geisser	1.786	1.000	1.786	.915	.376	.132	.915	.128
			Huynh-Feldt	1.786	1.000	1.786	.915	.376	.132	.915	.128
			Lower-bound	1.786	1.000	1.786	.915	.376	.132	.915	.128
		Error(Tests)	Sphericity Assumed	11.714	6	1.952					
			Greenhouse-Geisser	11.714	6.000	1.952					
			Huynh-Feldt	11.714	6.000	1.952					
			Lower-bound	11.714	6.000	1.952					
	2.00	Tests	Sphericity Assumed	15.559	1	15.559	3.893	.066	.196	3.893	.458
			Greenhouse-Geisser	15.559	1.000	15.559	3.893	.066	.196	3.893	.458
			Huynh-Feldt	15.559	1.000	15.559	3.893	.066	.196	3.893	.458
			Lower-bound	15.559	1.000	15.559	3.893	.066	.196	3.893	.458
		Error(Tests)	Sphericity Assumed	63.941	16	3.996					
			Greenhouse-Geisser	63.941	16.000	3.996					
			Huynh-Feldt	63.941	16.000	3.996					
			Lower-bound	63.941	16.000	3.996					
4_1c_NonMotv	1.00	Tests	Sphericity Assumed	3.200	1	3.200	1.618	.235	.152	1.618	.207
			Greenhouse-Geisser	3.200	1.000	3.200	1.618	.235	.152	1.618	.207
			Huynh-Feldt	3.200	1.000	3.200	1.618	.235	.152	1.618	.207
			Lower-bound	3.200	1.000	3.200	1.618	.235	.152	1.618	.207
		Error(Tests)	Sphericity Assumed	17.800	9	1.978					
			Greenhouse-Geisser	17.800	9.000	1.978					
			Huynh-Feldt	17.800	9.000	1.978					
			Lower-bound	17.800	9.000	1.978					
	2.00	Tests	Sphericity Assumed	2.700	1	2.700	.744	.403	.050	.744	.127
			Greenhouse-Geisser	2.700	1.000	2.700	.744	.403	.050	.744	.127
			Huynh-Feldt	2.700	1.000	2.700	.744	.403	.050	.744	.127
			Lower-bound	2.700	1.000	2.700	.744	.403	.050	.744	.127
		Error(Tests)	Sphericity Assumed	50.800	14	3.629					
			Greenhouse-Geisser	50.800	14.000	3.629					
			Huynh-Feldt	50.800	14.000	3.629					
			Lower-bound	50.800	14.000	3.629					
5_2c_NoMessages	1.00	Tests	Sphericity Assumed	4.500	1	4.500	3.000	.122	.273	3.000	.333
			Greenhouse-Geisser	4.500	1.000	4.500	3.000	.122	.273	3.000	.333
			Huynh-Feldt	4.500	1.000	4.500	3.000	.122	.273	3.000	.333
			Lower-bound	4.500	1.000	4.500	3.000	.122	.273	3.000	.333
		Error(Tests)	Sphericity Assumed	12.000	8	1.500					
			Greenhouse-Geisser	12.000	8.000	1.500					
			Huynh-Feldt	12.000	8.000	1.500					
			Lower-bound	12.000	8.000	1.500					
	2.00	Tests	Sphericity Assumed	.036	1	.036	.011	.916	.001	.011	.051
			Greenhouse-Geisser	.036	1.000	.036	.011	.916	.001	.011	.051
			Huynh-Feldt	.036	1.000	.036	.011	.916	.001	.011	.051
			Lower-bound	.036	1.000	.036	.011	.916	.001	.011	.051
		Error(Tests)	Sphericity Assumed	40.464	13	3.113					
			Greenhouse-Geisser	40.464	13.000	3.113					
			Huynh-Feldt	40.464	13.000	3.113					
			Lower-bound	40.464	13.000	3.113					

a. Computed using alpha =

Table 56

Benjamini-Hochberg alpha adjustments for significance of pre-post-test scores by condition by high/low grit

Condition	*Grit level*	*P- value*	*Adjusted α*
** Control-value*	*Low*	*.001*	*0.005*
Control-value	High	.736	0.04
Social Identity	Low	.232	0.02
Social Identity	High	.865	0.045
Self-efficacy	Low	.376	0.03
Self-efficacy	High	.066	0.01
Non-motivational messages	Low	.235	0.025
Non-motivational messages	High	.403	0.035
No messages	Low	.122	0.015
No messages	High	.916	0.05

**significant after Benjamini-Hochberg alpha adjustments*

In examining the means of the control-value condition (condition 1) between high and low grit participants, the data revealed that the low grit group had positive learning gains but the high grit group had negative learning gains (see Table 27 and Figure 27), although they did not have a statistically significant decline in performance.

A closer examination of the pre and post test scores were conducted for the control-value condition, and while there were two participants who scores seem to make them outliers (post test = 9 correct), these participants each had the same score of 9. These participants' data were not removed from the analysis as they were only one point outside of the normal range of post-test scores: 10-18 correct in the control-value condition.

Table 57

Means chart rANOVA analysis of high and low grit participants in control-value condition

Conditions	$\bar{x}_{PRE}$	[SD]	$\bar{x}_{POST}$	[SD]	$\bar{x}_{DIFF}$	df	F	Sig.	ηp2
1 Low Grit	12.38	1.69	14.88	2.04	+2.495	1, 22	35.00	0.001	0.833
2 High Grit	13.44	1.81	13.22	2.24	-0.22	1, 17	.444	0.736	0.007

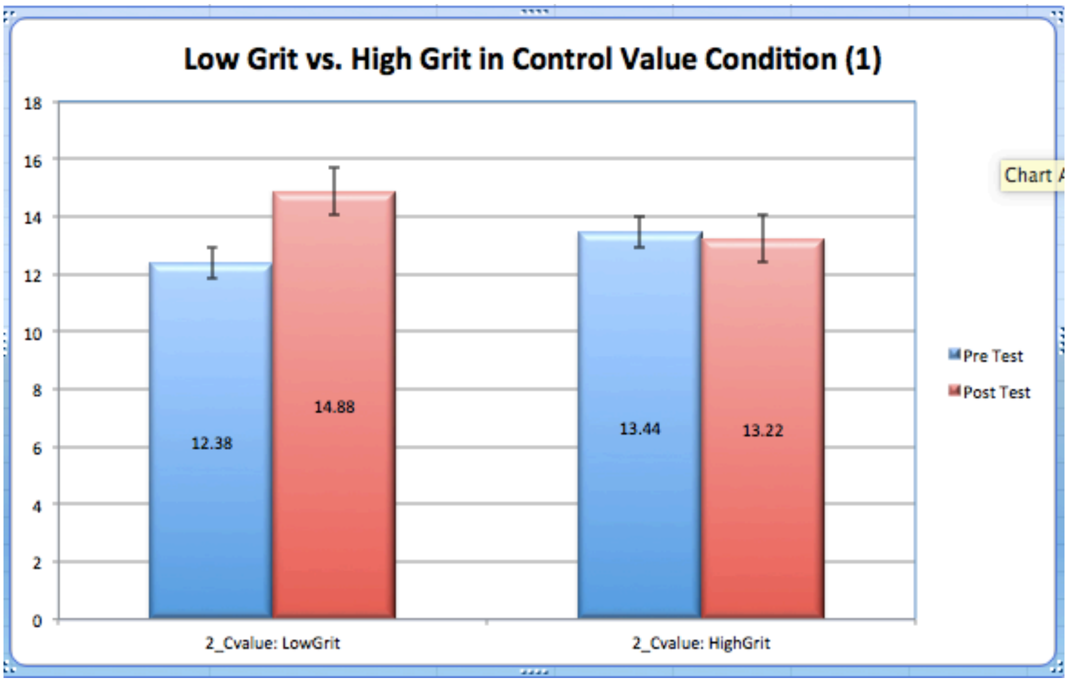


Figure 27. Pre-post test comparing high /low Grit groups in the control-value condition

Specifically, participants with low grit had +2.495 points in positive learning gains, whereas high grit participants had negative learning gains of -0.22 points, (see Figure 31). This seems to indicate that the control-value messages had a positive impact on participants with low grit scores, perhaps encouraging them to see the value in the experiment or the learning activity more broadly. For high grit participants, these participants might have seen the messages as unnecessary, annoying, or even frustrating -- perhaps even causing some disengagement with the experiment/learning activity.

Section 5. Summary Analyses Results

Data were not found to have violated the assumptions of normality for a mixed model, rANOVA. The assumptions necessary to run a two-way mixed model rANOVA's were also met.

In an examination of relevant covariates, since all effects were qualitatively the same when controlling for BROMP-detected frustration and system detected frustration, all subsequent analyses used the system detected frustration. The rationale behind this decision rested in the fact that the system-detected frustration was deemed a more continual measure of frustration.

Frustration was included in the model in order to determine whether it was a mediating factor that effected the main relationship of interest: the differences between motivational feedback conditions.

There was evidence in favor of the first hypothesis, that stated there would be a statistically significant difference between motivational feedback vs. non-motivational conditions when addressing frustration in a game-based learning environment. A further analysis indicated that there was a statistically significant difference between the self-efficacy condition (condition 3) and the non-motivational feedback messages (control group 1), as well as a statistically significant difference between the self-efficacy condition (condition 3) and the no message condition (control group 2).

There was no evidence in favor of the second hypothesis that stated there would be a statistically significant difference between conditions that provide intervention messages to address frustration vs. the condition where no messages are provided.

There was evidence to support the third hypothesis, that stated there would be statistically significant differences between motivational feedback conditions when addressing frustration in game-based learning environment. A comparison between conditions gave evidence that of all comparisons, there was a statistically significant difference between the social identity condition (condition 2) and the self-efficacy condition (condition 3). Doing a simple main effects analysis by condition, the self-efficacy condition was also the only motivational feedback condition that rendered a statistically significant difference in pre-post test outcomes when mediated by measures of frustration.

The fourth hypothesis had no evidence to support the claim that a participant's perceived presence in a game-based learning environment would yield statistically significant differences between motivational feedback conditions.

Lastly, there was evidence to support the fifth hypothesis that there would be a difference on learning outcomes between motivational feedback conditions based on a person's level of grit. Findings included that when splitting participants into high and low grit groups in the control-value condition, low grit participants had positive pre-post test outcomes as compared to the high grit participants who did not have statistically significant evidence of learning.

It is speculated that for high grit participants, motivational feedback messages that address the long term value of a learning experience has a negative effect – perhaps increasing frustration or some degree of disengagement.

Chapter V: DISCUSSION

Section 1. Overview

The first section will provide a summary of the results, including a table delineating the research questions, hypothesis, and analyses of results. Following this, second section will review the limitations of this experiment as well as recommendations for future work. Lastly, the third section provides a conclusion to the study and its findings.

Section 2. Summary of Results

Based on five primary hypotheses, the analyses and results are intended to inform research questions associated with motivational feedback delivered in a serious game environment to address learner frustration and promote learning in an intelligent tutoring system. The summary of the hypotheses testing (see Table 58) provides a summary of the results and how they related to the research questions posed in Chapter 1.

There was evidence in favor of the first hypothesis, that stated there would be a statistically significant difference between motivational feedback vs. non-motivational conditions when addressing frustration in a game-based learning environment. A further analysis indicated that there was a statistically significant difference between the self-efficacy condition (condition 3) and the non-motivational feedback messages (control group 1), as well as a statistically significant difference between the self-efficacy condition (condition 3) and the no message condition (control group 2). These results are in line with prior research that correlates motivational manipulations with greater cognitive processing (Locke & Braver, 2010; Maddox & Markman, 2010; Pessoa, 2009; Pessoa & Engelmann, 2010; Shohamy & Adcock, 2010).

There was no evidence in favor of the second hypothesis that stated conditions there would be a statistically significant difference between conditions that provide intervention messages to address frustration vs. the condition where no messages are provided. While prior research gave evidence that providing interventions in the form of messages has been shown to positively effect the learning of domain content in ITSs (Wagster, Tan, Wu, Biswas, & Schwartz, 2007; Roll, Aleven, McLaren, & Koedinger; 2011), these results seem to indicate that it is not merely providing the learner with any message that impacts cognition, particularly as it relates to addressing learner frustration. Rather, the context of the message appears to play an important role in addressing frustration and promoting learning.

There was evidence to support the third hypothesis, that stated there would be statistically significant differences between motivational feedback conditions when addressing frustration in game-based learning environment. A comparison between conditions gave evidence that of all comparisons, there was a statistically significant difference between the social identity condition (condition 2) and the self-efficacy condition (condition 3). Doing a simple main effects analysis by condition, the self-efficacy condition was also the only motivational feedback condition that rendered a statistically significant difference in pre-post test outcomes when mediated by measures of frustration.

Based on prior research that indicated that a “one size fits all” approach to affective feedback is unlikely to regulate emotional experiences such as frustration, (D’Mello, Strain, Olney, & Graesser, 2013), the results showed that the difference between the condition of self-efficacy and social identity was statistically significant, and that overall the condition of self-efficacy had a statistically significant difference in pre-post test outcomes as compared to the other three motivational feedback messages.

The fourth hypothesis stated that a participant’s perceived presence in a game-based learning environment will yield statistically significant differences between motivational feedback conditions. Indeed, capitalizing on the participants’ identity as a military population did not seem to motivate participants differently in comparison to the control conditions. These results might be mitigated by the fact that the 87.9% of the participants at USMA did not have prior military service, 54.0% of the participants were freshman, and the experiment was conducted in the beginning of the school year (September 2015).

It is likely that since the majority of participants have yet to attain active military status, their identity as members of the military had not been completely solidified and as a result would not respond positively to motivational feedback that targets their military identity. Alternatively, these results may also indicate that a motivational feedback message that targets long term goals and values is not effective when addressing an immediately frustrating event in a learning experience.

Lastly, the fifth hypothesis claimed that there would be a difference on learning outcomes between motivational feedback conditions based on a person's level of grit. This hypothesis was based on prior research that demonstrated the impact motivational feedback had depending upon whether groups were differentiated by low ability and high ability, or by being unmotivated and motivated (Burelson, 2006; Mayer et al., 2006; Rebolledo-Mendez et al., 2006). The results from the test of the fifth hypothesis also showed that in the control-value condition (condition 3), for those who had comparatively low grit, their outcome gains in pre-post test scores were positive, compared to the high grit participants who did not have statistically significant learning gains.

It is speculated that while the control-value messages may have a positive impact on participants with low grit scores, encouraging them to see the long-term value in the learning activity, this was not the case for high grit participants. For high grit participants, the control-value messages might have been perceived as annoying, distracting, or perhaps even more frustrating, likely contributing to some disengagement in the learning activity. However, given that these findings rest on a small sample size, the generalizability of these findings are accordingly very limited.

The next sections focus on the limitations of this study and will conclude with recommendations for future work.

Table 58

Summary of results associated research questions and hypotheses

Question	Associated Hypothesis	Implications of Results
RQ1: Is there a difference between motivational and non-motivational conditions when addressing learner frustration in a game-based learning environment?	This hypothesis stated that there will be a statistically significant difference between motivational feedback vs. non-motivational conditions when addressing frustration in a game-based learning environment.	Inclusion of self-efficacy motivational feedback was found to have a significant effect on learning outcomes as compared to non-motivational feedback and conditions with no messages.
RQ2: Is there a difference between conditions that provide a message than those that do not when addressing learner frustration in a game-based learning environment?	It is hypothesized that there will be a statistically significant difference between conditions that provide intervention messages to address frustration vs. the condition where no messages are provided.	There was not statistically significant difference in learning outcomes between conditions that do not provide messages than those that do.
RQ3: Is there a difference between motivational feedback conditions when addressing participants' frustration in a game-based learning activity?	It is hypothesized that there will be statistically significant differences between motivational feedback conditions when addressing frustration in game-based learning environment.	There was a statistically significant difference of pre-post test outcomes between motivational conditions, specifically in the self-efficacy condition compared to the social identity condition when factoring in the mediating effect of frustration.
RQ4: Does a participant's perceived presence in a game-based learning environment yield differences between motivational feedback conditions?	The fourth hypothesis stated that a participant's perceived presence in a game-based learning environment will mediate the differences between motivational feedback conditions.	There was no statistical significant effect of presence between conditions on pre-post test outcomes.

RQ5: Is there a difference on learning outcomes between motivational feedback conditions based on a person's level of grit ?	The fifth hypothesis claims that there will be a difference on learning outcomes between motivational feedback conditions based on a person's level of grit.	Grit had a significant moderating effect on pre-post test outcomes by condition. Grit was significantly associated with pre-post test gains for low grit participants in the control-value condition (condition 1) yielding positive outcomes, whereas there was no statistically significant difference in learning outcomes for the high grit participants who yielded marginally negative pre-post test gains.

Section 3. Limitations and Future Work

Limitations

The current study only compared three types of motivational theories (control-value, social identity, and self-efficacy) and examined their relationship to frustration, presence, grit, and learning. Additional affective states (such as confusion) were not considered in this work.

Another limitation of the study includes the demographics of the participants. The participant pool was not only a young population and mostly freshmen, but the overwhelming majority had not yet been active members of the US Army – an important element to consider as the domain of this experiment consisted of combat medical care designed for active military personnel. Also, difference in class year by condition may have been a limitation as a possible confound in significance testing. In short, it is not clear how well these findings will generalize beyond the cadet population tested.

Also, this study encountered difficulties in launching many of the tasks of the experiment, i.e., the transition from the PowerPoint to vMedic, and then from vMedic to surveys often required manual prompts or an outright reboot of the entire experiment. As a result of these technical difficulties, there were disruptions for many participants in the experiment. Sometimes these difficulties resulted in complete data loss, but mostly these difficulties delayed the transitions between experimental procedures for the participant. This may have influenced participants' degree of frustration.

Other limitations of this work include the fact that the feedback messages were brief and only occurred one time per each scenario. It is possible that longer feedback messages and/or repeated messages might have had a more significant impact on the learning outcomes of the participant. Lastly, a limitation of this work includes the lack of opportunity so study content retention over a greater period of time.

Future work

As this study was only conducted in one game-based learning domain (vMedic), conclusions from this study should be applied to other game-based learning domains to further verify the generalizability of these findings. Also, future studies could be designed so to stratify the assignment to conditions by pre-test scores.

Additionally, future studies should devise feedback messages based on other theories of motivation, perhaps capitalizing on the Hawthorne Effect (e.g., where the system explicitly states that the system as an entity is evaluating and judging the participant and will report back to another entity on the participant's success or failure); messages based on Expectancy Theory (e.g., messages that state that future scenarios will be less difficult if participants can master initial scenarios quickly and efficiently); messages based on Attribution Theory (e.g., messages that state the participant must put forth more effort to be successful).

Future studies should also test to see whether older, active members of the US Army would respond differently to the existing set of motivational messages – particularly messages derived from the social identity theory -- and explore the impact of additional motivational theory-based interventions on this older, active population.

Additionally, given the mostly homogenous nature of the population, future research should replicate this study on a more heterogeneous population with a greater diversity of skills, abilities, and traits such as grit. Given the interactive relationship between motivation, affect, and cognition, further studies are needed to examine what combinations of feedback designs and trait considerations would support cognitive performance across a greater diversity of populations.

Also, this study revealed that individuals that differed according to grit responded differently to motivational feedback messages in the control-value condition. It is suggested that future ITSs incorporate procedures to collect other trait measures as a way for tutoring systems to individually tailor and automate exactly what kind of motivational feedback messages would be an effective intervention.

Section 4. Conclusion

In conclusion, the results of this experiment support previous theories and empirical research that have recognized the need to identify and address affective states that lead to disengagement in learning (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Lehman, & Graesser, 2011; D'Mello Strain, Olney, & Graesser, 2013; Forbes-Riley, Litman, Friedberg, 2011; Gee, 2004, 2007; Picard et al., 2004). These results also contribute to the body of research that has given evidence that providing interventions in the form of feedback messages can positively effect the learning of domain content in ITSs (Wagster, Tan, Wu, Biswas, & Schwartz, 2007; Roll, Aleven, McLaren, & Koedinger; 2011).

Further, this work contributes to ongoing work on developing interactive, sensor-free affective detector models for intelligent tutoring systems. Specifically, this results of this study demonstrate how interaction based, sensor-free detectors embedded in technology-based learning environments (such as serious video games) can be used to trigger interventions that can lead to better learning outcomes. Perhaps most notable, this work was positioned within a broader research agenda led by Dr. Ryan Baker of Teachers College, to develop sensor-free affect detectors. As such, the findings of this dissertation study was used in a subseqent study led by Dr. Baker to test the functionality and effectiveness of these sensor-free affect detectors using the self-efficacy feedback messages developed in this dissertation study.

Future work can build from this work to develop more sensitive and timely interaction-based detectors and test other feedback interventions for more robust learning outcomes not only in medical combat care training, but other domains that require content and skill mastery.

Overall, self-efficacy based motivational feedback interventions were associated with better learning when addressing frustration. While this current study was limited to a military population, it is likely that this motivational approach would work outside of a military population. In particular, over 80% of this study's sample population had not having previously served in the military, rendering the identity of this group arguably closer to undergraduate population of a comparable higher education institution than a population of active military personnel.

While presence did not interact with conditions to yield a difference between learning outcomes by condition, there was an interactive effect of grit by condition, providing some marginal evidence that motivational messages were more effective for participants with comparatively low grit measures vs. high grit measures in the control-value condition.

These results contribute towards the body of cognitive performance theory and research by providing empirical evidence for effective approaches to address motivation in simulated learning environments, considerations on the mediating effect of frustration in relation to motivational feedback, as well as evidence regarding tailoring motivational feedback according to trait characteristics such as grit. Specifically, these results include evidence that the self-efficacy motivational feedback messages used to intercede in instances of high frustration can promote greater learning gains, and motivational messages based on the theory of control-value may be effective for low grit populations, but may have a negative impact on high grit populations.

In sum, the outcomes of this experiment have implications for both human to human instruction as well as for efforts in developing affect-sensitive ITSs. Specifically, the results of this study can be used to contribute to the ongoing effort of developing affect-sensitive feedback interventions to support learner engagement and promote learning gains in ITSs.

References

Abt, Clark C. *Serious Games*. Lanham, MD: University Press of America, 1987.

Adams, D.M., Mayer, R.E., MacNamara, A., Koenig, A., Wainess, R. (2012) Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104* (1), 235-249.

Allwood, J., Nivre, J., & Ahlsén, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, *9*(1), 1-26.

Amsel, A. (1992). Frustration theory: Many years later. *Psychological bulletin*, 112(3), 396.

Annetta, L.A., Minogue, J., Holmes, S.Y. Cheng, M.T. (2009) Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*, 74–85.

Arai, K., & Asmara, R. A. (2013). 3D Skeleton model derived from Kinect Depth Sensor Camera and its application to walking style quality evaluations. *International Journal of Advanced Research in Artificial Intelligence*, *2*(7), 24-28.

Aron, A. & Aron E. N. (1994) *Statistics for Psychology*. Englewood Cliffs, NJ: Prentice Hall Publishers.

Arroyo, I., Woolf, B.P., Cooper, D., Burleson, W., Muldner, K., Christopherson, R. (2009) Emotion Sensors Go To School. *Proceedings of the 14th International Conference on Artificial Intelligence In Education.*

Azevedo, R. (2002). Beyond intelligent tutoring systems: Using computers as METAcognitive tools to enhance learning?. Instructional Science, 30(1), 31-45..

Baddeley, A., (2012). How does emotion influence working memory? In S. Masmoudi, D. Y. David, A. Naceur, (Eds.). *Attention, Representation, and human performance:*

Integration of cognition, emotion, and motivation. New York: Psychology Press.

Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 1059-1068). ACM.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: when students game the system. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 383-390). ACM.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T., Graesser, A.C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. In *International Journal of Human-Computer Studies*, 68, 223-241.

Baker, R. S., DeFalco, J. A., Ocumpaugh, J., & Paquette, L. (2015). Toward Detection of Engagement and Affect in a Simulation-based Combat Medic Training Environment. In Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2) (p. 115).

Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., Rossi, L. (2012). Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In *Proceedings of the 5th International Conference on Educational Data Mining*, 126-133.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in" gaming the system" behavior in interactive learning environments. Journal of Interactive Learning Research, 19(2), 185.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.

Barrett, L. F., & Lindquist, K. A. (2008). The embodiment of emotion. In Semin, G., & Smith, E. R. (Eds.). *Embodied grounding: Social, cognitive, affective, and neuroscience approaches* (pp. 237-262). Cambridge: Cambridge University Press.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*. Series B (Methodological), 289-300.

Blair, B. L., Perry, N. B., O'Brien, M., Calkins, S. D., Keane, S. P., & Shanahan, L. (2015). Identifying developmental cascades among differentiated dimensions of social competence and emotion regulation. *Developmental psychology*, 51(8), 1062.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., ... & Zhao, W. (2015, March). Automatic detection of learning-centered affective states in the wild. In Proceedings of the 20th international conference on intelligent user interfaces (pp. 379-388). ACM.

Braver, T. (2015). (Ed.) *Motivation and Cognitive Control*. London, GB: Routledge.

Brody, L. R. & Hall, J. A. (2008). Gender and emotion in context. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), Handbook of emotions (3rd ed., pp. 395-408). New York: Guilford. Buck, R. (1985). Prime theory: An integrated view of motivation and emotion. *Psychol Rev*, 92 (3), 389.

Burleson, W. (2006). Affective learning companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance (Doctoral dissertation, Massachusetts Institute of Technology).

Burleson, W., Picard, R.W. (2004) Affective agents: Sustaining motivation to learn through failure and a state of stuck. *Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments.*

Buck, R. (1985). Prime theory: An integrated view of motivation and emotion. Psychological review, 92(3), 389.

Butler Jr., F. K., & Blackbourne, L. H. (2012). Battlefield trauma care then and now: a decade of Tactical Combat Casualty Care. *Journal of Trauma and Acute Care Surgery*, *73*(6), S395-S402.

Butz, C. J., Hua, S., & Maguire, R. B. (2004, September). A web-based intelligent tutoring system for computer programming. In Web Intelligence, 2004. WI 2004. Proceedings. IEEE/WIC/ACM International Conference on (pp. 159-165). IEEE.

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.

Carver, C. S. (2006). Approach, avoidance, and the self-regulation of affect and action. *Motivation and Emotion*, 30 (2), 105– 110.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychol Rev*, 97 , 19– 35. Chiew, Chapman, P. L., Cabrera, L. D., Varela-Mayer, C., Baker, M. M., Elnitsky, C., Figley, C., ... & Mayer, L. P. (2012). Training, deployment preparation, and combat experiences of deployed health care personnel: key findings from deployed US Army combat medics assigned to line units. Military medicine, 177(3), 270-277.

Ciompi, L., & Panksepp, J. (2004). Energetic effects of emotions on cognitions: Complementary psychobiological and psychosocial finding. In R. Ellis & N. Newton (Eds.), *Consciousness and emotions* (Vol. 1, pp. 23– 55). Amsterdam: Benjamins.

Cocea, M., Hershkovitz, A., & Baker, R. S. J. d. (2009) The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? Proceedings of the 14th International Conference on Artificial Intelligence in Education, 507-514.

Conati, C., Maclaren, H. (2005) Data-Driven Refinement of a Probabilistic Model of User Affect, *Proceedings of User Modeling 2005,* 40-49.

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

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. *Handbook of human-computer interaction*, 849-874.

Craig, S., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive-affective states during learning. *Cognition & Emotion*, 22(5), 777-788.

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

D'Mello, S. K., Craig, S. D., & Graesser, A. C. (2009). Multimethod assessment of affective experience and expression during deep learning. *International Journal of Learning Technology*, 4(3-4), 165-187.

D'Mello, S., Craig, S., Fike, K., & Graesser, A. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In Human-computer interaction. Ambient, ubiquitous and intelligent interaction (pp. 595-604). Springer Berlin Heidelberg.

D'Mello, S.K., Craig, S.D., Gholson, B., Franklin, S., Picard, R., Graesser, A.C., 2005. Integrating affect sensors in an intelligent tutoring system. In: Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces. ACM Press, New York, pp. 7–13.

D'Mello, S.K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., Graesser, A. C. (2008) Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User Adapted Interaction, 18* (1-2), 45-80.

D'Mello, S. K., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20(2), 147-187.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, 25(7), 1299-1308.

D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., ... & Graesser, A. (2008). AutoTutor detects and responds to learners affective and cognitive states. In Workshop on emotional and cognitive issues at the international conference on intelligent tutoring systems (pp. 306-308).

D'Mello, S. K., Lehman, B., & Graesser, A. (2011). A motivationally supportive affect-sensitive auto-tutor. In New perspectives on affect and learning technologies (pp. 113-126). Springer New York.

D'Mello, S., Picard, R. W., & Graesser, A. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, (4), 53-61.

D'Mello, S. K., Strain, A. C., Olney, A., & Graesser, A. (2013). Affect, meta-affect, and affect regulation during complex learning. In International handbook of metacognition and learning technologies (pp. 669-681). Springer New York.

Dalgleish, T., & Power, M. J. (Eds.). (1999). Handbook of cognition and emotion. Chichester, UK: Wiley.

De Lorenzo, R. A. (2001). Medic for the millennium: the US Army 91W health care specialist. *Military medicine*, 166(8), 685.de Wit, S., & Dickinson, A. (2015). In Braver, T. (2015).

(Ed.), *Motivation and Cognitive*

Control (pp. 143-163). London, GB: Routledge.

Deci, E. L., Eghrari, H., Patrick, B.C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62,* 119-142.

Deci, E.L., Ryan, R.M., 1985. The general causality orientations scale: self- determination in personality. Journal of Research in Personality 19, 109–134.

Deci, E.L., Ryan, R.M., 2000. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry* 11, 227–268.

DePillis, L. (2013). The return of the tourniquet: What we learned from war led to lives saved in Boston. *New Republic*, April 17, 2013. Retrieved on February 8, 2015 at http://www.newrepublic.com/article/112939/boston-marathon-bombing-return-tourniquet.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). Journal of personality assessment, 91(2), 166-174.

Dweck, C. S. (2002). Messages that motivate: How praise molds students' beliefs, motivation, and performance (in surprising ways).

Engadget. (2010) *Affectiva's Q sensor wristband monitors and logs stress levels, might bring back the snap bracelet*. New York: Newstex. Retrieved from http://ezproxy.cul.columbia.edu/login?url=http://search.proquest.com.ezproxy.cul.columbia.edu/docview/761419847?accountid=10226

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

Fernandez, R., Picard, R.W., 1998. Signal processing for recognition of human frustration. In: *International Conference on Acoustics, Speech, and Signal Processing*.

Field, A. (2005).Discovering Statistics Using SPSS (2nd edition). London: Sage.

Forbes-Riley, K., Litman, D., & Friedberg, H. (2011). Annotating disengagement for spoken dialogue computer tutoring. In New Perspectives on Affect and Learning Technologies (pp. 169-181). Springer New York.

Forgas, J.P. (2000). Managing moods: Toward a dual-process theory of spontaneous mood regulation. In *Psychological Inquiry*, 11, 172-177

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of counseling psychology*, 51(1), 115.

Frize, M., & Frasson, C. (2000). Decision-support and intelligent tutoring systems in medical education. *Clinical and investigative medicine*, 23(4), 266-269.

Gable, P., & Harmon-Jones, E. (2010). The motivational dimensional model of affect: Implications for breadth of attention, memory, and cognitive categorisation. Cognition and Emotion, 24(2), 322-337.

Goldberg, B. S. (2013). *Explicit Feedback Within Game-based Training: Examining the Influence of Source Modality Effects on Interaction* (Doctoral dissertation, University of Central Florida Orlando, Florida).

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., Billings, D. R., & Malone, N. (2012). Use of evidence-based strategies to enhance the extensibility of adaptive tutoring technologies. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (No. 1).

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.

Gollwitzer, P. M. (2012). Mindset theory of action phases. In P. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), Handbook of theories of social psychology (Vol. 1, pp. 526–545). London, UK: SAGE.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems.

Graesser, A., & Olde, B. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, 95(3), 524-536.

Gross, J. J. (2008). Emotion Regulation. In M. Lewis, J. M., Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotion* (pp. 497-512). New York: Guilford Press

Habgood, J.M.P. Ainsworth, S.E. (2011) Motivating children to learn effectively: Exploring the 206.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: controversies and consensus. *Trends in cognitive sciences*, 16(9), 458-466.

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. *Current Directions in Psychological Science*, 22(4), 301-307.

Haro, A., Essa, I., Flickner, M., 2000. Detecting and tracking eyes by using their physiological properties. In: *Proceedings of Conference on Computer Vision and Pattern Recognition*, June 2000.

Heilman, M., Collins-Thompson, K., Callan, J., & Eskenazi, M. (2006). Classroom success of an intelligent tutoring system for lexical practice and reading comprehension. In *INTERSPEECH.*

Heffernan, N. T., & Koedinger, K. R. (2002, June). An intelligent tutoring system incorporating a model of an experienced human tutor. In *Intelligent Tutoring Systems* (pp. 596-608). Springer Berlin Heidelberg.

Hershkovitz, A., Wixon, M., Baker, R. S. J. d., Gobert, J., & Sao Pedro, M. (2011) Carelessness and Goal Orientation in a Science Microworld. Poster paper.

Proceedings of the 15th International Conference on Artificial Intelligence in Education, 462-465.

Hickey, D. T., Ingram-Goble, A. A., Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology, 18*, 187–208.

Hinshaw, S. P. (2007). Moderators and mediators of treatment outcome for youth with ADHD: Understanding for whom and how interventions work. *Ambulatory Pediatrics*, 7(1), 91-100.

Holden, H., Sottilare, R., Goldberg, B., & Brawner, K. (2012). Effective learner modeling for computer-based tutoring of cognitive and affective tasks. Proceedings from *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).*

Hollands, F., & Bakir, I. (2015). Efficiency of Automated Detectors of Learner Engagement and Affect Compared with Traditional Observation Methods. NY, NY: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.

Holodynski, M. & Friedlmeier, W. (2006). *Development of emotions and emotion regulation*. New York: Springer.

Honey, M., & Hilton, M. (2011) *Learning Science Through Computer Games and Simulations.* Washington, DC: National Research Council.

Hoyle, R. H., Harris, M. J., & Judd, C. M. (2002). Research methods in social relations. USA: Wadsworth Thompson Learning, Inc.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York: Appleton-Century.

Hume, D. (1888). *A treatise of human nature* (L. A. Selby-Bigge, Ed.). Oxford: Oxford University Press. (Original work published 1739)

Jaccard, J., Wan, C. K., & Turrisi, R. (1990). The detection and interpretation of interaction effects between continuous variables in multiple regression. Multivariate Behavioral Research, 25(4), 467-478.

Jacobs, L., McSwain, N., Rotondo, M., Fabbri, W., Eastman, A., Butler, F., & Sinclair, J. (2013). Improving survival from active shooter events: The Hartford Consensus. American College of Surgeons and Federal Bureau of Investigation: Joint Committee to Create a National Policy to Enhance Survivability from Mass Casualty Shooting Events. April 2, 2013. Retrieved on 8 February 2015 at http://www.scribd.com/doc/136325175/Improving-Survival-From-Active-Shooter-Events

James, W. (1884). What is an emotion? *Mind*, 9, 188– 205.

James, W. (1950). *Principles of psycholo*gy. New York: Dover. (Original work published 1890)

Johnson, D. M., & Stewart II, J. E. (1999). Use of virtual environments for the acquisition of spatial knowledge: Comparison among different visual displays. *Military Psychology*, 11(2), 129.

Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies*, 65(8), 724-736.

Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014). Grit and hardiness as predictors of performance among West Point cadets. *Military Psychology*, 26(4), 327.

Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. Review of General Psychology, 2(3), 320.

Kensinger, E. A., & Schacter, D. (2008). Memory and emotion. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 601-617). New York: Guilford.

Keselman, H. J. (1998). Testing treatment effects in repeated measures designs: an update for psychophysiological researchers. Psychophysiology, 35(04), 470-478.

Ketelhut, D. J., Dede, C., Clarke, J., Nelson, B. (2006). A multi-user virtual environment for building higher order inquiry skills in science. Paper presented at the American Educational Research Association Conference, San Francisco, CA, USA

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city.

Koedinger, K. R., Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences.* Cambridge, MA: Cambridge University Press.

Kragh Jr., COL (ret), J. F., Dubick, M. A., Aden, J. K., McKeague, A. L., Rasmussen, T. E., Baer, D. G., & Blackbourne, L. H. (2015). US Military Use of Tourniquets from 2001 to 2010. Prehospital Emergency Care, 19(2), 184-190.

Kragh Jr., J. F., Walters, T. J., Westmoreland, T., Miller, R. M., Mabry, R. L., Kotwal, R. S., ... & Cain, J. S. (2013). Tragedy into drama: an American history of tourniquet use in the current war. *Army Inst Of Surgical Research.* Fort Sam Houston ,TX.

Jimura, K., Locke, H. S., & Braver, T. S. (2010). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. Proceedings of the National Academy of Sciences, 107(19), 8871-8876.

Johns, C., Nunez, D., Daya, M., Sellars, D., Casanueva, J., & Blake, E. (2000). The interaction between individuals' immersive tendencies and the sensation of presence in a virtual environment. In *Virtual Environments* 2000 (pp. 65-74). Springer Vienna.

Laming, D. R. (2000). On the behavioural interpretation of neurophysiological observation. Behavioral and Brain Sciences, 23(02), 209-209.

Larsen, J., Berntson, G., Poehlmann, K, Ito, T., & Cacioppo, J. (2008). The psychophysiology of emotion. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 180-195). New York: Guilford.

Lehman, B., Matthews, M., D'Mello, S., & Person, N. (2008, June). What are you feeling? Investigating student affective states during expert human tutoring sessions. In *Intelligent Tutoring Systems* (pp. 50-59). Springer Berlin Heidelberg.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. *Computers as cognitive tools*, 1993, 75-105.

Lester, J., McQuiggan, S., Sabourin, J. (2011) Affect Recognition and Expression in Narrative-Centered Learning Environments. In Calvo, R., D'Mello, S. (Eds.) *New Perspectives on Affect and Learning Technologies.* New York, NY: Springer.

Linnenbrink, E.A., & Pintrich, P.R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*. C.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. (2013). Sequences of frustration and confusion, and learning. In Educational Data Mining 2013.

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, and Behavioral Neuroscience*, 8 , 99– 112.

Locke, H. S., & Braver, T. S. (2010). Motivational influences on cognitive control: A cognitive neuroscience perspective. In R. R. Hassin, K. N. Ochsner, & Y. Trope (Eds.), *Self control in society, mind, and brain* (pp. 114– 140). New York: Oxford University Press.

Maddox, W. T., & Markman, A. B. (2010). The motivation-cognition interface in learning and decision-making. *Current Directions Psychological Science*, 19 (2), 106– 110. D oi:10.1177/0963721410 364008.

Mandler, G. (1984). *Mind and body: Psychology of emotion and stress*. WW Norton.

Mandler, G. (1999). Emotion. In B. M. Bly & D. E. Rumelhart (Eds.), *Cognitive science. Handbook of perception and cognition* (2nd ed.). San Diego, CA: Academic Press.

Marien, H., Aarts, H., & Custer, R. (2015). How goals control behavior: The role of action-outcome and reward information. In Braver, T. (2015). (Ed.), *Motivation and Cognitive Control* (pp. 165-185). London, GB: Routledge.

Matsumoto, D., Keltner, D., Shiota, M., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. . In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds*.), Handbook of emotions* (3rd ed., pp. 211-234). New York: Guilford.

Matsumoto, D., Yoo, S. H., Jakagawa, S., & 37 members of the Multinational Study of Cultural Display Rules (2008). Culture, emotion regulation, and adjustment. *Attitudes and Social Cognition*, 94, 925-937.

Mayer, R. E., Johnson, C. I. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research, 42*(3), 241-265.

McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., & Graesser, A. (2007). Facial Features for Affective State Detection in Learning Environments. In D. McNamara & G.

Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 467-472). Austin, TX: Cognitive Science Society.

McLaughlin, M., Chen, Y., Park, N., Zhu, W., Yoon, H., 2004. Recognizing user state from haptic data. In: *International Conference on Information Systems Analysis and Synthesis*.

McQuiggan, S. W., Lee, S., and Lester, J. C. 2007. Early prediction of student frustration. In A. Paiva, R. Prada, and R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction,* 698-709.

McQuiggan, S.W, Lester, J.C. (2007) Modeling and Evaluating Empathy in Embodied Companion Agents. *International Journal of Human-Computer Studies, 65* (4), 348-360.

McQuiggan, S.W., B.W. Mott, Lester, J.C. (2008) Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction, 18*, 81–123.

McQuiggan, S.W., Robison, J.L., Lester, J.C. (2010) Affective transitions in narrative-centered learning environments. *Educational Technology & Society, 13* (1), 40-53.

Meyer, D., & Turner, J. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18(4), 377-390.

Michael, D., and Chen, S. *Serious Games : Games That Educate, Train, and Inform. Boston*, US: Course Technology / Cengage Learning, 2005.

Milanak, M. E., & Berenbaum, H. (2014). The effects of context on facial affect recognition. *Motivation and Emotion*, 38(4), 560-568.

Mota, S., Picard, R.W., 2003. Automated posture analysis for detecting learner's interest level. In: *Workshop on Computer Vision and Pattern Recognition for Human–Computer Interaction*, June 2003.

Murray, H.A. (1938). *Explorations in personality*. New York: Oxford University Press.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In Authoring tools for advanced technology learning environments (pp. 491-544). Springer Netherlands.

National Association of Emergency Medical Technicians. (2016) Retrieved on 14 March 2016 at http://www.naemt.org/education/TCCC/Acout_TCCC.aspx .

Niedenthal, P. M. . (2008). Emotion concepts. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 587-600). New York: Guilford.

Ocumpaugh, J. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual.* Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for Educational Data Mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487-501.

Ocumpaugh, J., Baker, R.S.J.d., & Rodrigo, M.M.T. (2012). Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0. Technical Report. EdLab: New York, NY; Manila, Philippines: Ateneo Laboratory for the Learning Sciences

Panksepp, J. (1998a). *Affective neuroscience: The foundations of human and animal emotions*. New York: Oxford University Press.

Panksepp, J. (1998b). The periconscious substrates of consciousness: Affective states and the evolutionary origins of the SELF. *Journal of Consciousness Studies*, 5, 566– 582.

Panksepp, J. (2003). At the interface of affective, behavioral and cognitive neurosciences: Decoding the emotional feelings of the brain. *Brain and Cognition*, 52, 4– 14.

Panksepp, J. (2008). The affective brain and core consciousness: How does neural activity generate emotional feelings? In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 47-67). New York: Guilford.

Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014, June). Sensor-free affect detection for a simulation-based science inquiry learning environment. In Intelligent tutoring systems (pp. 1-10). Springer International Publishing.

Pardos, Z. A., Baker, R. S., San Pedro, M., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: investigating how affect and engagement during the school year predict end-of-year learning outcomes. test, 1(1), 107-128.

Pekrun, R. (1992). *Educational environments and intrinsic/extrinsic academic motivation: Expectancy value theory and longitudinal evidence.* In paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Pekrun, R., Elliot, A., & Maier, M. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98(3), 583-597.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational psychologist, 37(2), 91-105.

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends Cogn Sci*, 13 (4), 160– 166.

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Front Neurosci*, 4 , 17. doi:10.3389/fnins.2010.00017

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., ... & Strohecker, C. (2004). Affective learning—a manifesto. BT technology journal, 22(4), 253-269.

Plant, R.W., & Ryan, R.M. 1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. *Journal of Personality, 52*, 435-449.

Prendiner, H., Ishizuka, M. (2005) The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence, 19* (3-4), 267-285.

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, 36(8), 879-896

Qi, Y., Reynolds, C., Picard, R.W., 2001. The Bayes point machine for computer user frustration detection via pressure mouse. In: *Proceedings From the Workshop on Perceptive User Interface*.

Raybourn, E.M. (2007) Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with Computers, 19* (2), 206-214.

Rebolledo-Mendez, G., Du Boulay, B., & Luckin, R. (2006). Motivating the learner: an empirical evaluation. In *International Conference on Intelligent Tutoring Systems* (pp. 545-554). Springer Berlin Heidelberg.

Robison, J., McQuiggan, S., & Lester, J. (2009, September). Evaluating the consequences of affective feedback in intelligent tutoring systems. In *Affective Computing and Intelligent Interaction and Workshops*, 2009. ACII 2009. 3rd International Conference on (pp. 1-6). IEEE.

Rodrigo, M. M. T., & Baker, R. S. (2009). Coarse-grained detection of student frustration in an introductory programming course. In *Proceedings of the fifth international workshop on*

Computing education research workshop (pp. 75-80). ACM.

Rodrigo, M.M.T., Baker, R.S.J.d. (2011) Comparing Learners' Affect While Using an Intelligent Tutor and an Educational Game. *Research and Practice in Technology Enhanced Learning*, 6 (1), 43-66.

Rodrigo, M.M.T., Baker, R.S.J.d., D'Mello, S., Gonzalez, M.C.T., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sugay, J.O., Tep, S., Viehland, N.J.B. (2008a) Comparing Learners' Affect While Using an Intelligent Tutoring Systems and a Simulation Problem Solving Game. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 40-49.

Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B. (2007) Affect and Usage Choices in Simulation Problem Solving Environments. *Proceedings of Artificial Intelligence in Education 2007*, 145-152.

Rodrigo, M.M.T., Rebolledo-Mendez, G., Baker, R.S.J.d., du Boulay, B., Sugay, J.O., Lim, S.A.L., Espejo-Lahoz, M.B., Luckin, R. (2008b) The Effects of Motivational Modeling on Affect in an Intelligent Tutoring System. *Proceedings of International Conference on Computers in Education*, 57-64.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. Learning and Instruction, 21(2), 267-280.

Rowe, J., Shores, L., Mott, B., Lester, J. (2011) Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education, 21*(1-2), 115-133.

Rozin, P., & Cohen, A. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion*, 3, 68 75.

Ryan, R.M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*, 450-461.

Ryan, R.M., Connell, J.P., & Plant, R.W. (1990). Emotions in non-directed text learning. *Learning and Individual Differences, 2,* 1-17.

Ryan, R.M., Koestner, R., & Deci, E.L. (1991). Varied forms of persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotions*, *15*, 185-205.

Ryan, R.M., Mims, V., & Koestner, R., (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology, 45,* 736-750.

Ryle, G. (1951). *The concept of mind*. New York: Barnes & Noble.

Sabourin, J., Mott, B., Lester, J. (2011) Modeling Learner Affect with Theoretically Grounded Dynamic Bayesian Networks. *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction* (ACII-2011).

Sabourin, J., Rowe, J., Mott, B., and Lester, J. (2011). When Off-Task in On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. Proceedings of the 15th International Conference on Artificial Intelligence in Education, pp. 534-536.

Sansone & J.M. Harackiewicz (Eds.). San Diego, CA: Academic Press. (2000), 195-227.

Sartre, J.P. (1948). *The emotions: Sketch of a theory* (B. Frechtman, Trans.). New York: Philosophical Library. (Original work published 1939)

Sartre, J.P. (1956). *Being and nothingness* (H. Barnes, Trans.). New York: Washington Square Press. (Original work published 1943)

Schunk, DH & Zimmerman BJ. (2007). Influencing children's self-efficacy of reading and self-regulation of reading and writing through modeling. In *Reading and Writing Quarterly, 23,* 7-25.

Schwarz, N. (2007). Attitude construction: Evaluation in context. Social Cognition, 25, 638−656.

Schwarz, N. (2009). Mental construal in social judgment. In F. Strack, & J. Förster (Eds.), Social cognition: The basis of human interaction (pp. 121−138). New York: Psychology Press.

Shaffer J.P. (1995) Multiple hypothesis testing, Annual Review of Psychology 46:561-584, Annual Reviews.

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends Cogn Sci* , 14 (10), 464– 472.

Shuker, S. (2011). The immediate lifesaving management of maxillofacial, life-threatening hemorrhages due to IED and/or shrapnel injuries: "When hazard is in hesitation, not in the action." *Journal of Cranio-Maxillo-Facial Surgery*, 40: 534-540.

Slack, G. O. (2014). The intersectionality between the Short Grit Scale and four measures of Academic Self-Efficacy of African American males enrolled through a remedial program at a Historically Black College and University (Doctoral dissertation, PRAIRIE VIEW A&M UNIVERSITY).

Small, S. D., Wuerz, R. C., Simon, R., Shapiro, N., Conn, A., & Setnik, G. (1999). Demonstration of high-fidelity simulation team training for emergency medicine. *Academic Emergency Medicine*, 6(4), 312-323.

Smith, E., & Semin, G. (2004). Socially situated cognition: Cognition in its social context.

Advances in Experimental Social Psychology, 36, 53−117.

Smith, E., & Semin, G. (2007). Situated social cognition. Current Directions in Psychological Science, 16, 132−135.

Solomon, R. C. (2008). The philosophy of emotions. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 3-16). New York: Guilford.

Sotomayor, T. M. (2010). Teaching tactical combat casualty care using the TC3 sim game based simulation: a study to measure training effectiveness. *Studies in health technology and informatics*, *154*, 176-179.

Sottilare, R. A. (2005). Improving soldier learning and performance through simulation and training technologies. *Army AL&T Magazine*, *3*, 31-35.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Concept paper released as part of GIFT software documentation. Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED). Retrieved from: https://gifttutoring.org/attachments/152/GIFTDescription_0.pdf

Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012, December). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the *Interservice/Industry Training, Simulation, and Education Conference* (pp. 1-13).

Stein, N., & Levine, L. (1991). Making sense out of emotion. In A. O. W. Kessen, & F, Kraik (Eds.) (Ed.), *Memories, thoughts, and emotions: Essays in honor of George Mandler* (pp. 295-322). Hillsdale, NJ: Erlbaum.

Strayhorn, T. L. (2014). What role does grit play in the academic success of black male collegians at predominantly White institutions? *Journal of African American Studie*s, 18(1), 1-10.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. The social psychology of intergroup relations, 33(47), 74.

Thissen D., Steinberg L., & Kuang D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behavioral Stat., 27(1):77–83.

Tooby, J., Cosmides, L., Sell, A., Lieberman, D., & Sznycer, D. (2008). 15 Internal Regulatory Variables and the Design of Human Motivation: A Computational and Evolutionary Approach. In Handbook of approach and avoidance motivation (Vol. 251). Lawrence Erlbaum Mahwah, NJ.

TRADOC. (2011). The U.S. Army Learning Concept for 2015. TRADOC.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5), 497-503.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16,* 227–265.

VanLehn, K., Jordan, P. W., & Litman, D. J. (2007, October). Developing pedagogically effective tutorial dialogue tactics: experiments and a test-bed. In SLaTE (pp. 17-20).

Verma, J. P. (2016). *Repeated Measures Design for Empirical Researchers*. Hoboken, NJ: John Wiley & Sons.

Wagster, J., Tan, J., Wu, Y., Biswas, G., & Schwartz, D. (2007). Do learning by teaching environments with metacognitive support help students develop better learning behaviors.

In The twenty-ninth Annual Meeting of the Cognitive Science Society, Nashville, Tennessee (pp. 695-700).

Weinberg, S. L., & Abramowitz, S. K. (2002). Data Analysis for the Behavioral Sciences Using SPSS. New York: Cambridge University Press.

Wiener, N. (1948/1961). *Cybernetics; or Control and Communication in the Animal and in the Machine.* Cambridge, Mass.: MIT Press.

Witmer, B. G., & Singer, M. J. (1994). Measuring immersion in virtual environments. Tech. Rep. 1014, US Army Research Institute for the Behavioral and Social Sciences.

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

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognizing and responding to student affect*.* *International Journal of Learning Technology*, 4(3-4), 129-164.

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. In J. De Houwer, & D. Hermans (Eds.). *Cognition and emotion* pp. 211-275). New York: Psychology Press.

Zeidner, M. (1998) *Test Anxiety: The State of the Art.* Berlin, Germany: Springer.

APPENDIX A: SHORT GRIT SCALE

Short Grit Scale

Directions for taking the Grit Scale: Please respond to the following 8 items. Be honest – there are no right or wrong answers!

1. New ideas and projects sometimes distract me from previous ones.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

2. Setbacks don't discourage me.
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

3. I have been obsessed with a certain idea or project for a short time but later lost interest.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

4. I am a hard worker.
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

5. I often set a goal but later choose to pursue a different one.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

6. I have difficulty maintaining my focus on projects that take more than a few months to complete.*
   - ❑ Very much like me
   - ❑ Mostly like me
   - ❑ Somewhat like me
   - ❑ Not much like me
   - ❑ Not like me at all

7. I finish whatever I begin.
- ❑ Very much like me
- ❑ Mostly like me
- ❑ Somewhat like me
- ❑ Not much like me
- ❑ Not like me at all

8. I am diligent.
- ❑ Very much like me
- ❑ Mostly like me
- ❑ Somewhat like me
- ❑ Not much like me
- ❑ Not like me at all

Scoring:

1. For questions 2, 4, 7 and 8 assign the following points:
   5 = Very much like me
   4 = Mostly like me
   3 = Somewhat like me
   2 = Not much like me
   1 = Not like me at all

2. For questions 1, 3, 5 and 6 assign the following points:
   1 = Very much like me
   2 = Mostly like me
   3 = Somewhat like me
   4 = Not much like me
   5 = Not like me at all

Add up all the points and divide by 8. The maximum score on this scale is 5 (extremely gritty), and the lowest score on this scale is 1 (not at all gritty).

Grit Scale citation

Duckworth, A.L, & Quinn, P.D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, *91*, 166-174. http://www.sas.upenn.edu/~duckwort/images/Duckworth%20and%20Quinn.pdf

Duckworth, A.L., Peterson, C., Matthews, M.D., & Kelly, D.R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *9*, 1087-1101. http://www.sas.upenn.edu/~duckwort/images/Grit%20JPSP.pdf

APPENDIX B: SELF-EFFICACY SCALE

The General Self-Efficacy Scale (GSE)

<table>
<tr><td>Authors</td><td>Ralf Schwarzer & Matthias Jerusalem</td></tr>
<tr><td>Citation</td><td>Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user’s portfolio. Causal and control beliefs (pp. 35-37). Windsor, UK: NFER-NELSON.</td></tr>
<tr><td>Purpose</td><td>The scale was created to assess a general sense of perceived self-efficacy with the aim in mind to predict coping with daily hassles as well as adaptation after experiencing all kinds of stressful life events.</td></tr>
<tr><td>Population</td><td>The scale is designed for the general adult population, including adolescents. Persons below the age of 12 should not be tested.</td></tr>
<tr><td>Administration</td><td>The scale is usually self-administered, as part of a more comprehensive questionnaire. Preferably, the 10 items are mixed at random into a larger pool of items that have the same response format. Time: It requires 4 minutes on average. Scoring: Responses are made on a 4-point scale. Sum up the responses to all 10 items to yield the final composite score with a range from 10 to 40. No recoding or reverse coded items.</td></tr>
<tr><td>Description</td><td>The construct of Perceived Self-Efficacy reflects an optimistic self-belief (Schwarzer, 1992). This is the belief that one can perform a novel or difficult tasks, or cope with adversity -- in various domains of human functioning. Perceived self-efficacy facilitates goal-setting, effort investment, persistence in face of barriers and recovery from setbacks. It can be regarded as a positive resistance resource factor. Ten items are designed to tap this construct. Each item refers to successful coping and implies an internal-stable attribution of success.</td></tr>
<tr><td>Reliability</td><td>Cronbach’s Alpha: In samples from 23 nations, Cronbach’s alphas ranged from .76 to .90, with the majority in the high .80s. Factors: One Factor-The scale is unidimensional.</td></tr>
<tr><td>Validity</td><td>Criterion-related validity is documented in numerous correlation studies where positive coefficients were found with favorable emotions, dispositional optimism, and work satisfaction. Negative coefficients were found with depression, anxiety, stress, burnout, and health complaints. In studies with cardiac patients, their recovery over a half-year time period could be predicted by pre-surgery self-efficacy.
<table>
<tr><td colspan="2">Correlations between General Self-Efficacy Scale and Outcomes (correlations derived from a sample of n=180 university students; all correlations are significant, $p<.05$.</td></tr>
<tr><td>Extraversion</td><td>.49</td></tr>
<tr><td>Neuroticism</td><td>-.42</td></tr>
</table>
</td></tr>
</table>

	Action orientation	.43
	Hope for success	.46
	Fear of failure	-.45

Strengths	The measure has been used internationally with success for two decades. It is suitable for a broad range of applications. It can be taken to predict adaptation after life changes, but it is also suitable as an indicator of quality of life at any point in time.
Weaknesses	As a general measure, it does not tap specific behavior change. Therefore, in most applications it is necessary to add a few items to cover the particular content of the survey or intervention (such as smoking cessation self-efficacy, or physical exercise self-efficacy).
Bibliography (by year)	Jerusalem, M., & Schwarzer, R. (1992). Self-efficacy as a resource factor in stress appraisal processes. In R. Schwarzer (Ed.), *Self-efficacy: Thought control of action* (pp. 195-213). Washington, DC: Hemisphere. Rimm, H., & Jerusalem, M. (1999). Adaptation and validation of an Estonian version of the General Self-Efficacy Scale (ESES). *Anxiety, Stress, and Coping, 12,* 329-345. Schwarzer, R., & Scholz, U. (2000). Cross-Cultural Assessment of Coping Resources: The General Perceived Self-Efficacy Scale. Paper presented at the First Asian Congress of Health Psychology: Health Psychology and Culture, Tokyo, Japan.
Measure	

1	I can always manage to solve difficult problems if I try hard enough.
2	If someone opposes me, I can find the means and ways to get what I want.
3	It is easy for me to stick to my aims and accomplish my goals.
4	I am confident that I could deal efficiently with unexpected events.
5	Thanks to my resourcefulness, I know how to handle unforeseen situations.
6	I can solve most problems if I invest the necessary effort.
7	I can remain calm when facing difficulties because I can rely on my coping abilities.
8	When I am confronted with a problem, I can usually find several solutions.
9	If I am in trouble, I can usually think of a solution.
10	I can usually handle whatever comes my way.

Response Format	1 = Not at all true 2 = Hardly true 3 = Moderately true 4 = Exactly true

APPENDIX C: DEMOGRAPHICS SURVEY 2013

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 D: PRE and POST TEST SEPTEMBER 2013

NCSU TC Content Test

Content Test

For each of the following questions, choose the best answer.

1. You are treating a casualty while under fire. Which of the following can you perform before moving the casualty to a place of safety?

- ○ Perform cardiopulmonary resuscitation (CPR)
- ○ Apply a tourniquet to control severe bleeding on a limb
- ○ Perform needle chest decompression to relieve tension pneumothorax
- ○ Administer the combat pill pack to control pain and infection
- ○ All answers are incorrect

2. What medical term means bleeding, usually severe?

- ○ tourniquet
- ○ artery
- ○ hemorrhage
- ○ vein

3. How does evaluation and treatment of a casualty, in a care under fire situation, differ from a secure (tactical field care) situation?

- ○ While under fire, you only treat life-threatening bleeding from a limb
- ○ While under fire, you can focus more on the evaluation and treatment of the casualty
- ○ In a secure environment, you only treat life-threatening bleeding from a limb

4. 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?

- ○ Seek cover, return fire, play dead
- ○ Seek cover, return fire, administer self-aid
- ○ Play dead
- ○ Seek cover, return fire, administer buddy-aid

5. 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?

- ○ Nothing. Leave the tourniquet in place
- ○ Examine the wound to see if it is bleeding and can be controlled using other means

APPENDIX E: PRESENCE SURVEY

Intrinsic Motivation Inventory (IMI)

Scale Description

The Intrinsic Motivation Inventory (IMI) is a multidimensional measurement device intended to assess participants' subjective experience related to a target activity in laboratory experiments. It has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Ryan, 1982; Ryan, Mims & Koestner, 1983; Plant & Ryan, 1985; Ryan, Connell, & Plant, 1990; Ryan, Koestner & Deci, 1991; Deci, Eghrari, Patrick, & Leone, 1994). The instrument assesses participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension, and perceived choice while performing a given activity, thus yielding six subscale scores. Recently, a seventh subscale has been added to tap the experiences of relatedness, although the validity of this subscale has yet to be established. The **interest/enjoyment subscale is considered the self-report measure of intrinsic motivation**; thus, although the overall questionnaire is called the Intrinsic Motivation Inventory, it is only the one subscale that assesses intrinsic motivation, per se. As a result, the interest/enjoyment subscale often has more items on it that do the other subscales. The perceived choice and perceived competence concepts are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation, and pressure/tension is theorized to be a negative predictor of intrinsic motivation. Effort is a separate variable that is relevant to some motivation questions, so is used it its relevant. The value/usefulness subscale is used in internalization studies (e.g., Deci et al, 1994), the idea being that people internalize and become self-regulating with respect to activities that they experience as useful or valuable for themselves. Finally, the relatedness subscale is used in studies having to do with interpersonal interactions, friendship formation, and so on.

The IMI consists of varied numbers of items from these subscales, all of which have been shown to be factor analytically coherent and stable across a variety of tasks, conditions, and settings. The general criteria for inclusion of items on subscales have been a factor loading of at least 0.6 on the appropriate subscale, and no cross loadings above 0.4. Typically, loadings substantially exceed these criteria. Nonetheless, we recommend that investigators perform their own factor analyses on new data sets. Past research suggests that order effects of item presentation appear to be negligible, and the inclusion or exclusion of specific subscales appears to have no impact on the others. Thus, it is rare that all items have been used in a particular experiment. Instead, experimenters have chosen the subscales that are relevant to the issues they are exploring.

The IMI items have often been modified slightly to fit specific activities. Thus, for example, an item such as "I tried very hard to do well at this activity" can be changed to "I tried very hard to do well on these puzzles" or "...in learning this material" without effecting its reliability or validity. As one can readily tell, there is nothing subtle about these items; they are quite face- valid. However, in part, because of their straightforward nature, caution is needed in interpretation. We have found, for example, that correlations between self-reports of effort or interest and behavioral indices of these dimensions are quite modest--often around 0.4. Like other self-report measures, there is always the need to appropriately interpret how and why

participants report as they do. Ego-involvements, self-presentation styles, reactance, and other psychological dynamics must be considered. For example, in a study by Ryan, Koestner, and Deci (1991), we found that when participants were ego involved, the engaged in pressured persistence during a free choice period and this behavior did not correlate with the self-reports of interest/enjoyment. In fact, we concluded that to be confident in one's assessment of intrinsic motivation, one needs to find that the free-choice behavior and the self-reports of interest/enjoyment are significantly correlated.

Another issue is that of redundancy. Items within the subscales overlap considerably, although randomizing their presentation makes this less salient to most participants. Nonetheless, shorter versions have been used and been found to be quite reliable. The incremental R for every item above 4 for any given factor is quite small. Still, it is very important to recognize that multiple item subscales consistently outperform single items for obvious reasons, and they have better external validity.

On The Scale page, there are five sections. First, the full 45 items that make up the 7 subscales are shown, along with information on constructing your own IMI and scoring it. Then, there are four specific versions of the IMI that have been used in past studies. This should give you a sense of the different ways it has been used. These have different numbers of items and different numbers of subscales, and they concern different activities. First, there is a standard, 22-item version that has been used in several studies, with four subscales: interest/enjoyment, perceived competence, perceived choice, and pressure/tension. Second, there is a short 9-item version concerned with the activity of reading some text material; it has three subscales: interest/enjoyment, perceived competence, and pressure/tension. Then, there is the 25-item version that was used in the internalization study, including the three subscales of value/usefulness, interest/enjoyment, and perceived choice. Finally, there is a 29-item version of the interpersonal relatedness questionnaire that has five subscales: relatedness, interest/enjoyment, perceived choice, pressure/tension, and effort.

Finally, McAuley, Duncan, and Tammen (1987) did a study to examine the validity of the IMI and found strong support for its validity.

References

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, *62*, 119-142.

McAuley, E., Duncan, T., & Tammen, V. V. (1987). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport, 60,* 48-58.

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self- consciousness, self-awareness, and ego-involvement: An investigation of internally-controlling styles. *Journal of Personality*, *53*, 435-449.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461.

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in non-directed text learning. *Learning and Individual Differences*, *2*, 1-17.

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Varied forms of persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, *15*, 185-205.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology*, *45*, 736-750.

PRESENCE QUESTIONNAIRE

The Scales

THE POST-EXPERIMENTAL INTRINSIC MOTIVATION INVENTORY (Below are listed all 45 items that can be used depending on which are needed.)

For each of the following statements, please indicate how true it is for you, using the following scale:

1234567 not at all somewhat very true true true

Interest/Enjoyment

I enjoyed doing this activity very much□This activity was fun to do.□I thought this was a boring activity. (R)□This activity did not hold my attention at all. (R)□I would describe this activity as very interesting.□I thought this activity was quite enjoyable.□While I was doing this activity, I was thinking about how much I enjoyed it.

Perceived Competence

I think I am pretty good at this activity.□I think I did pretty well at this activity, compared to other students. After working at this activity for awhile, I felt pretty competent.□I am satisfied with my performance at this task.□I was pretty skilled at this activity.□This was an activity that I couldn't do very well. (R)

Effort/Importance

I put a lot of effort into this.□I didn't try very hard to do well at this activity. (R) I tried very hard on this activity.□It was important to me to do well at this task.□I didn't put much energy into this. (R)

Pressure/Tension

I did not feel nervous at all while doing this. (R) I felt very tense while doing this activity.□I was very relaxed in doing these. (R)

I was anxious while working on this task. I felt pressured while doing these.

Perceived Choice

I believe I had some choice about doing this activity.□I felt like it was not my own choice to do this task. (R) I didn't really have a choice about doing this task. (R) I felt like I had to do this. (R)□I did this activity because I had no choice. (R)□I did this activity because I wanted to.□I did this activity because I had to. (R)

Value/Usefulness

I believe this activity could be of some value to me.□I think that doing this activity is useful for ______________________ I think this is important to do because it can ______________________ I would be willing to do this again because it has some value to me.□I think doing this activity could help me to ______________________ I believe doing this activity could be beneficial to me.□I think this is an important activity.

Relatedness

I felt really distant to this person. (R)□I really doubt that this person and I would ever be friends. (R)□I felt like I could really trust this person.□I'd like a chance to interact with this person more often.□I'd really prefer not to interact with this person in the future. (R)□I don't feel like I could really trust this person. (R)□It is likely that this person and I could become friends if we interacted a lot. I feel close to this person.

Constructing the IMI for your study. First, decide which of the variables (factors) you want to use, based on what theoretical questions you are addressing. Then, use the items from those factors, randomly ordered. If you use the value/usefulness items, you will need to complete the three items as appropriate. In other words, if you were studying whether the person believes an activity is useful for improving concentration, or becoming a better basketball player, or whatever, then fill in the blanks with that information. If you do not want to refer to a particular outcome, then just truncate the items with its being useful, helpful, or important.

Scoring information for the IMI. To score this instrument, you must first reverse score the items for which an (R) is shown after them. To do that, subtract the item response from 8, and use the resulting number as the item score. Then, calculate subscale scores by averaging across all of the items on that subscale. The subscale scores are then used in the analyses of relevant questions.

PRESENCE QUESTIONNAIRE
(Witmer & Singer, Vs. 4.0)

Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

WITH REGARD TO THE EXPERIENCED ENVIRONMENT

1. How much were you able to control events?

|______| |______| |______| |______| |______| |______| |______|
NOT AT ALL SOMEWHAT COMPLETELY

2. How responsive was the environment to actions that you initiated (or performed)?

|______| |______| |______| |______| |______| |______| |______|
NOT RESPONSIVE MODERATELY RESPONSIVE COMPLETELY RESPONSIVE

3. How natural did your interactions with the environment seem?

|______| |______| |______| |______| |______| |______| |______|
EXTREMELY ARTIFICIAL BORDERLINE COMPLETELY NATURAL

4. How much did the visual aspects of the environment involve you?

|______| |______| |______| |______| |______| |______| |______|
NOT AT ALL SOMEWHAT COMPLETELY

5. How much did the auditory aspects of the environment involve you?

|______| |______| |______| |______| |______| |______| |______|
NOT AT ALL SOMEWHAT COMPLETELY

6. How natural was the mechanism which controlled movement through the environment?

|______| |______| |______| |______| |______| |______| |______|
EXTREMELY ARTIFICIAL BORDERLINE COMPLETELY NATURAL

7. How compelling was your sense of objects moving through space?

|______| |______| |______| |______| |______| |______| |______|
NOT AT ALL MODERATELY COMPELLING VERY

COMPELLING

8. How much did your experiences in the virtual environment seem consistent with your real world experiences?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT CONSISTENT — MODERATELY CONSISTENT — VERY CONSISTENT

9. Were you able to anticipate what would happen next in response to the actions that you performed?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — SOMEWHAT — COMPLETELY

10. How completely were you able to actively survey or search the environment using vision?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — SOMEWHAT — COMPLETELY

11. How well could you identify sounds?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — SOMEWHAT — COMPLETELY

12. How well could you localize sounds?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — SOMEWHAT — COMPLETELY

13. How well could you actively survey or search the virtual environment using touch?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — SOMEWHAT — COMPLETELY

14. How compelling was your sense of moving around inside the virtual environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT COMPELLING — MODERATELY COMPELLING — VERY COMPELLING

15. How closely were you able to examine objects?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL — PRETTY CLOSELY — VERY CLOSELY

16. How well could you examine objects from multiple viewpoints?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL SOMEWHAT EXTENSIVELY

17. How well could you move or manipulate objects in the virtual environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL SOMEWHAT EXTENSIVELY

18. How involved were you in the virtual environment experience?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT INVOLVED MILDLY INVOLVED COMPLETELY ENGROSSED

19. How much delay did you experience between your actions and expected outcomes?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NO DELAYS MODERATE DELAYS LONG DELAYS

20. How quickly did you adjust to the virtual environment experience?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL SLOWLY LESS THAN ONE MINUTE

21. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT PROFICIENT REASONABLY PROFICIENT VERY PROFICIENT

22. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL INTERFERED SOMEWHAT PREVENTED TASK PERFORMANCE

23. How much did the control devices interfere with the performance of assigned tasks or with other activities?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|

NOT AT ALL | INTERFERED SOMEWHAT | INTERFERED GREATLY

24. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL | SOMEWHAT | COMPLETELY

25. How completely were your senses engaged in this experience?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT ENGAGED | MILDLY ENGAGED | COMPLETELY ENGAGED

26. How easy was it to identify objects through physical interaction; like touching an object, walking over a surface, or bumping into a wall or object?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
IMPOSSIBLE | MODERATELY DIFFICULT | VERY EASY

27. Were there moments during the virtual environment experience when you felt completely focused on the task or environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NONE | OCCASIONALLY | FREQUENTLY

28. How easily did you adjust to the control devices used to interact with the virtual environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
DIFFICULT | MODERATE | EASILY

29. Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT CONSISTENT | SOMEWHAT CONSISTENT | VERY CONSISTENT

30. To what extent did you feel completely surrounded by and enveloped by the virtual environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL | SOME EXTENT | VERY MUCH

31. As you moved through the virtual environment and interacted with it, did you feel like you were inside the virtual environment, affecting or being affected by objects and events in that environment.

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL SOMEWHAT COMPLETELY

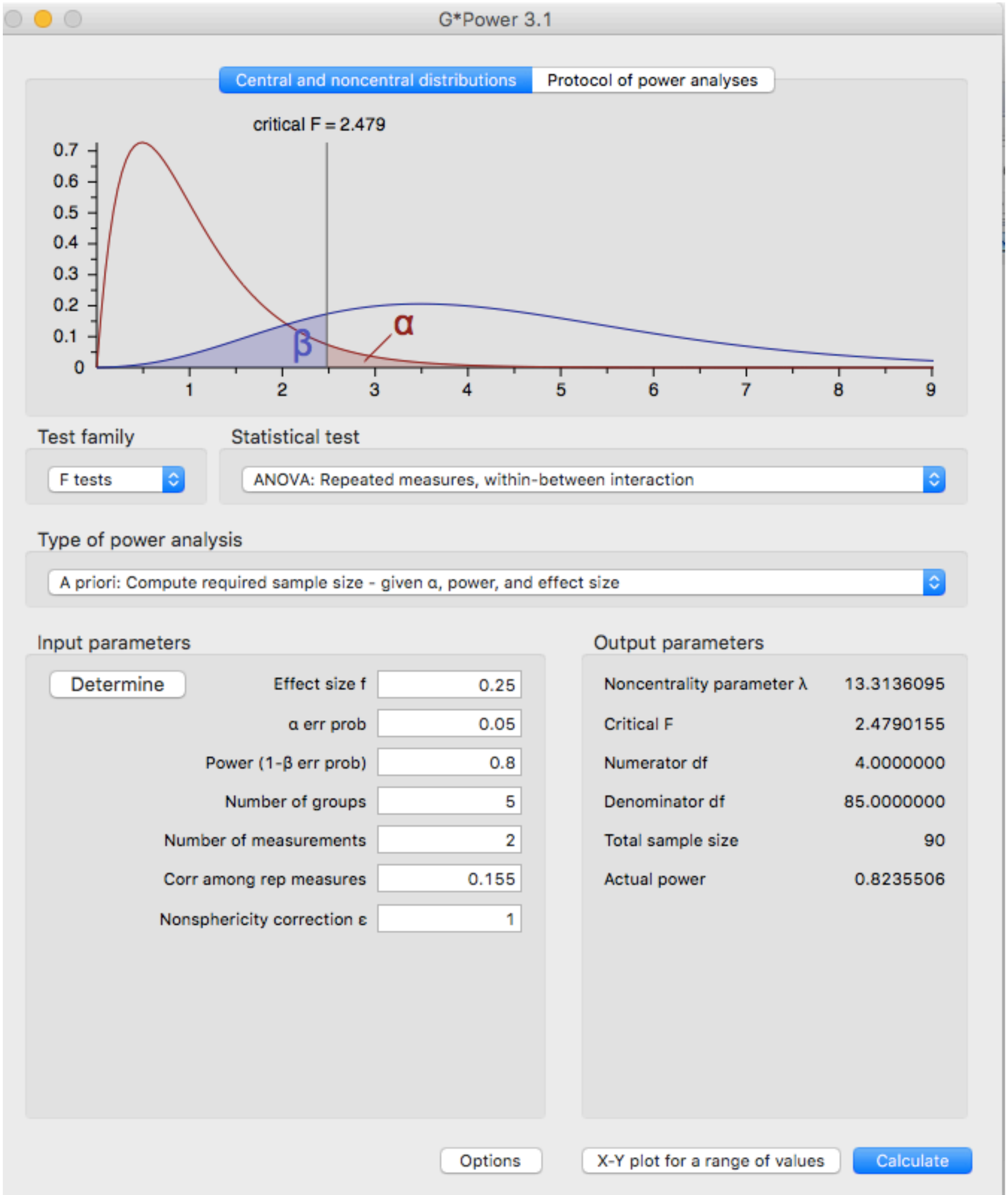
32. How much did your experience in the virtual environment seem like you were in a real place, able to directly sense and interact with the environment?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT AT ALL MODERATELY SO VERY MUCH

33. In the virtual environment, how strong was your sense of "being there"?

|_______| |_______| |_______| |_______| |_______| |_______| |_______|
NOT STRONG MODERATELY STRONG VERY STRONG

APPENDIX F: G*POWER3 ANALYSIS


G*Power 3.1
Central and noncentral distributions
Protocol of power analyses
critical F = 2.479
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
β
α
1
2
3
4
5
6
7
8
9
Test family
F tests
Statistical test
ANOVA: Repeated measures, within-between interaction
Type of power analysis
A priori: Compute required sample size - given α, power, and effect size
Input parameters
Determine
Effect size f 0.25
α err prob 0.05
Power (1-β err prob) 0.8
Number of groups 5
Number of measurements 2
Corr among rep measures 0.155
Nonsphericity correction ε 1
Output parameters
Noncentrality parameter λ 13.3136095
Critical F 2.4790155
Numerator df 4.0000000
Denominator df 85.0000000
Total sample size 90
Actual power 0.8235506
Options
X-Y plot for a range of values
Calculate

APPENDIX G: IRB APPROVAL USMA, WEST POINT, AND TEACHERS COLLEGE

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

MCUD 10 July 2015

MEMORANDUM FOR COL James Ness, Director, Engineering Psychology Programs, Dept of BS&L, United States Military Academy, West Point, New York 10096

SUBJECT: Notification of Initial IRB Approval for KACH IRB Protocol #15-0 22, IRBNet #413796-1 "Motivational feedback for sensor-free detected frustration within game-based training: Examining the influence of feedback messages on learning outcomes", PI: COL James Ness, PhD

1. Expedited review of the subject protocol (version dated 2 July 2015), the informed consent (version dated 2 July 2015), the Stimulus Questionnaire, the Motivation Feedback Demographics Questionnaire and the debriefing statement was conducted and approved by the IRB Chair on 9 July 2015 in accordance with 32 CFR 219.110 (b) (1), Category 4.

2. The KACH IRB was informed of the expedited review approval at the convened IRB meeting of 9 July 2015.

3. There are no outstanding human subjects' protection issues to be resolved. This study has been assessed as No Greater than Minimal Risk.

4. You may implement/start your study with the issuance of this KACH IRB approved letter dated 9 July 2015.

5. The study is approved for enrollment of up to 120 subjects.

6. The IRB Chair approved the initiation of this Minimal risk study for a one-year period, effective 9 July 2015 – 8 July 2016. The IRB stamped approved informed consent (version dated 9 July 2015) should be used when consenting subjects.

7. 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 8 June 2016 to ensure approval on or before 8 July 2016.

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

JEFFREY A. LAWSON
COL, MC
Chair, KACH IRB

TEACHERS COLLEGE
COLUMBIA UNIVERSITY

Teachers College IRB — *Approval Notification*

To: Jeanine De Falco
From: Karen Froud, IRB Chair
Subject: IRB Approval: 15-283 Protocol
Date: 08/07/2015

Please be informed that as of the date of this letter, the Institutional Review Board for the Protection of Human Subjects at Teachers College, Columbia University has given full approval to your study, entitled "*Motivational feedback for sensor-free detected frustration within game-based training: Examining the influence of feedback messages on learning outcomes* ," after a **Full Board Review**.

The approval is effective until **07/27/2016**.

The IRB Committee must be contacted if there are any changes to the protocol during this period. **Please note:** If you are planning to continue your study, a Continuing Review report must be submitted to either close the protocol or request permission to continue for another year. Please submit your report by **06/29/2016** so that the IRB has time to review and approve your report if you wish to continue your study. The IRB number assigned to your protocol is **15-283**. Feel free to contact the IRB Office (212-678-4105 or IRB@tc.edu) if you have any questions.

Please note that your Consent form bears an official IRB authorization stamp. Copies of this form with the IRB stamp must be used for your research work. Further, all research recruitment materials must include the study's IRB-approved protocol number. You can retrieve a PDF copy of this approval letter from the Mentor site.

Best wishes for your research work.

Sincerely,

Karen Froud, Ph.D.
Associate Professor of Neuroscience & Education
IRB Chair

APPENDIX H : CONSENT FORM

UNITED STATES MILITARY ACADEMY
CONSENT TO PARTICIPATE IN RESEARCH

KELLER ARMY COMMUNITY HOSPITAL
IRB APPROVED: JULY 9, 2015
APPROVAL EXPIRES: JULY 8, 2016

This consent form is valid only if it contains the IRB stamped date

Study Title: Motivational Feedback for Sensor-free Detected Frustration within Game-Based Training: Examining the Influence of Feedback Messages on Learning Outcomes

You are asked to participate in a research study conducted at the United States Military Academy by Principal Investigator (PI), COL James Ness, Ms. Jeanine DeFalco, and Ms. Vasiliki Georgoulas. 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

You have been invited to participate in a research study conducted at West Point, NY. This study will evaluate the effectiveness of automated feedback messages in the serious game vMedic. The results of this research will inform future designs of feedback messages for intelligent tutoring systems.

Participation is entirely voluntary. As vMedic is an animated video game comparable to video games currently available for entertainment (e.g., *Call of Duty*) it is appropriate for a mature audience as is the target population for this study. However, some images may be disturbing due to their graphic nature. You may refuse to participate or withdraw from the study at any time without penalty or loss of benefits to which you are otherwise entitled and will receive participation credit based on the amount of time that you participated.

EXPECTED DURATION OF PARTICIPATION

Participants in the study will participate in a single experimental session of approximately 60 minutes.

PROCEDURES

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

1. fill out a demographic survey
2. take a pre-test on hemorrhage control
3. complete a short course on hemorrhage control in the Tactical Combat Casualty Care (TC3)
4. play four scenarios in the serious video game, vMedic
5. complete two questionnaires:
   (a) the first questionnaire will ask questions about how you overcome obstacles and achieve goals
   (b) the second questionnaire will ask about how engaged you were playing the video game vMedic
6. take a post-test on hemorrhage control

Please note that some of the material in the courseware and game can be viewed as graphic in nature, as it involves material relevant to common battlefield injuries.

Version 1.0, dated 9 July 2015 Revised after IRB review

Teachers College, Columbia University
Institutional Review Board
Protocol Number: 15-283
Consent Form Approved Until: 07/27/2016

0:00 -- 5:00
Participant is asked to sign a consent form. After signing the consent form, participants will log-in to the GIFT system with an assigned participant ID number. GIFT is the intelligent tutoring system through which the entire experiment will be delivered to the participant. This task should take 5 minutes.

5:00 – 10:00 Once logged on, the participant is asked to complete a demographics questionnaire. This task should take 5 minutes.

10:00 – 20:00 GIFT will then present to the participant a pre-test on hemorrhage control that needs to be completed by the participant to the best of their ability. There will be 20 questions on the pre-test that includes material on hemorrhage control, tactical field care and care under fire. The pre-test will be used to determine the effectiveness of the hemorrhage control course. This entire process in estimated to take 10 minutes to complete.

20:00 -- 25:00
After the pre-test, GIFT will launch a PowerPoint presentation on hemorrhage control, tactical field care, and care under fire. It is estimated participants will spend an average of 5 minutes with this content.

25:00 -- 45:00
Following the PowerPoint, GIFT will launch vMedic and participants will have a tutorial on the TC3Sim game controls. This introduction will review interface components and allow participants to interact with environment elements prior to the start of the scenario-based training event.

Following the tutorial, the participant will go through four scenarios and apply what they know of hemorrhage control, tactical field care, and care under fire in the serious video game vMedic. The participant should expect to spend about 20 minutes on this task.

45:00 -- 50:00
After completing the four scenarios, GIFT will launch the first questionnaire.

After completing the four scenarios in vMedic, GIFT will launch the first questionnaire, the Grit questionnaire (Duckworth and Quinn, 2009). This questionnaire will ask how the participant achieves long term goals and deals with obstacles.

After this is completed, GIFT will launch a second questionnaire, the Presence questionnaire (Witmer and Singer , 1994, 2005). This questionnaire will ask about the experience the participant had in the simulated game environment.

50:00-60:00

Once this second questionnaire is completed, GIFT will launch a post-test consisting of 20 questions. This section should take approximately 10 minutes.

Following completion of the post-test, the experiment is complete and the participant can log out of the system.

After the participant is logged out of the system, proctors will debrief the participant on the experiment experience and encouraged to ask questions about the experiment.

During the debriefing, participants will be encouraged to provide any feedback about the experimental session. The experimental proctor will be present in the room where subjects are participating. Subjects will be monitored for adverse effects to the experimental stimuli. As participants will be sitting at a desk and using a computer the amount of risk is minimal. If participants are experiencing a negative reaction, the session will be ended and the individual will be escorted to the nearest health official located on the premises. An individual participant session should take an hour to complete.

POTENTIAL RISKS AND DISCOMFORTS

The risks of this study are minimal, and similar to those experienced when using standard computer-based military training materials. As vMedic is an animated video game comparable to video games currently available for entertainment (e.g., *Call of Duty*) it is appropriate for a mature audience as is the target population for this study. However, some images may be disturbing due to their graphic nature.

Participants' data will remain anonymized. Each participant upon completing the consent form will receive a participant ID number. Only de-identified datasets will be shared further. All data will be de-identified and at the completion of the study or upon withdrawal the code for relating the subject number to the volunteer will be destroyed. Only aggregated data will be reported. Consent forms will be stored, locked in a filing cabinet and will be destroyed by shredding after a period of five years has lapsed.

ANTICIPATED BENEFITS
There is no guarantee you will receive any benefit from this study other than knowing that the information may help scientists learn more about the long term benefits for examining the influence of feedback messages on learning outcomes.

MEDICAL CARE FOR RESEARCH RELATED INJURY
N/A

NUMBER OF SUBJECTS THAT WILL TAKE PART IN THIS STUDY
120 United States Military Academy Cadets

CONFIDENTIALITY
The subject identification Master List will be safeguarded by the Principal Investigator. All data will be de-identified and at the completion of the study or upon withdrawal the code from the Master List for relating the subject number to the volunteer will be destroyed. Only aggregated data such as survey data will be reported. Your responses and results in the experiment will be de-identified and maintained by the researchers. All files concerning the protocol will be maintained in a secured file cabinet in the Department of Behavioral Sciences and Leadership as required by regulation. The subject identification key will be safeguarded by the PI. Only de-identified datasets will be shared further. All data will be de-identified and at the completion of the study or upon withdrawal the code for relating the subject number to the volunteer will be destroyed. All data will be stored in a CAC enabled and password protected laptop. Consent forms will be stored and locked in a filing cabinet. Consent forms will be stored and locked in a filing cabinet separate from all other data and will be destroyed by shredding after a period of 5 years has lapsed. All files concerning the protocol will be maintained in a secured file cabinet in the Department of Behavioral Sciences and Leadership as required by regulation.

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
You will not be paid for your participation in this study. However, cadet volunteers will be recruited from the volunteer pools in PL100 and PL300 and will be offered extra credit for participation.

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. An alternative to obtaining extra credit can be obtained through either writing a research paper

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 (United States Military Academy)

If you have any questions about your rights as a volunteer in the research, please feel free to contact the USMA Human Protections Administrator at (845) 938-7370.

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

COL James Ness, 845-938-0239 (United States Military Academy)
Vasiliki Georgoulas, 845-938-3598 (United States Military Academy)

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) 938-6684.

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)

Teachers College, Columbia University
Institutional Review Board
Protocol Number: 15-283
Consent Form Approved Until: 07/27/2016

APPENDIX I: FEEDBACK MESSAGES

Condition 1: Control-Value Theory

1. “Studies have shown that between 17%-19% of deaths in Vietnam could have been prevented if tourniquets had been used,” (DePillis, 2013).

2. “A 2008 study from a hospital in Baghdad found an 87% survival rate with use of tourniquets,” (DePillis, 2013).

3. “There is no room for hesitation or consultation in facial injuries, and quick action (3-10 minutes) is critical to the survival and recovery of injured soldiers,” (Shuker, 2011).

4. “The number one cause of preventable deaths in active shooter events is blood loss, and the best way to stop blood loss is to properly apply a tourniquet,” (Jacobs et al., 2013).

5. “The first U.S. casualty to die in the war from enemy fire was a Special Forces Soldier, SFC Nathan Chapman, who died during medical air-evacuation on 4 January2002 from isolated limb exsanguination without tourniquet use,” (Kragh et al., 2013).

Condition 2: Social Identity Theory

1. As General Maxwell Thurman said, “Make good things happen for our Army."

2. Remember, solder, what General Patton said: “An Army is a team. It lives, sleeps, eats, and fights as a team.”

3. “Every single man in this Army plays a vital role,” said General Patton. “Don't ever let up. Every man has a job to do and he must do it.”

4. General MacArthur once said: “Duty, Honor, Country, are three hallowed words that dictate what you ought to be, what you can be, what you will be.”

5. General Patton said that the soldier is both a citizen and the Army, and the highest obligation and privilege of citizenship is the bearing arms for one’s country.

Condition 3: Self-Efficacy Theory

1. In this important combat situation, your best outcomes will be achieved if you persist.

2. You can succeed in this because you've been trained to succeed under all conditions.

3. Tell yourself that you will succeed because failure is not an option in this high stakes combat zone.

4. Difficult doesn't mean impossible. It means work harder till your combat mission is achieved.

5. In all combat situations, success comes from overcoming the things you thought you couldn't.

Control Condition 1 – Non motivational feedback messages

1. "Battlefield care emerged in Europe when Post-Revolutionary France established a system of prehospital care that included a corps of litter-bearers to remove wounded individuals from the battlefield," (Chapman et al., 2012).

2. "The modern combat medic has its roots in the American Civil War, when enlisted soldiers served as hospital stewards." (De Lorenzo, 2001).

3. "As of 10 September 2001, the unreliable, World War II–era U.S. Army tourniquet was the only widely fielded tourniquet in the U.S. military," (Kragh et al., 2013).

4. "In 2003, in the farmlands around Fort Bragg, Amanda Westmoreland became a tourniquet maker by melting and bending plastic tourniquet components in her living room, packaging and distributing thousands of assembled tourniquets early in the war against Iraq," (Kragh et al., 2013).

5. "The use of a tourniquet went from a means of last resort to a means of first aid and became the prehospital medical breakthrough of the wars in Afghanistan and Iraq," (Kragh et al., 2013).

Control Condition 2: NO MESSAGES

APPENDIX J: DEMOGRAPHIC QUESTIONNAIRE SEPTEMBER 2015

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 K: PRE-TEST SEPTEMBER 2015

Test #	Output # from Event Reporting tool in GIFT	Question	Choices of Answers
Pre test 1	S_Pre-test_SQ_396_Q300	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:	Blocked Airway. Cardiac arrest. Hypothermia. Shock.
Pre test 2	S_Pre-test_SQ_397_Q55	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?	Yes No
Pre test 3	S_Pre-test_SQ_398_Q70	The look-listen-feel method is used to:	See if the casualty is in shock. Approximate the amount of blood loss. Test the casualty's level of consciousness. Determine if the casualty is breathing

Pre test 4	S_Pre-test_SQ_399_Q178	What has historically been a problem with requests for medical evacuation?	Proper classification. Over classification. Priority classification. Routine classification.
Pre test 5	S_Pre-test_SQ_400_Q226	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?	You can now safely remove the tourniquet. You can now reevaluate the casualty's wound to see if other measures, such as a pressure dressing, would be more appropriate. You cannot remove a tourniquet once it has been applied.
Pre test 6	S_Pre-test_SQ_401_Q259	Which of the following statements are true? (Select all that apply)	Do not attempt to salvage a casualty's rucksack, unless it's critical to the mission Always attempt to salvage a casualty's rucksack Don't waste time taking a casualty's weapon and ammunition Take the casualty's weapon and ammunition if possible
Pre test 7	S_Pre-test_SQ_402_Q62	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?	As soon as you can Only after you have performed a full examination of the casualty Only after you have completed your treatment of the casualty Only if the casualty requires evacuation

Pre test 8	S_Pre-test_SQ_403_Q263	Hemorrhage control is the most important aspect of saving lives during Care Under Fire phase for what reasons?	A Soldier can go into shock and die quickly after injuring a large blood vessel Hemorrhage is the easiest thing to treat on the battlefield Hemorrhage is the leading cause of preventable death in combat Hemorrhage rarely leads to infection
Pre test 9	S_Pre-test_SQ_404_Q102	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?	Between the wound and the elbow. Directly over the elbow. A little above the elbow. Two inches distal to the shoulder joint.
Pre test 10	S_Pre-test_SQ_405_Q128	Which of the following is a sign or symptom of tension pneumothorax?	Skin becomes warmer and dry. The casualty develops progressive respiratory distress. You can still feel the casualty's pulse at his wrist. The casualty's breathing has returned to normal.
Pre test 11	S_Pre-test_SQ_406_Q113	When you check for breathing, you should:	Watch the casualty's chest to see if it rises and falls. Listen for sounds of breathing. Feel for any exhaled breath blowing against your face. All answers are correct.
Pre test 12	S_Pre-test_SQ_407_Q51	You have been wounded and are still under enemy fire. You are unable	Call for help Play dead

		to return fire and there is no safe cover nearby. What should you do?	
Pre test 13	S_Pre-test_SQ_408_Q52	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?	Yes No
Pre test 14	S_Pre-test_SQ_409_Q47	You are going to the aid of an injured soldier while under fire. What should be your first action upon reaching the soldier?	Check the soldier for responsiveness Check the soldier's pulse Check the soldier for breathing Check the soldier for shock
Pre test 15	S_Pre-test_SQ_410_Q103	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?	Apply an Emergency Bandage to the wound. Apply a tourniquet two inches above the amputation site. Apply a pressure dressing to the stump. Apply a tourniquet two inches above the elbow.
Pre test 16	S_Pre-test_SQ_411_Q106	The lower part of the casualty's arm has been amputated. You have applied a tourniquet. How is the stump treated?	The stump is dressed and bandaged. The stump is left exposed to facilitate drainage.

Pre test 17	S_Pre-test_SQ_412_Q199	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?	Open the casualty's airway (head-tilt/chin-lift). Perform needle chest decompression. Apply a tourniquet to a limb with severe bleeding from a wound. Insert a nasopharyngeal airway. All listed actions can be performed before moving the casualty to a safe location.
Pre test 18	S_Pre-test_SQ_413_Q260	Pulse can be used to indicate the extent of blood loss	True False
Pre test 19	S_Pre-test_SQ_414_Q249	During casualty care under fire, you should do which of the following?	Always drag casualties out the line of fire Never attempt to move a casualty because it is too dangerous Use any means available to move the casualties as quickly as possible Always administer care before attempting to move a casualty
Pre test 20	S_Pre-test_SQ_415_Q39	What are the three most common medically preventable causes of death on the modern battlefield?	extremity hemorrhage, tension pneumothorax, airway obstruction -extremity hemorrhage, tension pneumothorax, gunshot wound -amputation of a limb, tension pneumothorax, gunshot wound -amputation of a limb, infection, airway obstruction

APPENDIX L: POST TEST SEPTEMBER 2015

Test #	Output # from Event Reporting tool in GIFT	Question	Choices of Answers
Post test 20	S_Post-Test_SQ_436_Q202	You are in a tactical field care situation (not under enemy fire). A soldier is lying on his back. He is breathing and alert. He has no serious wounds to his extremities or head. You see an entrance wound on the casualty's chest. What should you do now?	Seal the chest wound and check for other open chest wounds on his back. Apply an Emergency Bandage to the wound on his chest and begin rescue breathing (mouth-to-mouth resuscitation). Perform needle chest decompression. Insert a nasopharyngeal airway into the casualty's nostril.
Post test 19	S_Post-Test_SQ_437_Q310	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?	Flex the casualty's knees so that they are raised and his feet are flat on the ground. Place a nasopharyngeal airway in each nostril. Place a field pack or other object under his feet so that the feet are elevated slightly above the level of his heart. Have the casualty drink a full canteen of warm, salted water.
Post test 18	S_Post-Test_SQ_438_Q54	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?	Nothing. Leave the tourniquet in place Examine the wound to see if it is bleeding and can be controlled using other means Place another tourniquet above the first tourniquet and leave both tourniquets

			in place Place another tourniquet above the first tourniquet and remove the first tourniquet
Post test 17	S_Post-Test_SQ_439_Q84	You are accompanying an unconscious casualty during evacuation. What should you do?	Monitor the casualty's breathing. Monitor the bleeding from the casualty's wounds. Reinforce dressings, as needed. All answers are correct.
Post test 16	S_Post-Test_SQ_440_Q61	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?	Tactical evacuation care Tactical field care Care under fire

Post test 15	S_Post-Test_SQ_441_Q222	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?	Roger, Roger, I have a request for evacuation. Over. Please dispatch (an air) (a ground) ambulance to the following location. (State location.) I require medical assistance ASAP. Over. I have a MEDEVAC request. Over.
Post test 14	S_Post-Test_SQ_442_Q83	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?	None of the below. A tactical field care environment allows you to focus more on the evaluation, treatment and evacuation of the casualty. A tactical field care environment limits you to only to the treatment of life-threatening bleeding from a limb and movement to safety.
Post test 13	S_Post-Test_SQ_443_Q255	How long can you leave a tourniquet on without having to worry about the loss of a limb?	10 Minutes 30 Minutes 1 Hour 2 Hours 5 Hours
Post test 12	S_Post-Test_SQ_444_Q105	Once you have tightened an improvised tourniquet, you must:	Secure the windlass so that the tourniquet will not unwind. Apply an Emergency Bandage over the windlass. Remove the windlass and tie the tails in a nonslip knot.

Post test 11	S_Post-Test_SQ_445_Q104	Which one of the following statements gives a proper rule for tightening a tourniquet?	A tourniquet should be loose enough so that you can slip two fingers under the tourniquet band. A tourniquet should be loose enough so that you can slip the tip of one finger under the tourniquet band. 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. 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.
Post test 10	S_Post-Test_SQ_446_Q42	Your unit is in ground combat. You see a soldier fall as though he has been shot. Your primary duty is to:	Continue firing at the enemy Stop firing and go to the fallen soldier Create a diversion before approaching the wounded Retreat back
Post test 9	S_Post-Test_SQ_447_Q44	When performing care under fire, which of the following actions can be performed before moving the casualty to a safe location? (Choose all that apply)	Perform cardiopulmonary resuscitation (CPR) Applying a tourniquet to control bleeding Perform needle chest decompression to relieve tension pneumothorax Administer the combat pill pack to control pain and infection

Post test 8	S_Post-Test_SQ_448_Q45	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?	Seek cover, return fire, play dead Seek cover, return fire, administer self-aid Play dead Seek cover, return fire, administer buddy-aid
Post test 7	S_Post-Test_SQ_449_Q41	How does evaluation and treatment of a casualty, in a care under fire situation, differ from a secure (tactical field care) situation?	While under fire, you only treat life-threatening bleeding from a limb While under fire, you can focus more on the evaluation and treatment of the casualty In a secure environment, you only treat life-threatening bleeding from a limb
Post test 6	S_Post-Test_SQ_450_Q50	When should you plan how to move a wounded soldier out of enemy fire?	Before you leave your place of safety, to go to the wounded soldier As soon as you reach the wounded soldier As soon as you have treated the life-threatening conditions As soon as you have treated all of the casualty's injuries

Post test 5	S_Post-Test_SQ_451_Q197	Which of the following describes a combat lifesaver?	A nonmedical soldier who provides lifesaving measures as his primary mission. A nonmedical soldier who provides lifesaving measures as his secondary mission. A medical soldier who provides lifesaving measures as his primary mission. A medical soldier who provides lifesaving measures as his secondary mission.
Post test 4	S_Post-Test_SQ_452_Q208	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?	Six inches above the wound. Two inches above the wound. Directly over the wound. Two inches below the wound. Six inches below the wound.
Post test 3	S_Post-Test_SQ_453_Q254	Blood sweeps are performed prior to measuring blood pressure or taking the casualty's pulse.	True False
Post test 2	S_Post-Test_SQ_454_Q250	Which of the following statements are true about "Care Under Fire"? (Select all that apply)	Medics should expect to return fire in a combat situation Casualties should return fire if able Airway management should be administered Medics should direct the casualty to move to cover and apply self aid if able

Post test 1	S_Post-Test_SQ_455_Q57	Which of the following is NOT part of care under fire?	Moving the casualty to safety Checking the casualty's level of consciousness -Treating an open chest wound Applying a tourniquet