DISSERTATION

THE ROLE OF ENGAGEMENT IN SYNTHETIC LEARING ENVIRONMENTS

Submitted by

Tristan Quinn Nelson

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2015

Doctoral Committee:

Advisor: Kurt Kraiger

Alyssa Gibbons
Chris Henle
Lucy Troup

ABSTRACT

THE ROLE OF ENGAGEMENT IN SYNTHETIC LEARING ENVIRONMENTS

Synthetic learning environments (SLEs) are often lauded for their ability to “motivate” trainees. However, little empirical evidence exists to support the popular claim that SLEs work because they are motivating. The reason for the lack of evidence supporting these claims lies in the often inadequate definition and measurement of the motivation experienced by these trainees. The present study makes a case for switching the focus from the nebulous term “motivation” to a more defined and measurable construct of training engagement which consists of one’s personal investments of physical, cognitive, and emotional energies. An integrated SLE model is outlined and used as a theoretical explanation for why and how SLEs impact trainee engagement and training outcomes. Study One explores the antecedents of engagement among a sample of undergraduate students playing an educational videogame. Study Two examines the comparative levels of engagement between two training conditions (SLE and E-learning control group) and explores the mediating role of engagement in the relationship between SLE characteristics and training outcomes. Results indicate some support for the integrated SLE model demonstrating that the user judgments of meaningfulness and availability predict trainee engagement. Furthermore, trainees in the SLE condition seem to experience significantly higher levels of engagement compared to their control group counterparts. However, training outcomes were uninfluenced by training condition and engagement did not play a mediating role. Theoretical contributions, limitations and directions for future research are discussed.

AKNOWLEDGEMENTS

I would like to dedicate this dissertation to my wife Chailyn. Thanks for sticking with me through the good and the bad and for being patient during the long and sometimes trying experience that is graduate school. Thanks to my advisor Kurt Kraiger for helping me turn this into a manageable and sometimes even enjoyable task. Thanks to my office mate Kyle Sandell for letting me ask dumb questions any time of the day and for allowing me to bring tuna fish for lunch. I also want to thank my committee members Chris Henle, Alyssa Gibbons, and Lucy Troup for supporting me and guiding my work. Finally, I would like to thank Dr. Sandra Winkler for allowing me to collaborate on this project and for being so helpful with all of my requests. Study Two in this dissertation was funded by a grant awarded to Dr. Sandra Winkler by the Agency for Healthcare and Research Quality AHRQ grant # R24HS022021.

TABLE OF CONTENTS

Introduction

Your hands are sweating; your pulse is racing; your eyes dart back and forth searching for possible threats as you rummage through a tattered backpack searching for something to defend yourself with. You quickly locate a 9mm semi-automatic handgun but your heart sinks as you realize you only have four rounds left. You load the gun with shaky hands and begin to creep forward in a crouching position. You know that if you keep out of sight and stay quiet you might be able to escape the building alive. You pause every few steps to look behind you and listen carefully. The darkness obscures your vision making each black doorway you pass feel like an undefinable nightmare; an abys from which anything could emerge. You can hear the distant moans of those that did not survive the outbreak; those that want to end your life, but you cannot see any of them yet. You move on through the dark past tables, and chairs, boxes and bodies. Suddenly you freeze; you can hear something or someone shuffling ahead of you. Your stomach turns as you contemplate your chances of survival. Should you run, should you stand and fight, or should you swallow your fear and hope you are passed by unnoticed? Refusing to be torn apart while cowering, you jump to your feet and fire at the creature. Your reflexes are quick your aim accurate, but four rounds are not enough. The dark figure moves forward, grasps you by the neck, and ends your life in a furious and bloody struggle. The game is over! Struggling with feelings of anger, fear, and frustration, you pitch your inordinately expensive controller across the room and quickly regret it.

This is the experience of millions of individuals who have played The Last of Us – a popular and highly acclaimed survival horror videogame. The Last of Us is a prime example of a game that draws players into a world that elicits physical, cognitive, and emotional reactions. Such an experience can have a relatively lasting impact on those who play it as evident by the numerous fan pages, discussion forums, books, and independent videos created by former players. However, the million dollar question for many training professionals and educators is can this type of motivation and interest be effectively harnessed for the purpose of learning. Those who extoll the merits of synthetic learning environments (including computer-based simulation games, virtual worlds etc.) claim it is possible and that this motivation is what makes them effective.

Background

Although videogames and other synthetic environments were originally thought to be the domain of children, this has changed as the videogame generations of the 1970s and 1980s have grown up. For example, Slagle (2006) reported that 40% of U.S. adults play computer or videogames, and we now know that the average age of videogame players is 30 with the percentage of videogame players over the age of 36 exceeding the percentage of players under 18 (Entertainment Software Association, 2013).

This adult interest in videogames has resulted in their influence on popular culture. For example, Resident Evil, a popular videogame on from CAPCOM, has been adapted to the big screen five times with another on the way in 2016 (Resident Evil Collection, n.d.). The impact of Resident Evil on popular culture can be seen in part by the veritable obsession that society currently has with zombies. The AMC TV show The Walking Dead, movies like World War Z and Warm Bodies, and tablet/phone games like Plants vs. Zombies are only a few examples of this obsession that has its roots in CAPCOM's popular videogame series (Barber, 2013). As another example of videogame influence on popular culture, you might consider the music. Since 1987, full orchestras dedicated to videogame-themed music have played to sellout crowds in North America and Asia playing the scores from games including Final Fantasy, Call of Duty, and Metal Gear Solid (Davis, 2005).

One of the most recent examples of how videogames impact culture is through their influence on education and learning. Learning design in adult education is beginning to integrate the use of videogames and other immersive learning environments for instruction in universities, the military, and business organizations. For example, the U.S. military has spent more than $32 million on the development of America's Army (Edwards, 2009) – a videogame designed as both a recruitment and training tool with a focus on teamwork and adherence to Army values (America's Army, n.d.). Likewise, Cisco Systems has developed an arcade of games to help train employees in wireless network setup, binary code, and an array of other skills (Learn while you play, n.d.).

Videogame influences on Learning

It has become apparent that videogames for training and development purposes have gained popularity in both practice and research over the past decade. References to the use of these games have begun to permeate books, journal articles, popular magazines and professional conference presentations (Kapp, 2012). Research indicates that as far back as 2006, one third of companies were planning to use videogames for training in the future (Meister, 2006) and that more recently 55 percent of Americans report wanting to work for a company that uses videogames to increase productivity (Saatchi & Saatchi, 2011). In sum, videogames for training have become a 'hot-topic' among academics and practitioners.

So why are academics and practitioners so interested in using videogames for training purposes? First, many contend that videogames are fun and often training is not (Garris, Ahlers, & Driskell, 2002). Companies and employees both like the idea of breaking away from the stereotypical ‘boring training’ to implement something that is attention grabbing and fun. Second, videogames are perceived as being naturally motivating and/or “addictive.” Evidence of this addictive quality can be seen among those who play Massively Multiplayer Online Role Playing Games (MMORPGs) such as World of Warcraft. These gamers spend an average of 20 hours per week playing their game and many gamers play for stretches as long as 10 hours (Yee, 2006).

Third, videogames have many characteristics that make them useful training tools. For example, they are active; meaning that players do not passively experience the game content, they interact with it. In Portal 2 you don’t just hear about how to solve complex spatial puzzles, you learn by actually moving your body through the puzzle to solve it. Research has shown that training is particularly effective when learners are actively engaged with the material (Bell & Kozlowski, 2008) – a natural consequence of videogames. Videogames are also ripe with feedback. We know that specific and immediate feedback is critical for learning (Gagné & Medsker, 1996), and videogames include feedback as a natural consequence to gameplay. For example, if you make a mistake in Super Mario Brothers you either shrink or die. Last, academics and practitioners are interested in videogames for training because they may be an effective way to lower attrition rates which are notoriously high for other forms of e-learning (Flood, 2002). The idea is that because videogames have addictive motivational qualities trainees are more likely to persist and the high attrition rates common in other types of non-instructor-led training will be lessened.

In sum, people are interested in using videogames for training purposes because videogames can be fun, entertaining, motivating, addictive, and because they possess characteristics of high quality training. Many believe that if such qualities could be harnessed so that trainees put this same time and energy into learning training content, then improvements to learning and transfer outcomes would likely follow.

Definition

Before discussing the research on videogame training, and its relatives, it is important to provide the reader with a definition. Although the videogame training literature is rapidly expanding there is still significant confusion related to their definition. For example, the following terms are often used interchangeably: serious games, instructional games, computer games, simulations, training videogames, virtual world trainings etc. (Hays, 2005). While there are some notable differences among a few of these terms (videogame training, simulations, and virtual worlds), I believe that they share many characteristics and can all be categorized under the umbrella term "synthetic learning environments" (SLE) (Cannon-Bowers & Bowers, 2010).

To illustrate, consider the distinctions made between simulations and training videogames. Some hold that the defining feature of simulations is their grounding in reality. However there are multiple examples of computer simulations that do not recreate real-world scenarios with great fidelity (North, Sessum, & Zakalev, 2003; Taylor & Chi, 2006). Likewise there are numerous examples of training videogames that mirror real-world scenarios quite closely (Air marshaller, n.d.; Garris & Allers, 2001; Kelly, 2014).

Similarly distinctions have been made between training videogames and virtual worlds used for training purposes. This distinction has been more implicit than explicitly stated. For example, a literature search for the terms serious games, videogame training, or simulation games often does not retrieve research involving virtual worlds for training. Outside of training, the differences between videogames in general and virtual worlds have been more explicitly stated. Some suggest that virtual worlds differ from simulations in three ways. First, some contend that videogames are goal oriented while virtual worlds are non-goal oriented (e.g., no point system and no discernable wins). Second, they contend that unlike videogames, virtual worlds host many "players" interacting at the same time in a social context. Third, they argue that unlike videogames, virtual worlds are live and changing even when an individual player leaves (Boskic, 2015). While all of these distinctions have some merit, I believe that the lines between videogames (or simulation games in this case) and virtual worlds has become far too blurred for practical application.

Take Grand Theft Auto 5 as an example. This popular videogame features a protagonist who moves through the fictional city of Los Santos completing "missions" that may involve anything from simple deliveries to targeted assassinations. Money is awarded for each completed mission and new missions and areas are opened as the player progresses. Although this would seem to fit clearly into the traditional definition of a videogame Grand Theft Auto 5 includes many features that are characteristic of virtual worlds. For example, you can interact with multiple other avatars whose activity continues even after you turn off your console. You can also play the game without any discernable goals or objectives. For instance, some players gamble, buy homes or businesses, go sightseeing, hunting or to even go to the movies (Buffa, 2014). Hence, each of the supposed differentiators between videogames and virtual worlds has been challenged by Grand Theft Auto 5 and many other games like Titanfall, Destiny, and the Sims.

Likewise there are many examples of virtual worlds crossing the boundaries into the videogame realm. For instance, Second Life (one of the best known and most used of all the virtual worlds) features islands where zombies can be fought off with an array of destructive weapons. World of Warcraft, another prominent virtual world, involves "leveling-up" an avatar (i.e., increasing skills and abilities), exploring the landscape, fighting various monsters, completing quests, and interacting with non-player characters (NPCs) or other players. Truly, videogames can be virtual worlds and Virtual Worlds can be videogames. Any differentiation that might be made depends on how the videogame or virtual world is structured and used.

Up to this point I have been using the term "videogame" for much of my discussion. However, the present study examines data from both videogame and virtual world trainings and leans on research from the areas of videogame training, simulations, and virtual world training. To the extent that these three technology-based training media share their defining characteristics, I believe I can rely on research from each area. This considerable overlap is shown in table 1.

In accordance with Cannon-Bowers and Bowers (2010) I consider videogame training, simulations, and virtual world training to be overlapping constructs described by the umbrella term synthetic learning environments (SLE). SLEs are defined as a variety of technology-based training media or approaches that have as an essential feature the ability to augment, replace, create, and/or manage a learner's actual experience with the world by providing realistic content and embedded instructional features.

Research on the effectiveness of synthetic learning environments

So, is the excitement around SLEs misplaced or do they really result in positive training outcomes? There have been a number of studies over the past 20 years that have sought to answer this question, and although the literature is somewhat fragmented, the general conclusion is yes. For example, multiple empirical studies have demonstrated that virtual worlds can be effective training environments (Burton, Martin, & Robins, 2013; Finkelstein et al., 2005; Heiphetz & Woodill, 2010; Jacobson, 2006). However, the utility of these environments, above and beyond other forms of training, is more debatable. For example, in their review of the literature, Dalgarno and Lee (2010) stated that much of the research involving virtual worlds for learning is “show and tell” in that many studies simply present case studies that fail to make meaningful comparisons to other forms of training or provide links between the characteristics of virtual world training and meaningful outcomes. Similarly, some experts have claimed that comparisons between SLEs and control groups cannot be made since the training design is inextricably tied to the media making such comparisons essentially meaningless (Clark, 1994). On the other hand, some have argued convincingly that virtual worlds may be particularly useful for particular types of training activities or learning goals. Dalgano and Lee (2010) outlined five “learning affordances” of virtual worlds which include their ability to facilitate learning tasks that lead to improved transfer, and their ability to facilitate experiential learning tasks not possible in the real world. In sum, research on virtual worlds seems to indicate that they can be effective for training if well designed and appropriately used (Jacobson, 2006).

In terms of videogame training research there seems to be a larger research base from which to draw including review articles and meta-analyses. A review by Randel, Morris, Wetzel, and Whitehill (1992) examined 67 studies of which 38 showed no performance differences between instructional games and conventional learning approaches. Alternately, 27 of the studies favored the videogame condition, and three favored the conventional condition. Final conclusions from this review were that the positive effects of videogame learning are most likely to be present when the content is specific and the objectives are clearly defined. Vogel et al. (2006) performed a meta-analysis on 32 qualifying studies comparing videogame training conditions to a control. Results showed a strong positive overall effect size in favor of videogame training for cognitive gains and attitudes when compared to conventional training conditions. These findings led the authors to conclude that videogames do have a positive effect on both learning outcomes and attitudes toward training.

A more recent meta-analysis by Sitzmann (2011) helped identify some important moderating variables that can impact the effectiveness of videogame training. After analyzing 65 studies meeting the basic inclusion criteria, Sitzmann concluded that videogames result in better retention as well as better declarative and procedural learning outcomes compared to more conventional approaches. However, Sitzmann also noted that the effectiveness of videogame training depends on its ability to actively engage the trainee. Videogames that failed to actively engage learners actually resulted in poorer outcomes when compared to conventional training. Likewise, Sitzmann found that trainees actually learned less from the videogame training conditions compared to the comparisons when the comparison conditions were designed to be actively engaging.

Finally, Girard, Ecalle, and Magnan (2013) criticized the Sitzmann (2011) meta-analysis for her inclusion of games that were “too old” to be considered true training videogames, and for her broad definition that allowed the inclusion of studies concerned more with computer-assisted learning rather than pure training videogames. For their meta-analytic study, Girard et al. used more restrictive inclusion criteria. This resulted in a final sample size of only nine. With an insufficient sample size the authors were unable to make definitive conclusion from their findings. However, results showed that of the 11 games analyzed, three were more effective than the comparison group, seven showed no effect, and one showed mixed effects.

In summary, research to this point has shown that virtual worlds and videogames can be effective learning environments. This is likely due to a combination of factors. First the SLE must be used to train individuals on material or skills that are appropriate (e.g., learning tasks that are not possible in real life such as emergency preparedness). Second, the SLE must encourage the active participation of the learners. Third, SLEs must be well designed keeping in mind the basic principles of training design. If each condition is met, SLEs may result in training performance comparable to or better than their comparisons.

Why do synthetic learning environments work?

As discussed above, research suggests that SLEs are commonly as effective, and are sometimes more effective than conventional methods of training (Kapp, 2012). What we don't know is why they are effective. There are numerous books, journal articles, and other professional publications that suggest the positive effects for SLEs are due to high levels of trainee motivation. For example, Vogel et al. (2006) explained their results supporting the effectiveness of videogame training by declaring: "This result agrees with the current overall theory stating that interactive experiential activities that increase motivation also show increased learning outcomes." Similarly, Ricci, Salas, and Cannon-Bowers, (1996) concluded: "One of the primary reasons given as an explanation for the effectiveness of computer-based gaming is the motivational appeal of participating in gaming as compared with a more traditional learning format." Many other publications have made similar arguments suggesting that motivation is responsible for the effectiveness of SLEs (Byers, 2007; Clark, 2007; DeRouin-Jessen, 2008; Garris et al., 2002; Girard et al., 2013; Malone, 1981; Randel et al., 1992; Rastegarpour & Marashi, 2012; Sitzmann, 2011).

This argument has been proposed by the authors of various theoretical articles. Malone (1981) contended that intrinsic motivation can be increased through the use of game characteristics such as curiosity, challenge, and fantasy. Malone also argued that this increase in intrinsic motivation results in more time spent in training, more effort devoted to learning, and an increased likelihood of using the knowledge and skills in the future. In sum, the crux of Malone's theory is that videogames are successful for training purposes through their ability to increase the level of learner motivation.

A more recent theoretical article attempts to explain the motivational process that can occur when games are used for instruction (Garris et al., 2002). The authors outlined an input-process-outcome model of games (see Figure 1) showing the mediating link between training and game content and eventual training outcomes. This mediating link known as the game cycle is essentially the motivation process. This process begins with affective judgments where trainees evaluate their interest, enjoyment, task involvement and confidence in the game. These judgments then influence the user's behavior which may include engagement in the gameplay, effort exertion, concentration etc. In turn, these behaviors influence game feedback which has two major functions. First, it provides the trainee with information about their progress toward their goals, which can motivate further action. Second, feedback helps the trainee make judgments about the level of challenge and whether or not they are enjoying, or benefiting from the game (user judgments). At this point we find ourselves back at the top of the cycle and the process repeats. Hence, the game cycle is a perpetual process in which judgments impact user behaviors which impact feedback which once again influence the user's judgments. The game cycle illustrates that playing a game does not simply increase motivation in its own right, but that there is a complex interplay of judgments, behaviors, and feedback which determine one's level of motivation (Garris et al., 2002).

In sum, both Malone (1981) and Garris et al. (2002) stressed the importance of motivation in the effectiveness of SLEs. In fact, in Garris et al.'s model, the motivational process (the game cycle) is responsible for mediating the relationship between game characteristics and learning content and actual learning outcomes including skill-based, cognitive and affective learning outcomes.

Evidence for motivational differences?

While many have suggested that motivational differences are responsible for the effectiveness of synthetic leaning environments, there has been almost no research on this topic. To date there have been only two empirical studies directly comparing levels of motivation among trainees in traditional training verses those in a SLE. Interestingly, and contrary to common belief and theory, neither study found evidence for higher levels of motivation among trainees in the SLEs. First, Parchman, Ellis, Christinaz, and Vogel (2000) compared videogame training with classroom instruction and two other computer-based instructional strategies. Their findings showed that the game condition exhibited higher motivation on only one of the four motivation subscales – as did one of the computer-based conditions. No other differences were found. Despite these results, some have cited this article for evidence that videogame training results in higher levels of motivation compared to traditional methods (e.g., DeRouin-Jessen, 2008). Obviously such conclusions are unwarranted.

Additionally, Parchman et al.'s (2000) findings may be uninformative for two reasons. First, motivation was measured using the ARCS motivation questionnaire (Keller, 1987) which was designed to measure motivation to learn. As discussed below, motivation to learn may not be relevant in this context. Second, the ARCS questionnaire is an instrument used for formative evaluation in the instructional design process. In other words, instead of providing a measure of a trainee's overall level of motivation, the ARCS offers information about how the instructional design might be improved to increase motivation to learn. For example, consider the following items from the ARCS: "present[s] an example that does not seem to exemplify a given concept;"

or “model[s] enthusiasm for the subject taught.” Although these items might be useful for making changes to instructional materials, their utility as indicators of trainee motivation or engagement is probably minimal.

The second empirical study to examine motivational differences between a SLE and a comparison group was an unpublished dissertation (DeRouin-Jessen, 2008). The author assigned participants to one of four conditions - three game conditions (varying on levels of fantasy and reward) and one control condition (a computer-based PowerPoint presentation). Once again, results were contrary to expectations. Trainees in all four conditions reported equal levels of motivation, and satisfaction, completed a similar number of practice exercises, and performed equally well on both declarative knowledge and skill-based learning tests. These results led the author to conclude that “adding the “bells and whistles” of game features to a training program won’t improve learner motivation or training outcomes” (DeRouin-Jessen, 2008, p. iv). Clearly this conclusion runs contrary to the many publications asserting that motivational differences are the driving force behind the success of SLEs.

The purpose of the present study is to examine more closely, both conceptually and empirically, trainee motivation in SLEs. Many seem to assume that motivation is the driving force behind the success of SLEs but research to this point has failed to demonstrate any meaningful differences. Is motivation really the driving force and if so, why doesn’t the extant research support this? I believe the answer lies both in the confusion surrounding the definition and measurement of motivation as well as the theoretical frameworks used to understand motivation in SLEs.

Confusion around motivation: We are actually talking about engagement

Many of the authors claiming that motivation is the driving force in the success of SLEs fail to actually define the term motivation (often because they simply make the claim as if it were common knowledge). This failure to define motivation appears to have contributed to considerable confusion in the literature. When discussing motivation, some authors seem to refer to a trainee's interest in the novelty of the training while some seem to be referring to a trainee's willingness to return to the training of their own accord (Garris et al., 2002; Ricci et al., 1996). Still others seem to refer to intrinsic or even extrinsic motivation (Clark, 2007; Malone & Lepper, 1987). Truly there seems to be little agreement concerning what motivation is in the context of synthetic learning environments. This confusion is present not only in the broader training, education, and gaming literature, but also in the two empirical studies comparing motivation in SLEs to control groups (DeRouin-Jessen, 2008; Parchman et al., 2000). For example, the non-significant findings of Parchman et al. (2000) can likely be attributed to the assessment of the wrong construct – motivation to learn.

Motivation to learn is "a specific desire on the part of the trainee to learn the content of the training program" (Noe & Schmitt, 1986, p. 501). Research has shown that individual characteristics (e.g., locus of control, conscientiousness, cognitive ability), job characteristics (job involvement, organizational commitment), and even situational characteristics (e.g., climate, manager support) all influence motivation to learn (Colquitt, LePine, & Noe, 2000). However, neither the definition nor the research models proposed by Noe and Colquitt (2002) or Noe and Schmitt (1986) would suggest that changing the delivery method of the training content would have any impact on motivation to learn. For instance, it is difficult to imagine a graphic design student having a greater desire to learn Adobe Illustrator just because the class is offered in a game format instead of face-to-face. The medium used to present the training material does not make one more or motivated to learn. In hopes of ensuring that future researchers in this area do not continue to confuse motivation to learn with more relevant types of motivation I propose, and will test, the following hypothesis:

H1: There will be no significant differences in <u>*trainee motivation to learn*</u> *when comparing a SLE training group to a control group.*

Next, in the case of DeRouin-Jessen (2008), I believe he failed to fully capture the construct of motivation that the literature in general would suggest should vary between SLEs and comparison trainngs. DeRouin-Jessen employed a measure of intrinsic motivation which can be defined as the desire to exert effort on a task in the absence of external constraints or contingencies (Deci, 1975; Deci & Ryan, 1985). This desire to "exert effort" (intrinsic motivation) is only one component of the more holistic construct of engagement (Kahn, 1990) which involves the combination of physical investments (i.e., effort exertion or intrinsic motivation) in addition to cognitive and emotional investments. Therefore, if one measures only intrinsic motivation they miss out on the trainees' cognitive and emotional investments in the training. Indeed, research indicates that measuring the combined investment of physical, cognitive and emotional energies can create a more holistic and informative explanation for job (or training) performance than each (or only one) component alone (Phillips, Horstman, Vye, & Bransford, 2014; Rich, LePine, & Crawford, 2010). In reality, the insignificant results from DeRouin-Jessen may have been a consequence of measuring only part of the construct of interest. In sum, I believe this construct of interest is engagement which can be defined as "a personal investment of the self into individual work (in this case training) tasks that are performed on a job (or in training)" (Kahn, 1990, p. 700).

Truly, engagement seems to be the construct to which much of the SLE literature refers. For instance, the SLE literature has referred to "motivation" as the "harnessing of the trainees' full self" (Garris et al., 2002). Likewise, Kahn (1990) described engagement as a motivational concept that harnesses the employees' full self in terms of physical, cognitive, and emotional energies. Think about how this happens with popular videogames like Halo. The player is physically invested (their pulse is racing, their eyes are focused on the screen, and they are exhibiting a high level of hand-to-eye coordination), cognitively invested (they are constantly checking their heads-up displays, health levels, enemy locations and considering strategy) and emotionally invested (they may swear, scream, or throw the controller when "killed"). Indeed, engagement seems to capture what many authors are referring to when they claim SLEs work because they are "motivating."

In conclusion, although many articles concerning SLEs talk about "motivation," most fail to define the term. Furthermore, of the two studies that actually compare levels of motivation, one used a measure of motivation to learn – which should not be expected to differ based on training media; while the other used a measure of intrinsic motivation – which may not fully capture the construct being described in the general literature. To move theory and research forward with respect to motivation in SLEs, I propose using engagement (i.e., the physical, cognitive, and emotional investments of trainees) to examine the differences between SLEs and comparison training formats.

Defining Engagement

Till now, many SLE researchers have done a poor job defining motivation and many have used the terms motivation and engagement interchangeably. In a recently published special issue of *Simulation and Gaming*, experts concluded that this problem has contributed to considerable confusion in the literature (Filsecker & Kerres, 2014; Phillips et al., 2014). They also petitioned SLE researchers to move their focus beyond vague conceptualizations of motivation in general, or even intrinsic motivation alone, to a more complete conceptualization of engagement including physical, cognitive and emotional processes. For this reason, I believe it is crucial to clearly define engagement as it is conceptualized in the present study, outline how this definition differs from the many others in SLE research, and differentiate intrinsic motivation from engagement.

In the present study engagement will be defined as "a multidimensional motivational concept reflecting the simultaneous investment of an individual's physical, cognitive, and emotional energy in active, full [training] performance" (Rich et al., 2010, p. 619). This definition is rooted in the field of Industrial/Organizational (I/O) Psychology with Kahn (1990) serving as the theoretical foundation. Kahn conceptualized engagement as a motivational construct referring to the allocation of one's personal resources to their work role performance as well as how intensely and persistently those resources are applied. Furthermore, Kahn believed that these personal investments could take three forms: physical, cognitive and emotional. Both Kahn's theory and a measure of the three dimensions of engagement (see Rich et al., 2010) played important parts in the relative explosion of engagement research in the area of I/O Psychology (Christain, Garza & Slaughter, 2011). Nonetheless, to my knowledge this is the first extension of the theory and this measure to the context of training. Thus, instead of focusing on one's personal investment of physical, cognitive, and emotional resources into their *work role*, I am focusing on personal investments to one's *role as a trainee*. This extension may make intuitive sense to those who can recall a class or training in which they were physically engaged with the instructor or classmates (talking, using gestures, drawing diagrams etc.), cognitively engaged with the material (strategizing, problem solving etc.) and emotionally engaged (feeling happy or irritated with success or failure arguing a point). Indeed, I believe that Kahn (1990) can be extended to capture trainees' personal investments in learning contexts; especially those related to SLEs.

It is important to differentiate engagement as conceptualized by Kahn from other ways the term has been applied to learning. The notion of engagement in education or engagement with SLEs is complex and has been used in many different and potentially conflicting ways (Whitton & Moseley, 2014). Some authors have speculated that there are almost as many definitions as there are articles on engagement (Sharek & Wiebe, 2014). For example, engagement has been defined as immersion (Biocca & Levy, 1995), attention (O'Brien & Toms, 2008), flow (Csikszentmihalyi, 1997), effort (Dow, Mehta, Harmon, Macintyre, & Mateas, 2007), presence (Lombard & Ditton, 1997), enjoyment (Mayes & Cotton, 2001), or even as a combination of school attendance, involvement in academics, and so forth for those in educational research (Karweit, 1989). Likewise there have been a number of conceptual models of engagement proposed. Hickley and Anderson (2007) proposed a multilevel model distinguishing between different levels of engagement: immediate, close, proximal and distal. Whitton (2011a) proposed a model outlining the major contributing factors to engagement: challenge, control, immersion, interest, and purpose. Whitton and Moseley (2014) suggested a model based on six approaches to constructing engagement: participation, attention, captivation, passion, affiliation, and incorporation. Additionally, numerous methods of measuring engagement have been proposed including physical measures such as attention to the screen (O'Brien & Toms, 2008), cardiac interbeat intervals (IBIs) and facial electromyography (EMG) (Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006), self-report measures (O'Brien & Toms, 2008), or even mixed approaches such as game-clock clicks paired with cognitive load measures (Sharek & Wiebe, 2014).

Ultimately, there are multiple definitions, models and measurement approaches when it comes to engagement. While I recognize the merit of each of these, it is impossible to apply all of them to any one study. Additionally, I believe that both the theory and measure of engagement I use in the present study align quite well with recommendations of SLE researchers. For example, Filsecker and Kerres (2014) encouraged SLE researchers to move away from simply examining intrinsic motivation and begin exploring engagement as a multidimensional concept including physical, cognitive and emotional dimensions. Likewise, Phillips et al. (2014) encourage SLE researchers to avoid collapsing physical, cognitive and affective engagement into one shared category (as is the case with some measures of flow, immersion and commitment). Therefore, the dimensions of physical, cognitive, and emotional engagement that are the building blocks of Kahn's (1990) theory align nicely with recent recommendations from experts in SLE research.

Having defined engagement in the context of training, it is important to differentiate intrinsic motivation from engagement. This is because some SLE publications actually use the term engagement along with, or in place of motivation (Filsecker & Kerres, 2014). For example, although Garris et al. (2002) describes the motivational process of training videogames, they often use the terms "motivation" and "engagement" interchangeably. In fact, when referring to the motivational process of their input-process-output model, the authors actually call it engagement and suggest that this engagement is what leads to specific learning outcomes (Garris et al., 2002). Taking a closer look at both constructs (motivation and engagement); it becomes clear why the confusion between the two exists in the videogame training literature. Pinder

(1998) asserted that motivation determines the form, direction, intensity and duration of work-related (or in this case training related) behavior. In the same vein, Rich et al. (2010) stated that "Kahn's engagement concept is motivational because it refers to the allocation of personal resources to role performance and also to how intensely and persistently those resources are applied" (pg. 619). Thus, both constructs are concerned with the direction, intensity and duration of work (or in this case training) related behavior.

So how does engagement differ from intrinsic motivation? Intrinsic motivation has been formally defined as the desire to exert effort on a task in the absence of external constraints or contingencies (Deci, 1975; Deci & Ryan, 1985). Basically it is a natural inclination toward assimilation, mastery, spontaneous interest and exploration, and is something every person is born with (Ryan & Deci, 2000). For example, young children have a natural inclination to explore and experience their surroundings regardless of external rewards. Thus, intrinsic motivation is an inherent quality of all humans. However, each individual may be more or less intrinsically motivated to perform a given task based on personal interests, and feelings of competence, autonomy and relatedness that might be experienced by performing the task (Gagné & Deci, 2005). In the case of SLEs, intrinsic motivation would be defined as a trainee's desire to exert effort on the training in the absence of external constraints (e.g., managers' request).

Measuring intrinsic motivation has traditionally been done using the "free choice" measure described by Ryan and Deci (2000). In the classic form of this measure, participants are exposed to a task under different conditions (reward vs. no reward) after which they are given a period of free time where they are left alone in the experimental room and are neither asked to engage in or avoid the task. Assuming there are no extrinsic reasons to perform the task, the more time participants engage in the task, the more intrinsically motivated they are said to be (i.e., the task is providing opportunities to experience feelings of autonomy, control, and relatedness and so they make a choice to continue to exert effort on it).

Basically, intrinsic motivation involves the choices one makes concerning the allocation of effort to any given task. For instance, if I am intrinsically motivated to read, my intrinsic motivation related to that task might be measured through observations of the time spent reading in the absence of any other rewards outside of my own personal satisfaction. In this sense, intrinsic motivation is similar to the physical subcomponent of engagement because it involves the physical effort devoted to any given task (e.g., physically picking up the book and persisting in reading for a period of time). Alternatively, engagement concerns this same allocation of effort (physical engagement) in addition to the cognitive energy (cognitive engagement) and affective energies (emotional engagement) devoted to that task or work role. To illustrate the differences, imagine yourself binge watching Netflix on an uneventful Saturday afternoon. In this case (which some of you may have actually experienced), you might be said to be intrinsically motivated (or physically engaged) since you keep pressing the “play next episode” button but you are not necessarily cognitively engaged (you are bored and only half paying attention) or emotionally engaged (you stare apathetically at the screen). In conclusion, while intrinsic motivation may occasionally go hand-in-hand with engagement this is not always the case. Hence, the measurement of trainee engagement in SLEs could help capture a more complete picture of one’s personal investments in the training (i.e. physical investments/intrinsic motivation, cognitive investment, and emotional investment).

A Revised Framework for Understanding Engagement in SLEs

To this point I hope to have convinced the reader of the following six points. First, SLEs can be as effective, if not more effective, than non SLE comparisons. Second, many researchers believe that the success of SLEs is related to their ability to “motivate” trainees above and beyond what is possible with other training media. Third, there is almost no research to support this claim and the two empirical studies that do exist actually provide evidence to the contrary. Fourth, the failure of these two studies to find motivational differences may be due either the measurement of the wrong construct (motivation to learn) or failure to measure the entire construct of interest (engagement). Fifth, the appropriate construct is engagement as opposed to motivation to learn or intrinsic motivation alone. Sixth, engagement as it is defined above is distinct from intrinsic motivation and may be better suited for the motivational comparisons made between SLEs and other forms of learning. Ultimately, I believe that the reason we do not observe a more consistent pattern of SLEs outperforming their comparison trainings is due to our misunderstandings regarding the link between SLE characteristics and eventual learning outcomes. In other words, I believe that SLE characteristics can lead to improved learning outcomes when we understand the mediating role of engagement.

One of the main purposes of the present study is to outline a framework for understanding the role of engagement on the relationship between SLE characteristics and learning outcomes. It is my opinion that the model outlined by Garris et al. (2002) is essentially useful, and that it can be applied to all SLEs. However, I believe there are at least two modifications that might strengthen its ability to explain why and when SLEs will lead to expected outcomes. First, the game cycle needs to be clearly identified as a process of engagement. Although Garris et al. (2002) used the term engagement to describe the game cycle, it is used loosely; motivation and engagement are used interchangeably and no operational definition of engagement is offered. Thus, the model can be improved through use of the definition of engagement provided in the preceding section.

Second, I believe Garris's model can be strengthened by integrating the game cycle process with Kahn's original work on engagement and its antecedents. Specifically, Garris's user judgments (see Figure 2a) can be better understood as antecedents to engagement (see Figure 2b). Likewise, Garris's user behaviors can be better understood as displays of engagement (physical, cognitive, and emotional). Thus, I propose the integrated SLE model displayed in Figure 3. This model proposes that instructional content and game characteristics serve as the input to the "SLE cycle" process. The SLE cycle process is composed of three main components: user judgments, user behaviors, and system feedback. User judgments are trainee evaluations of meaningfulness, safety, confidence and availability. User behaviors are trainee investments of physical, cognitive, and emotional resources. System feedback involves any feedback provided by the game, instructors, and other trainees. The final component of the integrated model is learning outcomes. In the following paragraphs I elaborate on each of the components from the integrated SLE model shown in Figure 3 with special emphasis on user judgments and user behaviors (the areas of overlap between Garris's and Kahn's models). I conclude with a brief summary of the utility of this integrated model.

Input. Garris's model (Figure 2a) asserted that both instructional content and game characteristics serve as the input for the game cycle process. Game characteristics are fantasy, rule/goals, sensory stimuli, challenge, mystery, and control. Assuming at least some of these defining characteristics are present and effectively implemented in a SLE they should be expected to impact user judgments in the game cycle. For example, active learner control might be expected to have an impact on one's judgment of the meaningfulness of the training (e.g., my time is being well spent because I am actually practicing the skills I need). Likewise, mystery might impact judgments of availability (explained in the following paragraph) by encouraging the learner to devote more cognitive resources to uncover clues to help solve the game's mystery. Thus, I agree with the input portion of Garris's model and support the idea that both instructional content and SLE characteristics have an impact on user judgments (meaningfulness, safety, availability, and confidence) within the SLE cycle (see Figure 3).

User judgments. According to the integrated game model, the first component of the SLE cycle is user judgments. While involved in an SLE, users make judgments related to meaningfulness, safety, availability, and confidence. According to Kahn, meaningfulness is defined as a "sense of return on investments of itself in role performances" (p. 705). For my model, and in the context of a SLE, meaningfulness is defined as a sense of return on investments of one's self in training performances. Thus meaningfulness answers the following questions: What is the value of this training? Am I learning something that will help me in my job? Does the training make me feel more excited or energized (physically, cognitively, or emotionally)? Am I entertained or interested in the training? For example, a trainee in a SLE might feel that the training is providing them with an enjoyable mental break from the day's normal work schedule. In this situation, the trainee would be expected to report a relatively high level of meaningfulness due to his/her perception of the return on investment they are receiving from their training participation. Likewise this high level of meaningfulness should be expected in influence that trainee's user behavior (physical, cognitive, and emotional engagement) (see Figure 3). This return on investment is evident in many gamers who play videogames or log into a virtual world to take an emotional break from the regular hassles of the day. For these individuals the synthetic environment is providing them with a return on their investment in the form of emotional or even cognitive energy.

Next, according to Kahn, safety is defined as “a sense of being able to show and employ self without fear of negative consequences to self-image, status, or career” (p. 705). For my model and in the context of a SLE, safety is defined as a sense of being able to show and employ one’s self in a training context without fear of negative consequences to self-image, status, training performance, or career. Safety generally answers the question: “How safe do I feel in this training environment in terms of my ability to participate?” As an example, consider a trainee who has just been made the butt of a joke due to his or her question related to the training content. In such a situation, the trainee will likely judge the safety of the training environment to be low which may impact his or her user behaviors (physical, cognitive, or emotional engagement). Judgments of safety are commonly made among those playing games such as Call of Duty and Grand Theft Auto 5. These gamers decide how fully to invest themselves in the game context based on perceptions of safety among their peers (e.g., if I hook up my headset and contribute to the conversation will I experience negative consequences to my self-image, status, etc.?). Interestingly, research shows that SLEs are optimal environments for fostering feelings of safety. For example, computer-mediated environments in general are known for their ability to reduce concern about status differentials between authority figures and other group members, which helps to encourage participation and open discussion (Bikson & Eveland, 1990; Sproull & Kiesler, 1986; Zigurs, Poole, & DeSanctis, 1988).

Availability is the third user judgment and was defined by Kahn (1990) as “a sense of possessing the physical, emotional, and psychological resources necessary for investing self in role performances” (p. 705). For my model and in the context of a SLE, availability is defined as a sense of possessing the physical, emotional, and psychological resources necessary for investing self in training performances. For example, a trainee who has just had an argument with his/her spouse will probably have fewer emotional and cognitive resources to devote to training, thus impacting their user behaviors (psychical, cognitive, and emotional engagement). This is a common occurrence for completive players of games like NBA 2K15 who rely on their ability to “get their head in the game” (Khan, 2014). Those with relatively low judgments regarding their physical, cognitive, and emotional resources to devote to the game are unlikely to perform their best. The idea of psychological availability has been discussed by others as well. For example Kanfer and Ackerman (1989) demonstrated that the cognitive resources devoted to any task are dependent on one’s overall cognitive ability, the difficulty of the task, and the motivating mechanisms built into the task. Thus, one might expect the availability (at least the cognitive portion) of SLE trainees to depend on not only their overall cognitive ability, but the difficulty of the training tasks and the engaging mechanisms built into the SLE. Hence, a well-designed SLE has the potential to impact trainee availability by moderating the training task difficulty as well as impacting the devotion of physical, cognitive, and emotional resources through the SLE cycle.

The last user judgment is confidence, which was not part of Kahn’s original theory. However, Garris et al. (2002) made a compelling case for the importance of confidence as a user judgment. Although Garris et al. failed to define confidence, their description indicates that it is essentially a feeling of self-confidence derived from being able to practice skills and acquire knowledge in an environment where risk is minimized. Thus, it involves one’s feelings of self-confidence related to the training tasks. For my model and in the context of a SLE, confidence is defined as one’s level of confidence related to their ability to perform the training tasks. For example, a trainee who feels a high level of confidence in his or her ability to fly an airplane in a simulation is more likely to invest themselves through user behaviors (psychical, cognitive, and emotional engagement). In videogames like Uncharted 2: Among Thieves, the impact of confidence can be clearly seen. This game guides players through a series of small missions meant to ensure success by providing instructions and minimal consequences for failure (the game picks up exactly where the player last failed so there is no progress lost). Each small mission helps players gain confidence in their ability to control their avatar and execute complicated movements like taking cover, attacking while hanging onto a ledge etc. In circumstances like these, a growing sense of confidence contributes to the players' decision to invest more physical, cognitive, and emotional energy (i.e., engagement) in the game. Interestingly, research shows that SLEs may be optimal environments for fostering feelings of confidence since they act as low risk training environments for dangerous, complex, or stressful tasks (Driskell & Johnston, 1998). Similarly, some have suggested that World of Warcraft be used as a leadership incubator helping individuals get used to making decisions, failing, and using that information to improve. In this situation, the game builds confidence in the players abilities to explore options, make decisions and even bounce back from potential setbacks instead of experiencing what some have called "decision paralysis" (Reeves, Malone, & O'Driscoll, 2008).

In summary, Garris et al. (2002) originally outlined four user judgments. These were interest, enjoyment, confidence, and task involvement (degree to which individuals concentrate on and become absorbed in an activity). However, I believe that the first three of these (interest, enjoyment, and confidence) can be more fully captured by using the antecedents to engagement outlined in the framework above – specifically meaningfulness, safety, availability, and confidence (retained from Garris et al. (2002)). The fourth user judgement mentioned by Garris et al. (task involvement) is not represented in the user judgment section of the integrated SLE model because it is a user behavior and is fully captured by the combination of physical, cognitive, and emotional engagement which is discussed next.

User Behaviors. According to the SLE cycle shown in Figure 3, user judgments influence user behaviors – the second component of the cycle. Thus, user judgments are formed from initial and ongoing game play and determine the direction, intensity, and quality of subsequent user behaviors. These behaviors come in the form of physical, cognitive, or emotional investments of one's self in the training context (i.e., engagement). According to Kahn (1990), engagement is defined as "the simultaneous employment and expression of a person's preferred self in task behaviors that promote connection to work and to others, personal presence (psychical, cognitive and emotional), and active, full role performances" (p. 700). For my model, and in the context of a SLE, I define engagement similar to Rich et al. (2010) as "a multidimensional motivational concept reflecting the simultaneous investment of an individual's physical, cognitive, and emotional energy in active, full [training] performance." Thus, user behaviors take one of three forms: physical engagement, cognitive engagement, or emotional engagement. Physical engagement involves one's investments of physical energy into training performance. For example, a trainee who is physically engaged in a virtual firearm simulation (for examples see: http://www.gunfightersimulator.com/education.html) will likely have tensed muscles, an elevated pulse, and quick reactions to on-screen threats (e.g., fast draw and reload). Cognitive engagement involves one's investments of cognitive energy into training performance. For example, in the case of the firearm simulation the trainee would be making assessments of their physical position, possible threats, and areas useful for taking cover. Finally, emotional engagement involves one's investments of emotional energy in to their training performance.

Again, consider a trainee in a virtual firearm simulation. This trainee might display a high level of emotional engagement by swearing or shouting after being "shot" by one of the virtual threats.

In summary, although Garris et al. (2002) should be applauded for their work developing the components of the game cycle is clear that the user behavior section of their model was less developed than was the user judgment section (e.g., no cited research, no defined or measureable behaviors, elusions to engagement definition etc.). Hence, the use of Kahn's theory of engagement, and the measure of its physical, cognitive, and emotional components (Rich et al., 2010), allows for a much more complete and developed conceptualization of user behaviors. Therefore whereas Garris et al. said user behaviors consist of persistence (in gameplay), and the exertion of intense effort and concentration, I conceptualize user behavior as the investment of physical, cognitive, and emotional energy in training performance.

System Feedback. The final component of the SLE cycle is system feedback. In the cycle, user judgments lead to user behaviors which lead to system feedback. Kahn's engagement theory does not provide any parallel to this part of the game cycle which may well be a weakness of his theory. Indeed, feedback has been shown to be an important factor in regulating performance, motivation, and engagement (Crawford, LePine, & Crawford, 2010; Fried & Ferris, 1987; Saks 2006). In a SLE, feedback may take the form of scores, comments from other avatars or the instructor, goal attainment, or milestones met. If system feedback suggests consistent failure, the game may be deemed too hard while the opposite might cause the trainee to evaluate the game as being too easy. Thus, a balance must be met where the trainee feels challenged but not overwhelmed. For example, the trainee in the virtual firearm simulation discussed above might receive feedback in the form of points awarded for each room within a building cleared successfully. Similarly, feedback could be as simple as whether the trainee lived or died during the simulation. In both cases, the trainee's physical, cognitive, and emotional engagement will have a significant impact on this feedback. If they were physically, cognitively, and emotionally engaged system feedback should reflect this in the form more rooms cleared or a better survival rate.

Overview of the SLE cycle. Before moving on, it is useful to clarify the cyclical nature of the SLE cycle. Once system feedback is received, the cycle repeats with feedback influencing user judgments. Thus the SLE cycle involves user judgments (meaningfulness, safety, confidence, and availability) that serve as the antecedents to user behaviors (cognitive, emotional, and physical engagement), that in turn affect system feedback (indicators of progress including goals met, comments etc.). To illustrate one loop in the cycle, imagine a trainee who begins with an assessment that the SLE is relatively meaningful (user judgment). In other words, they believe the content will be useful to them in their job, and they feel that their time was well spent in that they learned something and enjoyed the fantasy aspects of the game. Concurrently, they will also make judgments about the safety and confidence related to the SLE and their personal level of availability. In this case, let's assume the trainee feels a high level of safety (they have not interacted with any other avatars yet and have had no feedback from the SLE) and they are experiencing a relatively high level of availability (e.g., they aren't overly sleepy or cognitively or emotionally drained). At this point, their judgments will lead to user behaviors including physical, cognitive, and emotional engagement. In this example, we are assuming the trainee made positive user judgments and so we could expect to see relatively high levels of each type of engagement.

Next, the trainee would receive system feedback. Let's assume that although they are receiving feedback that they are getting closer to attaining one of many goals, they also receive negative feedback from another avatar (a fellow trainee) who expresses annoyance with some of their actions. This source of feedback could impact the next cycle through a more negative assessment of safety. A lower judgment of safety could then impact the level of emotional (or cognitive or physical) engagement which could likewise affect system feedback again. Assuming our hypothetical trainee assesses the SLE as being lower in safety than they originally thought, they will likely be less emotionally engaged which could result in poorer performance and eventually negative feedback from the instructor. After this the cycle would begin again. In conclusion, user judgments (meaningfulness, safety, confidence, and availability), impact user behaviors (cognitive, emotional, and physical engagement), which impact system feedback.

Output. According to the integrated SLE model, the SLE cycle links the model's input (learning content and game characteristics) to its output (learning outcomes). These outcomes are the same as those outlined in Garris et al. (2002) and are categorized using the framework developed by Kraiger, Ford, and Salas (1993) that considers skill-based, cognitive, and affective training outcomes. Skill-based outcomes may include performance of technical or motor skills. Three cognitive outcomes may be impacted by the SLE cycle, namely declarative, procedural, and strategic cognitive outcomes. First, declarative outcomes involve knowledge of facts and data required to perform a tasks. Second, procedural outcomes involve knowledge about how to perform a task. Third, strategic outcomes involve the ability to apply rules and strategies to distal or novel cases. Finally, affective training outcomes involve beliefs or attitudes about an object or activity.

Model Summary. In all, I agree with much of the input-process-output game model described by Garris et al. (2002) but believe that integrating Kahn's original conceptualization of engagement and its antecedents sheds light on portions of the relatively underdeveloped game cycle. By combining portions of Garris et al. and Kahn's models I have created the integrated SLE model that overcomes the weaknesses of Garris et al. model and provides a clearer structure for testing. Testing this model will help us move forward in our understanding of what characteristics of SLEs lead to positive user judgments, user behaviors, and system feedback and how the cyclical interaction of these three components impact learning outcomes. It is my hope that this model will help researchers and practitioners ensure that SLEs truly have an impact on trainee engagement and ultimately training outcomes.

Testing Aspects of the Model

In addition to proposing the integrated SLE model, the other purpose of this study is to provide empirical support for at least some of the predictions it makes. First, I hope to demonstrate a link between user judgments and user behaviors for trainees in a synthetic learning environment. The theoretical grounding for this connection is discussed in detail by Kahn (1990). However, his discussion is focused on the impacts of meaningfulness, safety, and availability on employee investments of themselves in *work role* performances not *training* performances. Additionally, Kahn offers only qualitative evidence for the connection between these antecedents (meaningfulness, safety, and availability) and engagement. Nonetheless, subsequent research demonstrated a predictive link between Kahn's antecedents (meaningfulness, safety, and availability) and the three forms of engagement (Edmondson, 1999; May, Gilson, & Harter, 2004; Rich et al., 2010). Still, none of these studies examined this relationship in a training or SLE context.

So why should user judgments, as outlined in the integrated SLE model (consisting of meaningfulness, safety, availability, and confidence), be expected to predict physical, cognitive, and emotional engagement in training? First, meaningfulness was defined above as a sense of return on investments of one's self in training performances. Previous research indicates that training performance is partially dependent on the meaning a trainee attaches to the training (Noe & Colquitt, 2002). If the content or presentation of the learning content fails to be perceived as useful, trainees are likely to underperform. I believe that this underperformance is the result of the trainees' low level investment of physical, cognitive, and emotional resources (i.e., low engagement). Most of us have probably experienced similar circumstances in which training failed to convince us of its utility and return on our investment of time and energy. In such a case we probably did not devote much physical, cognitive, or emotional energy.

Next, safety was defined above as a sense of being able to show and employ one's self in a training context without fear of negative consequences to self-image, status, training performance, or career. Again, personal experience should serve to illustrate the link between safety and user behavior (i.e., engagement). Anyone who has felt self-conscious about asking questions, making comments, or expressing emotions during a college course or company training has experienced firsthand the impact that this can have on engagement. It is hard to be fully involved if you don't feel safe enough to participate and it's hard to perform at your best if you don't participate to the fullest.

Availability was defined above as a sense of possessing the physical, emotional, and psychological resources necessary for investing self in training performances. Numerous research studies demonstrate the importance of cognitive investments in training and work performance (Colquitt et al., 2000; Schmidt & Hunter, 2004). Caffeine addicts worldwide might attest to the importance of physical resources when it comes to training and work performance. Similarly, anyone who has attempted to focus on work while emotionally drained has had a similar experience. In each case, a lack of cognitive, physical, or emotional resources can dramatically and negatively impacts one's performance (including training performance).

Finally, confidence was defined above as one's level of confidence related to their ability to perform the training tasks. Research has demonstrated a link between trainees' opportunities to practice a task (i.e., opportunities to gain confidence) and training performance outcomes (Noe & Colquitt, 2002). Similarly, research has shown that levels of self-efficacy predict training performance (Gist, Schwoerer, & Rosen, 1989; Goldstein & Ford, 2002). Thus, evidence suggests that confidence also predicts training performance.

In summary, for each user judgment (meaningfulness, safety, availability, and confidence) evidence and logic indicates their ability to predict training performance. Furthermore, I believe that the reason each predicts performance is through its impact on physical, cognitive, and emotional engagement. However, it should also be noted that the present study makes no specific predictions regarding how each of the user judgments relate to each aspect of engagement.

H2: User judgments (meaningfulness, safety, availability, and confidence) will predict user behaviors in the form of engagement (physical, cognitive, and emotional).

Second, I hope to demonstrate a link between engagement in SLEs and training outcomes. Brown and Leigh (1996) demonstrated across multiple samples that employees who work harder also exhibit higher levels of job performance, while others have shown a relationship between engagement in one's work role and job performance has been demonstrated in the literature (Christian et al., 2011; Rich et al., 2010). Thus, it would stand to reason that those who work harder (i.e., invest more physical, cognitive, and emotional resources) in a training context would also exhibit higher levels of training performance. More specifically related to learning contexts, there is evidence that other conceptualizations of engagement have a direct link to training performance. For example, when engagement is defined in a behavioral sense (i.e., participation, time on task, etc.) there is evidence of a strong link to learning outcomes (Kuh, Kinzie, Schuh, & Whitt, 2011). Similarly, research has shown links between intrinsic motivation (one component of engagement) and learning (Cordova & Lepper, 1996). However, to this point research exploring the link between subjective reports of engagement and actual learning outcomes has been sparse (Whitton & Moseley, 2014). Thus the present study will further contribute to the literature by exploring the relationship between trainees self-reports of engagement and training outcomes.

H3: Engagement (the combination of physical, cognitive, and emotional) will predict training outcomes.

Third, I will explore the comparative levels of engagement between those in a SLE and those in a control condition. It is clear that simply being in a SLE does not automatically result in higher levels of engagement for all trainees (Whitton, 2011b; Whitton & Moseley, 2014). However, as discussed above, SLEs that lead to user judgments of meaningfulness, safety, availability, and confidence are likely to trigger user behaviors of engagement through their impact on user judgments. So, while there are many individual differences that could impact a particular trainee's tendency to be engaged in a SLE (e.g., technological savvy, experience with videogames, preferences for game styles, perceptions of learning using games etc.) (Ricci et al., 1996), I contend that when designed with forethought to correct principles of training and based on solid game characteristics (e.g., fantasy, sensory stimuli, mystery, control etc.) SLEs have an advantage over other forms of training to positively impact perceptions of meaningfulness, safety, availability, and confidence. These perceptions then influence trainee engagement. In other words among trainees who are at least somewhat technologically savvy and open to the idea of SLEs, effectively executed game characteristics have the potential to impact meaningfulness, safety, availability, and confidence above and beyond that which is possible given just the instructional content. This then impacts engagement in all its forms. Research has already shown that well-designed SLEs can result in high levels of intrinsic motivation and reports of enjoyment and interest among learners (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Druckman, 1995). Similarly, others have shown that those in SLEs have a tendency to report higher levels of intrinsic motivation and enjoyment than their control condition counterparts (Cordova & Lepper, 1996; Ricci et al., 1996). Thus, assuming that the sample of trainees is somewhat technologically savvy, it would not be unreasonable to expect those in a SLE to report higher levels of engagement than those in a control condition.

H4: Those in the SLE will report higher levels of engagement (the combination of physical, cognitive, and emotional) than those in a control training condition.

Finally, I propose that the relationship between SLEs and learner outcomes is partially mediated by learner engagement. As described above, organizations have become interested in videogame training because they hope to harness their “motivational” characteristics which we have redefined as engagement (Jana, 2006). This objective has even been referred to as the “holy grail” of trainers (Garris et al., 2002). It seems that both practitioners and scientists believe that the assumed differences in engagement between SLEs and comparison conditions are the driving force behind their success. This assumption is echoed by other researchers who believe that engagement explains the relationship between game characteristics and learning outcomes (Filsecker & Kerres, 2014; Whitton, 2011a). This relationship makes logical sense if one accepts hypotheses 4 above. If on average well-designed SLEs result in higher levels of engagement with the training material and this engagement leads to higher training performance, it follows that the impact that SLEs have on learning outcomes is at least partially explained by their influence on trainee engagement. Having said this, it is not reasonable to expect engagement to explain the entire relationship since other factors (e.g., active participation, access to the training etc.) that are common in SLEs likely have a large impact (Sitzmann, 2011). However, it is likely that engagement will explain a significant portion of the relationship (Ricci et al., 1996). Till now, this mediating relationship has not been empirically tested.

H5: Trainee engagement will partially mediate the relationship between training condition and trainee performance.

In conclusion, I believe that the model outlined by Garris et al. (2002) is essentially correct but that modifications to the game cycle using Kahn's theory of engagement can provide a more solid theoretical foundation for the links within. This extension of the model should make it easier to test and may eventually provide researchers and practitioners with a framework for understanding how to create engagement in SLEs and how to ensure that engagement leads to eventual training outcomes. As a first step, I wish to provide some support for this model by testing a few of its key relationships. I will do this through data collected as part of two studies. Study One will be used to examine the link between user judgments and user behaviors (engagement) and will be based on a sample of undergraduate students playing a computer game hosted on the Steam game network. Study 2 will be used to study the effect of SLE characteristics on user behaviors (engagement), examine the links between user behaviors (engagement) and training outcomes, and test the mediating effect of engagement for the relationship between the model's input (specifically SLE characteristics) and outcomes. Study

Two will be based on a sample of amputee trainees in one of two conditions, a virtual world condition hosted in Second Life, or a more traditional e-learning condition hosted online.

Method: Study One

Study One was used to address Hypothesis 2 regarding the link between user judgments (meaningfulness, safety, availability, and confidence) and engagement (physical, cognitive, and emotional). Data for Study One were collected as part of a larger study comparing multiple conditions of an educational game. An observed predictive relationship between participants' scores on meaningfulness, safety, confidence, and availability and engagement would provide support for Hypothesis 2 and the link between user judgments and user behaviors as conceptualized in the revised input-process-output game model described above.

Participants

Participants were 106 students who were either recruited from a psychology department research participant pool or an introductory management course of a large western university. Participants received 1.5 hours of research credit as compensation for their time and effort. Participants were between 18 and 60 years old (M_{age} = 20.8; SD = 4.4), mostly male (65%), white/Caucasian (76%), and in their first year of college (41%). Average rated experience with videogames was 2.8 out of 5 with a standard deviation of 1.1 indicating that the majority of participants had at least some experience with videogames.

Procedure

The data analyzed here were part of a larger study involving comparisons among multiple SLE conditions. However, I only wished to demonstrate a predictive relationship between trainees' perceptions of meaningfulness, safety, availability, and confidence and their reports of physical, cognitive, and emotional engagement. Thus, Study One involved administering each of these measures to a group of SLE trainees. Because there was no reason to expect this relationship to be moderated by condition, I examined the data collapsing across SLE conditions.

Participants in this study were required to participate in the educational game and complete all measures in a laboratory with one facilitator and up to two fellow game players (in separate rooms). All communication between players was done via a headset. Prior to gameplay, participants read a short instructional manual. Participants then engaged in a collaborative educational videogame known as Moonbase Alpha. This game is a science and technology educational game developed and published by NASA in 2010 and available on the Steam gaming network. During the game, players act as astronauts at a space station based on the moon. Gameplay revolves around the need to respond to a meteor impact that has just disabled the oxygen supply system for their moon base. Players must navigate the map to locate damaged or destroyed equipment, and repair or replace the equipment within the given time limit using an assortment of tools.

Participants completed each measure immediately after playing the game. All responses were collected using Qualtrics survey software.

Measures.

Measures for this study are provided in the appendix. To measure meaningfulness, safety, availability, and confidence I developed a measure based on the theoretical work of Kahn (1990) and a measure developed by May et al. (2004). May et al.'s scale was originally developed to explore the link between meaningfulness, safety, and availability and psychological engagement in one's work role. Adaptations of many of the items were required to change the focus from the work role to the training role. Some items were removed due to their inapplicability to the context of a SLE. Also, a number of items were added to each subscale to better capture each constructs as explained in the sections above. Most notably, modifications and additions were made to account for the influences of the components of the game cycle (Garris et al., 2002). For example, items related to enjoyment or entertainment were added to the measure of meaningfulness to account for the combination of Garris et al. and Kahn's conceptualizations of the construct.

To test the dimensionality of this modified measure I conducted a confirmatory factor analysis (CFA) on all 24 items representing each of the 4 types of user judgments. The initial CFA revealed relatively weak loadings for items 1 and 4 on the meaningfulness factor ($\lambda = .30$, $\lambda = .18$ respectively). Before removing items 1 and 4 fit indices were χ^2 [246] = 519.63, CFI = .88, TLI =. 87, RMSEA = .08. After removing these items the fit indices improved to a reasonable level χ^2 [203] = 423.36, CFI = .91, TLI =. 89, RMSEA = .08. I tested a single factor model which indicated poor fit χ^2 [209] = 1097.82, CFI = .63, TLI =.58, RMSEA = .16 and compared it to the four factor solution which fit the data significantly better than the one factor alternative $\Delta\chi^2$[6] = 6, $p < .001$. Thus, the final user judgment model consisted of 22 items; 11 items assessing meaningfulness, four items assessing psychological safety, four items assessing confidence, and three items assessing psychological availability. Internal consistency reliabilities for each subscale were within the acceptable range with $\alpha = .95$ for meaningfulness, $\alpha = .79$ for safety, $\alpha = .92$ for confidence, and $\alpha = .77$ for availability.

To measure engagement I used an adaptation of the Rich et al. (2010) employee engagement measure. This measure assesses the physical, emotional, and cognitive components of engagement proposed by Kahn (1990). My adapted measure maintains both the 18 original items and the Likert-type rating scale (*1 = strongly disagree 5 = strongly agree*) used for the original measure. Modifications included only minor wording changes to make training the focus of each item as opposed to one's job or work. The following represents sample items from the physical, emotional, and cognitive sub-scales respectively: "I worked with intensity on this training;” “I was enthusiastic in this training;” “During this training, I was absorbed by the material.” In the present study, the internal consistency reliabilities for each dimension were similar to those reported by Rich et al., (2010) with $\alpha = .92$ for physical engagement, $\alpha = .93$ for cognitive engagement, $\alpha = .92$ for emotional engagement and $\alpha = .94$ for the overall engagement scale.

Results: Study One

Study One sought to address Hypothesis 2 regarding the predictive link between user judgments (meaningfulness, safety, availability and confidence) and engagement (physical, cognitive, and emotional). All three engagement sub-factors were collapsed because no theories or predictions were made regarding the interrelationships of these sub-factors and each of the user judgements. Table 2 shows means, standard deviations for each study variable, as well as correlations among variables. Correlations show that meaningfulness, confidence, and availability were significantly correlated with engagement. Unexpectedly, safety was not correlated with engagement at all. Furthermore, there was a relatively large correlation between confidence and meaningfulness.

To address Hypothesis 2, a standard multiple regression was performed with overall trainee engagement regressed on the four user judgment variables. First though, an evaluation of assumptions led to the removal of five univariate outliers. All other assumptions including normality, linearity, homoscedasticity, and independence of errors were checked and found to be within commonly established limits (Field, 2013).

The regression analysis revealed that user judgments do predict overall trainee engagement $R^2 = .56$, F (1, 105) = 31.74, $p < .001$. This was a substantial R^2 value suggesting that over half the variance in trainee engagement can be accounted for by user judgments. However, it seems that only two of the user judgments (psychological meaningfulness and availability) were significant predictors of trainee engagement.

Table 3 shows the raw regression coefficients, standard errors, standardized beta coefficients, p-values, and semi-partial correlations for each of the four user judgments. Contrary to expectations, neither psychological safety nor confidence was found to be a significant predictor of engagement. However, it is interesting to note that while the bivariate and semi-partial correlations between engagement and both psychological meaningfulness and availability are quite similar, the bivariate and semi-partial correlations between engagement and confidence are drastically different. This seems to indicate that although confidence may be a useful predictor of trainee engagement, it does not seem to account for variance above and beyond that already captured by meaningfulness and availability. In reality, the relatively high correlation between meaningfulness and confidence (r = .48) would suggest that meaningfulness likely captures all relevant variance in engagement that could be accounted for by confidence.

A notable limitation of the data collected in Study One is its reliance on a university student sample. Student samples such as these may not accurately represent the diversity in age, race, educational background or level that is inherent in most workplaces (Sears, 1986). Furthermore, such a sample is more likely to display a feeling of apathy toward the training content since it is in no way related to their work as students or otherwise. Both of these issues can be problematic for training research meant to be applied to a diverse workplace in which trainees are generally motivated to learn and improve their skills related to their job.

In sum, Study One found that user judgments do predict user behaviors (i.e. engagement). However, only meaningfulness and availability act as significant predictors of engagement. Safety does not seem to be related to engagement at all. On the other hand, confidence is strongly related to engagement but does not seem to account for unique variance beyond that already captured by meaningfulness and availability.

In an attempt to overcome the issues discussed above, Study Two relies on a more mature sample of amputees completing a personally relevant training program. Although the sample size for Study Two is very small, there are many characteristics that make it an appropriate and particularly attractive sample. First, the sample is composed of more mature individuals than the traditional university sample – overcoming some of the limitations of undergraduate samples discussed by Sears (1986) (e.g. less-crystallized attitudes, stronger tendencies to comply with authority, and more unstable peer group relationships). Second, the training content is particularly relevant to the participants in that it provides information and skills directly related to a salient issue in their lives. Finally, participants in Study Two were recruited from all over the United States making the sample a much better representation of the population to which I wish to generalize.

Methods: Study Two

Study Two was used to address Hypotheses 1, 3, 4, and 5. Hypothesis 1 examined the comparative levels of motivation to learn between a SLE and control group. Hypothesis 3 examined the comparative levels of engagement between an SLE and control training group. Hypothesis 4 examined the ability of training engagement to predict training outcomes, and Hypothesis 5 examined the mediating function of engagement.

Data for this study were collected as part of a lager study entitled “Dissemination of Amputee and Prosthetic Evidence-based Medicine (DAP-EM)” which was funded by the Agency for Healthcare and Research Quality. This was a three-year study aimed at determining the optimal methods to disseminate physical, emotional, technology-based, evidence-based health information to individuals with upper and/or lower limb amputation(s). This study compared a traditional e-learning format to the use of a virtual world for training participants on topics including amputation, comorbidity management, physical and psychosocial aspects of living with an amputation, pain interference, choosing a prosthesis, and other prosthetic training. To ensure the quality of both conditions, the principal investigator assembled a team of subject matter experts, a professional e-learning programmer, and professionals in virtual world programing to develop the training materials.

Participants

Participants were recruited via flyers, amputee-related website postings, and word-of-mouth communications from the study’s principal investigator to various prosthetists, M.D.s, physical and occupational therapists, and prosthetic vendors all of whom alerted amputees that might be eligible for participation. To qualify, individuals had to have at least one major amputation (upper limb or lower limb) not including finger or toe amputations. In all 37 individuals were enrolled in the study, however, many did not complete either the e-learning or virtual world training. Ultimately, 22 participants completed the entire training with 14 in the e-learning control condition and eight in the virtual world condition. Participants in either condition received $75 as compensation for completing the intervention, and post-tests.

Participants were between 31 and 75 years old (M_{age} = 50.6; SD = 12.0), mostly male (66%), and white/Caucasian (95%). General control variables are reported in Table 4 and show that participants were quite comfortable with technology, spent many hours on the computer each week, played few videogames, and were relatively unfamiliar with virtual worlds. In fact, of the entire sample only one participant reported spending any time in virtual world environments. Thus, even though the sample was relatively comfortable with technology, they were not very experienced with regard to videogames or virtual worlds.

Training Conditions

Both training conditions provided essentially identical content covering a wide range of topics and skills applicable to those with upper or lower limb amputations. Content included information about prosthetic technology, amputation, physical and psychosocial aspects of living with an amputation, and skills related to managing secondary conditions such as pain and skin conditions.

Those in the e-learning group served as the control for the study. The materials for this condition were designed and developed by e-learning professionals to ensure a high standard of quality. Control group participants completed a series of three modules which were accessed online. Module one provided information about amputation and took about one to two hours to complete. Modules two and three presented short video clips of amputees discussing their experiences (e.g., selecting prostheses, caring for skin or other secondary conditions, managing pain etc.). Modules two and three took anywhere from one to three hours to complete. Each module featured a combination of pictures, videos, text, and external web links through which participants navigated using buttons and menus (See Figure 4).

In contrast to the control condition, participants in the SLE condition were given access to the virtual health adventures island within Second Life (Figure 5). Second Life is one of the most commonly-used virtual worlds for training (Heiphetz & Woodill, 2010) and allows paying users (such as researchers) to buy virtual real-estate. This real-estate can then be customized for the needs of the user.

In this case, the virtual health adventures island was developed as a training location for study participants. The island served as a virtual world training and support space where amputees could connect with other avatars (those of fellow amputees, caregivers, and friends or loved ones). The island provided amputees with opportunities to connect with fellow amputees and to explore their sense of post-amputation using a 3D representation of themselves. To access the island participants were required to have an internet connection and computer capable of rendering the environment. Just as learners in the control group could access the e-learning content at any time, access to the island was unrestricted, allowing participants to explore the learning content freely, any time of the day or night.

Learning content was located at 17 stations across the island and was identical to the content provided in the control condition. For example, one station featured a museum of prosthetics where participants could learn about the history of prosthetics while viewing 3D representations of them in the museum. Other sources of information included video demonstrations of how to put on prosthesis or how to get dressed (the same videos presented to control participants). Participant avatars were customizable in term of everything from body type to hair, eyes, facial features, and clothes. Additionally, avatars were encouraged to not only access the learning content but practice skills (e.g., dressing etc.) using their avatars, and to interact with other avatars using the same space.

Participants in the Second Life condition began their experience with a “buddy” who helped with orientation to the environment by teaching them how to move, interact with objects and other avatars, and how to cloth and modify their avatars. Those in the Second Life condition were required to participate for a period of at least one month. Participants were asked to visit all 17 stations but were given no other guidelines regarding the time they were supposed to spend in Second Life or how to spend that time.

Procedure

A randomized control design was employed to compare the e-learning control group to the virtual world/SLE group on four outcomes of interest: self-efficacy, psychosocial status, pain interference, and functional status. Hypothesis 3 stated that engagement would be positively related to training outcomes. Because data were collected as part of a larger study, I was limited to training outcomes of interest to the primary researchers. Below I describe each of these four outcomes in more detail, including how they were assessed. Note here though that trainee learning should have a positive impact on each of the four. Each of these outcomes was measured within a few days before (pretest) and after (posttest) the training intervention (thus change scores were available and used for each analysis). Individuals in both the control and SLE conditions received information and training on the same content and were asked to complete all of the same measures. Demographics were measured prior to the intervention as part of the pretest. Engagement was measured after training as part of the post-test assessment

Measures

Demographic and control variables including age, gender, level of education, level of limb loss, and current frequency of prosthetic use were collected prior to the training interventions. Training engagement was measured immediately following completion of the training program and was assessed using the same adaptation of Rich et al., (2010) that was employed in Study One. Internal consistency reliability for the three dimensions for this study were all within an acceptable range with α = .84 for physical engagement, α = .96 for cognitive engagement, α = .94 for emotional engagement and α = .96 for the overall scale.

Motivation to learn. Motivation to learn was assessed using an adaptation of Noe and Schmidt's (1986) motivation to learn measure. The modifications simply changed the focal point of the items to make sense in the current situation. For example, the item "I will try to learn as much as I can from Springfield" was changed to "I tried to learn as much as I could from the training." This measure consisted of eight items rated on a five-point rating scale (*1 = strongly disagree, 5 = strongly agree*). Thus, score totals may range from 8 – 40 with higher scores indicating higher motivation to learn. Sample items include: "I was motivated to learn the information presented in the training program" and "the reason I stuck with the training program was because I wanted to learn how to improve my knowledge and skills related to prosthetics." Internal consistency for this scale was similar to that reported by Noe and Schmitt (1986) with α = .85.

Self-efficacy. The first study outcome was self-efficacy consisting of general self-efficacy and self-efficacy to manage secondary conditions. Trainees who learn most from the training should be expected to report the highest levels of both general self-efficacy and self-efficacy to manage secondary conditions. General self-efficacy was measured using the General

Self-Efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995). The GSE is a 10-item scale that assesses the strength of an individual's belief in his or her own ability to respond to novel or difficult situations and to deal with any associated obstacles or setbacks (e.g. secondary condition related to amputation). The GSE uses a four point Likert-type rating scale where participants rate the truthfulness of 10 statements (*1 = not true at all and 4 = exactly true*). Scores can range from 10 – 40 with a higher score indicating greater self-efficacy. A sample items is "I can solve most problems if I invest the necessary effort." Internal consistency reliability for this scale was good with α = .82 for both pre and post measures.

Portions of the Stanford University School of Medicine Patient Education Self-efficacy Scales were used to measure self-efficacy to manage secondary conditions (Lorig, Sobel, Ritter, Laurent, & Hobbs, 2001; Lorig et al., 1996; Schwarzer & Jerusalem, 1995). Six items from the Stanford Scales were adapted for use in this study (3 items from each). Slight wording changes were made to 3 of the items to either simplify wording or focus the content on amputation specifically. The Stanford measures use a 10 point Likert-type scale to rate confidence in each of the six statements (*1 = not at all confident and 10 = totally confident*). Thus, total scores can range from 6-60, with a higher score indicating greater self-efficacy. A sample item is "How confident are you that you can keep the conditions associated with amputation from interfering with the things you want to do?" Internal consistency reliability was acceptable at α = .77 pre-training and good at α = .94 post-training.

Psychosocial Health. The second outcome is psychosocial health and includes self-reports of both perceived social support and psychosocial adjustment. In both cases, trainees who learn from training and experience a subsequent increase in prosthetic use should be expected to report higher levels of perceived social support and psychosocial adjustment (Gallagher &

Maclachlan, 2004). To measure social support the Multidimensional Scale of Perceived Social Support (MSPSS) was used (Zimet, Dahlem, Zimet, & Farley, 1988). This scale is designed to measure perceptions of support from three sources: family, friends, and a significant others. The scale is comprised of a total of 12 items, with four items for each subscale all rated on a seven-point Likert-type scale with *1 = very strongly disagree to 7 = very strongly agree*. Thus scores range from 12 to 84 with higher scores representing more support. A sample item is “there is a special person who is around when I am in need.” Internal consistency reliability for this scale was good with α = .92 pre-training and α = .95 post-training.

A subscale from the Trinity Amputation and Prosthesis Experience Scales, Revised (TAPES-r) was used to measure psychosocial adjustment (Gallagher & MacLachlan, 2000). This 15 item subscale provides a measure of general psychosocial adjustment for amputees using prosthesis. This measure uses four-point Likert-type scale to rate agreement with each of the 15 statements (*1 = strongly disagree and 4 = strongly agree*). Thus, score totals may range from 15 to 60 points with higher scores representing better adjustment. A sample item is “I have adjusted to having an artificial limb.” The TAPES-r has been shown to be a psychometrically sound measure with high internal consistency reliability (Gallagher & MacLachlan, 2000). In this study, internal consistency reliability for the TAPES-r was good for both pre and post training measurements with α = .92 and α = .90 respectively.

Pain Interference. The third outcome measure was pain interference which was measured using a subset of items from the West Haven-Yale Multidimensional Pain Inventory (WHYMPI) (Kerns, Turk, & Rudy, 1985). Trainees that perform better on the training and experience higher levels of social support are likely to experience lower levels of pain interference (Williams et al., 2004). The WHYMPI is a well-known and widely used measure of chronic pain. Nine items were selected from the interference subscale and three items from the affective distress subscale. Thus, a total of 12 items were used each rated on a six-point Likert-type scale with various anchors (see appendix). Score averages ranged from 0 to 6 with higher average scores signaling more pain interference. A sample item is “in general, how much does your pain problem interfere with your day to day activities?” each item is rated from *0 = no interference to 6 = extreme interference*. Good internal consistency reliability for the combination of these two subscales was found for both pre and post-training measurements with $\alpha = .96$ and $\alpha = .97$ respectively.

Functional Status. The fourth outcome measure was functional status which was measured by using portions of the Orthotics Prosthetics User’s Survey (OPUS) (Heinemann, Bode, & O’Reilly, 2003). Trainees who learn more in the training should report higher levels of functional status due to the skills and knowledge acquired in training regarding how to perform activities of daily living (ADLs). The OPUS is a self-report measure consisting of five modules designed to assess common rehabilitation goals in prosthetic and orthotic practice: Lower Extremity Functional Status (LEFS), Upper Extremity Functional Status (UEFS), Client Satisfaction with Device (CSD), Client Satisfaction with Services (CSS), and, Health-Related Quality of Life (HRQoL). For the purposes of this study only the LEFS and UEFS modules were used. Both modules assess one’s functional abilities in relation to actions that are commonly performed in everyday life. Activities are rated using a four-point Likert-type rating scale (*1 = very easy and 4 = cannot do this activity*). The LEFS consists of 20 items such as “walk around indoors” and “walk up a steep ramp.” The UEFS consists of 28 items such as “open door with knob” and “wash face.” The LEFS and UEFS have demonstrated good psychometric properties in the past (Burger, Franchignoni, Heinemann, Kotnik, & Giordano, 2008; Jarl, Heinemann, &

Norling, 2003) and exhibited high internal consistency reliabilities in the present study with α = .99 for the UEFS and α = .98 for the LEFS on both the pre and post-training measurements.

Results: Study Two

In preparation for the analyses, I closely examined the data for univariate outliers. Three potentially problematic cases were identified based on boxplots, histograms, and standardized scores greater than three. These cases were subsequently removed resulting in a final sample size of 19. Unfortunately, the small sample size in the present study resulted in largely non-normal distributions for most of the study variables. Log and square root transformations were attempted but did not seem to improve normality. Thus, in the subsequent analyses I rely on non-parametric tests when possible to obviate the need to meet many of the basic assumptions of parametric tests including normality.

Overall means, standard deviations, and correlations among the main study variables are displayed in Table 5. It should be noted that there are six outcome measures because both self-efficacy and psychosocial health were measured with two scales each. Means and standard deviations for each of the six training outcome variables are based on change scores representing the difference between post-tests and pre-tests. It is interesting to note that while significant improvements in psychosocial adjustment were observed from pre to post measurement (as made evident by a relatively large positive value of 3.8) the other training outcomes actually decreased or remained relatively unchanged from pre- to post-measurement.

The correlations in Table 5 show that engagement and motivation to learn were significantly correlated as were functional status and psychosocial adjustment. Contrary to expectations no other correlations were significant. Particularly surprising are the small correlations between trainee engagement and each of the training outcome variables with the exception of social support which was actually negatively correlated with engagement. Overall these results do not seem to support the study's hypotheses.

Hypothesis 1 stated that there would be no significant differences in trainee motivation to learn when comparing a SLE training group to a control group. Although there are arguments against testing hypotheses of no effect, Cortina and Folger (1998) suggested it may be a useful pursuit. To test Hypothesis 1, I performed a t-test and computed multiple effect size estimates. Results showed that, consistent with my hypothesis, there was no significant difference between the SLE and control group on motivation to learn t (17) = -1.89, p = .08, BCa 95% CI [-.927, .050]. Nevertheless, the effect size estimate was relatively large favoring the e-learning condition (Cohen's d = .84). However, it seems that the large effect size estimate may simply be an artifact of having very little variance in the motivation to learn variable (σ^2 = .27). For example, the numerator of the equation for Cohen's d is the difference between the two means being compared, while the denominator is the pooled standard deviation. In other words, the smaller the variance, the larger Cohen's d. Thus, in the present case the unduly small variance may have inflated the estimate of Cohen's d. In reality the small variance in this variable is not surprising given the motivation that might be expected in a sample of amputees in and amputee related training.

In an attempt to further explore Hypothesis 1 I also performed a series of non-parametric tests to examine the differences between conditions. First, a Mann-Whitney test can be particularly useful for small samples that are unable to meet some of the assumptions inherent in parametric tests such as the t-test. In the present case, my inability to achieve normality suggests that such a non-parametric test might be warranted. Again, consistent with my hypothesis, the Mann-Whitney test revealed no significant differences between the two groups U = 61.00, z = 1.65, p = .100. To further confirm my finding, I performed Kolmogorov-Smirnov Z test, which some have suggested to be a more powerful test when dealing with samples smaller than 25.

Results again confirmed that there was no significant difference in motivation to learn between the two training groups *D* = 1.10, *p* = .177. In conclusion, results indicated no significant differences between the two training groups in terms of motivation to learn. Thus, hypothesis 1 was supported

Hypothesis 3 stated that engagement (the combination of physical, cognitive, and emotional) would predict each of the six training outcomes (general self-efficacy, self-efficacy to manage secondary conditions, social support, psychosocial adjustment, pain interference, and physical function). To test this hypothesis I first calculated difference scores (pre to post) for each of the outcome variables. I then employed simple linear regression which allowed me to obtain estimates of R^2 for each of the outcome variables. Contrary to expectations, none of the six separate regressions revealed a predictive relationship between engagement and training the outcomes. Results for each of the regressions, including the F test and significance value, unstandardized coefficient, and the t and corresponding significance value are reported in Table 6. Though contrary to the study hypotheses, these results are not surprising given the small, non-significant correlations between engagement and each of the outcome variables as reported in Table 5 above.

Because the reliability of difference scores have been criticized as being less reliable than using each score alone, I decided to also run a repeated measures MANOVA to further explore hypothesis 3. Results confirmed the findings reported above. There was no effect for time on the combination of outcome variables, Wilks Lambda = .88, *F* (5,16) = .42, *p* = .827, meaning that outcomes did not change significantly from time one to time two. Additionally there was no time by engagement interaction, Wilks Lambda = .83, *F* (5,16) = .62, *p* = .688. In all, these results confirm that there was no change in any of the outcome variables from pre to post measurement and ultimately engagement is not a significant predictor of these outcomes.

Hypothesis 4 stated that those in the SLE would report higher levels of engagement (the combination of physical, cognitive, and emotional) than those in a control training condition. To test this hypothesis an independent samples t-test with bootstrapped bias corrected accelerated (BCa) confidence intervals was employed. Bootstrapping was used to reduce the impact of potential biases, including non-normality of the study variables. Results showed that participants in the Second Life training group reported higher levels of engagement in the training (M = 4.6, SD = .3), than those in the e-learning/control training group (M = 4.1, SD = .7). This difference, -.56, BCa 95% CI [-1.15, 0.02] was marginally significant t (17) = -2.02, p = .060 and represented a large effect d = 1.04. However, it should be noted that the variance was also relatively small for engagement (σ^2 = .40) which, similar to the case described above, may contribute to the large effect size.

Because of the relatively non-normal distribution of engagement, I decided to also perform a Mann-Whitney test to examine the differences in engagement between the two groups. The Mann-Whitney U test showed that indeed there is a significant difference between the two conditions in terms of their overall level of engagement U = 66, z = 2.03, p = .045. This result corresponds with an r = .47 representing a medium effect size. Furthermore, it is important to note that this effect size estimate is not biased by the restricted variance mentioned above since the Mann-Whitney test is non-parametric and its r is estimated from a z score produced from the ranking process. All in all, these results provide support for Hypothesis 4 and indicate the trainees in the Second Life condition experienced significantly higher levels of engagement than their control (e-learning) condition counterparts.

Finally, Hypothesis 5 stated that trainee engagement would partially mediate the relationship between training condition and trainee performance. A popular approach to mediation analysis is the Barron and Kenny method (Barron & Kenny, 1986). However, for the present study, I planned to use the PROCESS macro developed by Hayes and Preacher (Hayes, 2013). This method/macro is preferred to the traditional Barron and Kenny approach for a number of reasons. First, the Barron and Kenny approach infers mediation based on multiple regression analyses which can inflate the type 1 error rate. Second, their approach relies on a set of criteria which do not always hold. For example, research has shown that it is possible to observe a change from a significant X to Y path to a non-significant X to Y path after adding the mediator (which generally indicates mediation) without observing a substantial change in the absolute size of the X to Y coefficient. Third, the Barron and Kenny method offers no formal significance test for the indirect effect (although the Sobel method may be used for this purpose (Sobel, 1982). Finally, in some cases their method may suffer from low statistical power. In contrast, to the Barron and Kenny approach, the PROCESS method/macro overcomes the issues just mentioned be reducing the chance of a type 1 error, offering a formal significance test for the indirect effect, and demonstrating higher statistical power. Instead of relying on multiple regression analyses and subjective comparisons of regression coefficients (as the Barron and Kenny approach does), the PROCESS method/macro relies on bootstrapping to provide point estimates and confidence intervals which are used to assess the significance of the mediation effect. The fact that the PROCESS method/macro uses bootstrapping is particularly beneficial when considering mediation analyses for small sample sizes – as is the case in the present study.

Having said this, one of the basic requirements for a mediation model (both the Barron & Kenny approach and PROCESS) is the existence of a significant relationship between the predictor and the dependent variable(s). For the present hypothesis, condition serves as the predictor variable with the six outcome measures serving as the dependent variables and engagement as the mediator. Unfortunately many of the basic assumptions for mediation have not been met with the current data set. First, as reported in Table 5, condition is not related to any of the dependent variables. Furthermore, engagement is not related to any of the outcome variables either. The violations of these two basic assumptions for mediation render any mediation analysis meaningless. Thus, I will perform no such analysis and conclude that there is no empirical support for the mediating effect of engagement.

Discussion

SLEs have been publicized as an effective method for organizations to train employees. Interestingly, across years and numerous publications, the ability of SLEs to produce learning outcomes has been primarily attributed to their motivational capabilities (Whitton, 2011). However, the few empirical studies that have researched motivation in SLEs have failed to find any differences between SLEs and control groups in terms of overall trainee motivation (DeRouin-Jessen, 2008; Parchman et al., 2000). This may be due to the literature's failure to clearly define and measure "motivation." One primary purpose of the present study was to demonstrate that the SLE literature should not be concerned with motivation to learn or intrinsic motivation alone, but should be concerned instead with trainee engagement (specifically the combination of physical, cognitive, and emotional engagement). Another primary purpose of the present study was to propose and test portions of the integrated SLE model. This model suggested that SLEs lead to training outcomes via the SLE cycle which consists of user judgments leading to user behaviors which lead to system feedback and finally a repeated cycle.

In response to these two main purposes, I sought to answer five hypotheses. Hypothesis One stated that there would be no significant differences in trainee motivation to learn when comparing a SLE training group to an e-learning control group. In direct contrast to previous studies conceptualizing motivation as motivation to learn, I found no evidence for differences between the SLE and control groups. While this no-effect hypothesis could not be formally tested, the non-significant t-tests, including a bootstrapped confidence interval including 0, and a non-significant Mann-Whitney test provide reasonable support for the first hypothesis. This finding suggests that it is inappropriate to conceptualize the "motivation" that trainees experience from videogame training, as motivation to learn. Indeed, my results would suggest that motivation to learn is not impacted by the training medium. This aligns with previous research indicating that motivation to learn may be impacted by individual characteristics (e.g. locus of control, conscientiousness, cognitive ability), job characteristics (job involvement, organizational commitment), and even situational characteristics (Colquitt et al., 2000), but is likely unaffected by training media. In the present study it seems that a trainee's desire to learn the content of the training program (i.e. motivation to learn) was unaffected by the media used to present the training material. Consequently, future research on SLEs should avoid making comparisons between SLE and control groups in terms of their motivation to learn.

Hypothesis Two stated that user judgments (meaningfulness, safety, availability, and confidence) would predict user behaviors (the combination of physical, cognitive, and emotional engagement). Results indicated that this was the case. Specifically, the combination of meaningfulness, safety, availability, and confidence predicted 56% of the variance in overall engagement. However, it is interesting to note that only the coefficients for meaningfulness, and availability were significant. Safety and confidence on the other hand, were not found to be meaningful predictors. Interestingly, safety was nearly perfectly uncorrelated with engagement. This finding is inconsistent with Kahn's (1990) theory of engagement which defines psychological safety as "a sense of being able to show and employ one's self without fear of negative consequence to self-image, status, or career.

So why didn't safety predict engagement in my study? According to Kahn, there are four main influences on safety: interpersonal relationships, group and intergroup dynamics, management style and processes, and organizational norms. Supportive and trusting interpersonal relationships allow individuals to try and perhaps fail without fearing consequences, and as a result have the ability to promote engagement. Similarly, group dynamics that encourage the expression of individual's voices, and a relatively equal distribution of authority and power across group members, can promote engagement. Engagement can also be promoted by a supportive, resilient, and clarifying management style that creates a perception that is okay to try and fail. Finally, organizational norms can impact individual perceptions of safety in that when they are followed (and one feels within the normal boundaries set by the group) that person feels safer than when outside of those norms. Thus, understanding the organizational norms set by one's group and feeling within those bounds can increase perceptions of safety and consequently impact levels of engagement. In the present study however, it is important to note that participants played the learning game for a relatively short period of time (25 minutes). In this short period it is unlikely that the learning groups formed any of the requisite interpersonal relationships, group or intergroup dynamics, management styles or process, or organizational norms. In other words, each of the primary cues that impact one's perceptions of psychological safety were either minimally developed (interpersonal relationships, group dynamics, and organizational norms) or were completely non-existent (intergroup dynamics, and management styles). For this reason, it may not be terribly surprising that safety was unrelated to reports of engagement. In retrospect, it may not have been appropriate to include the safety subscale in this study. Future researchers planning to employ measures of user judgments should carefully consider the applicability of the safety sub-scale given the trainee's opportunities to develop and interpret the major cues related to psychological safety.

As mentioned above, results from Study One also indicated that confidence was not a meaningful predictor of engagement. However, upon examining the zero-order and semi-partial correlations in Table 1, it is apparent that although confidence was significantly related to engagement, its unique predictive contribution beyond that already captured by meaningfulness and availability was inconsequential. In other words, it seems that once meaningfulness and availability were accounted for there was little to no variance left for confidence to explain. Assumptions of multicolinearity were checked and were found to be within normal limits based on generally accepted rules of thumb which suggest researchers investigate tolerance values greater than .2 and VIF values greater than 10 (Field, 2013). However, another rule of thumb suggests that multicolinearity may be an issue when the average of all VIF values is greater than one. This was the case in the present dataset with an average VIF of 1.31 and an individual VIF of 1.55 for confidence (the largest VIF of all variables in the model). Thus, both the VIF values and the correlation of r = .48 between meaningfulness and confidence may be evidence that multicolinearity was an issue. However, it is important to note that the CFA results for the user judgment scale did not indicate serious problems with multicolinearity based on the between factor correlations. For example, confidence correlated with meaningfulness at $r = .52$ and with availability at $r = .31$. Neither correlation is particularly large given that each sub-scale has been designed to measure trainee judgments related to a training intervention. In the end, I believe multicolinearity was probably not an issue in the present study. However, it is clear that confidence does not explain variance that is not already captured by the combination of meaningfulness and availability.

In hindsight, this may actually make logical sense when considering the items used in each sub-scale. For example, an item from the meaningfulness sub-scale is “My time playing this game was NOT well spent” (reverse coded) while an item from the confidence sub-scale was “I felt confident in my ability to deal with problems that arose during the game.” If a trainee feels that their time is being well spent it is very likely they also feel they are gaining some useful skills and increasing their ability to deal with problems in the training. Future, research should continue to explore the relationship between confidence and meaningfulness. As this was an early attempt at the development of a user judgment scale for use in SLE contexts, further development and refinement of the sub-scales is surely warranted.

Next, contrary to Hypothesis 3, I found no evidence that engagement (the combination of physical, cognitive, and emotional) predicts any of the training outcomes (self-efficacy, psychosocial status, pain interference, physical function). Although these findings were surprising, there are a number of reasons why I might have obtained such results. First, it may be the case that training engagement is truly unrelated to training outcomes. In other words, one's reports of physical, cognitive, and emotional investments in the training may be relatively unimportant compared to all the other variables that might influence training outcomes such as the quality of the training, pre-training self-efficacy, cognitive ability, conscientiousness, and opportunities for feedback (Noe & Colquitt, 2002); not to mention all the environmental and social factors unrelated to the training and unaccounted for in the present study. Having said this, it is important to note that Noe and Colquitt and others (e.g., Mathieu, Tannenbaum, & Salas, 1992) have indicated that training motivation is a key determinant of learning outcomes which directly impact training transfer and job performance. Thus, even though my results could be taken as evidence that training engagement (a motivational construct) really makes no difference when it comes to training outcomes, previous studies and theoretical models would refute such a conclusion.

So, if training engagement really is related to important training outcomes why was there no supporting evidence found in the present study? There are two main explanations for these findings. First, there may have been issues with transfer of training among the study participants. According to models of learning and training transfer, training design characteristics and trainee characteristics directly impact learning outcomes (cognitive, skill, and effective outcomes) which then impact transfer outcomes (Baldwin & Ford, 1988). However, the relationship between learning outcomes and transfer outcomes is moderated by a number of variables including the opportunity to use the skills or knowledge acquired, whether or not what is learned in training is rewarded in the workplace, a supportive transfer climate, through comprehension of the principles being taught, and the strength of the link between training content and job content (Machin, 2002). In the case of the present study, it seems that some of these conditions may have been met while I can only speculate about others. For instance, there were probably ample opportunities to use the skills or knowledge acquired in training "back on the job." In this case it is important to remember that there was no actual job environment per-se. Participants were expected to transfer learning back to their day-to-day lives. Therefore, in the case of amputees, it is hard to imagine them not having ample opportunities to use their skills in their regular daily activities. On the other hand, it is quite unlikely that participants experienced any reward for using their new skills beyond those natural consequences that follow certain actions. This means that although they may have experienced some natural rewards for following certain training protocols (e.g. experiencing relief from a secondary condition due to knowledge gained in the training) there may have been instances where following a training protocol (e.g. how to put on a shirt) was not rewarded or encouraged by anyone in the transfer setting. In this example, the trainee may have finally given up transferring his/her knowledge of how to don a shirt correctly because it was taking longer than the method they commonly use (Schmidt & Bjork, 1992).

Perhaps an even more concerning issue is that the present study did not measure learning outcomes. Both training conditions were designed to increase trainees' declarative and procedural knowledge related to amputation and prostheses. However, no measures of actual learning were employed. Without such measures, I cannot be sure that learning actually took place; and if participants did not learn the content from the training it makes no sense to expect differences in the results level outcomes that were measured (self-efficacy, psychosocial adjustment, pain interference, functional status) (Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). Furthermore, research suggests that near transfer may be more easily achieved than far transfer – which represents a deeper level of learning (Barnett & Ceci, 2002). To clarify, near transfer is the application of learning to situations similar to those in which the original learning took place while far transfer is the application of learning to situations dissimilar to those in which the original learning took place (Laker, 1990). In the present study, my outcome measures would generally be viewed far transfer measures. This means that even though training may have resulted in learning and that learning may have resulted in near transfer, my measures of far transfer may have failed to detect any of these changes. In all, I cannot be confident that participants learned from the training and even if they did, I cannot be confident that their learning was successfully transferred to their day-to-day lives. Thus, the counterintuitive findings for the relationship between engagement and the training outcomes may be due to issues related to learning and transfer rather than a non-existent relationship between the two constructs.

A second potential explanation for the non-existent relationship between engagement and the measured training outcomes is that participants reinterpreted their functional status, self-efficacy, psychosocial adjustment, and pain interference as a result of the training. It is not unreasonable to think amputation/prosthetics training capable of calling one's attention to various deficiencies related to each of the outcomes of interest. For example, one of the many videos used in the training involves an upper limb amputee demonstrating how to perform yard work including how to operate a weed eater. For participants who live in an apartment, condo, or townhome and do not have yardwork responsibilities, this instructional video may only highlight another activity with which they could/should struggle. As a result, it is not unreasonable to expect such a participant to report an unchanged or even slightly lower level of self-efficacy and/or functional status. Similarly, the training includes a video about remaining active in which a lower-limb amputee demonstrates how he operates his Harley Davidson trike. Again, it may not be unreasonable to imagine a participant feeling comparatively less socially integrated or self-efficacious in comparison to the amputee in the video because he/she does not have the means to purchase an expensive recreational vehicle that gives them access to a relatively exclusive social circle of motorcycle enthusiasts (Barrett, 2008). As a consequence, this participant may actually rate his/her social support, psychosocial adjustment, or self-efficacy the same or lower following the training intervention.

This reinterpretation of a measure is known as gamma change (Golembiewski, Bilingsley, & Yeager, 1976). Gama change is a redefinition of a relevant psychological space as the consequence of an intervention. This is compared to alpha change which represents real change and beta change which results from the reinterpretation of a scale. For example, if seeking to measure happiness, alpha change would represent an actual true change in one's rating of happiness (let's say from 1 to 5). Beta change, on the other hand, could occur even without any apparent change to the rating. For example, I rate my happiness at 5 before winning the lottery. After winning, I still rate my happiness as a 5 but I am now significantly happier than I was before. In the case of beta change, I recalibrated the intervals used to measure happiness such that the intervals are much wider after winning than they were before. In contrast to the other two types of change, gamma change would involve a redefinition of the concept of happiness. For example, before news coverage of a natural disaster I rated my happiness as a 4 but rated it as a 5 following the disaster. In this case, nothing has changed in my own life but my interpretation of what it takes to be happy has changed (e.g. my family and I are healthy and safe which is pretty much all I need). In the present study gamma change may have been responsible for the apparent lack of change in the outcome variables. As a result of the training interventions, participants may have redefined their conceptualizations of self-efficacy, psychosocial adjustment, pain interference, and functional status. This reinterpretation may have resulted in the negative or non-existent changes from pre to post measurement. Thus, engagement may be unrelated to the study's outcome variables simply because of the reinterpretation that took place among participants. Unfortunately it is extremely difficult to establish the existence of gamma change or to distinguish it from either alpha or beta change. Further, gamma change is extremely difficult to measure because it is likely to be masked by common measurement instruments whose conceptualization and operationalization are typically rooted in concepts of alpha change (Golembiewski et al., 1976). As a result, I can do little more than speculate as to the role of gamma change in this study.

Next, I found support for Hypothesis 4 which stated that those in the SLE group would report higher levels of engagement (the combination of physical, cognitive, and emotional) than those in a control training condition. T-test results showed marginally significant differences in engagement favoring the SLE group. This finding was further substantiated by a significant result from the non-parametric Mann-Whitney test. In both cases, estimates of effect size were moderate to large indicating that the difference in engagement was meaningful. These results are quite informative given the relative dearth of evidence concerning the impact of SLEs on "motivation" – or more specifically engagement. In contrast to DeRouin-Jessen (2008), the present study did find evidence that the characteristics of SLEs have the potential to impact trainees' levels of engagement. As such, this study serves as one of the first to empirically demonstrate the long held belief that SLE's are (when done right) more engaging than alternate forms of training.

As stated above, we know that simply being in a SLE does not automatically result in higher levels of engagement (Whitton, 2011b; Whitton & Moseley, 2014). Indeed, individual differences including perceptions of SLEs, game style preferences, and experience playing videogames, simulations, or virtual worlds can impact a trainees' tendency to be engaged in a SLE. Having said that, the findings from this study indicate that when designed with forethought and based on solid game characteristics (e.g., fantasy, sensory stimuli, mystery, control etc.) SLEs have an advantage over other forms of training to positively impact perceptions of meaningfulness, safety, availability, and confidence which then impact user behaviors in the form of physical, cognitive, and emotional engagement.

Finally, no support was found for Hypothesis 5 stating that trainee engagement would partially mediate the relationship between training condition and trainee performance. This finding runs contrary to the integrated SLE model outlined in this paper. According to the model the SLE Cycle serves as the primary mediating mechanism between the training input (instructional content and SLE characteristics) and training outcomes. Engagement scores served as the primary measure for the SLE Cycle, and as such were expected to partially mediate the relationship between condition (representing the presence or absence of SLE characteristics) and training outcomes (self-efficacy, psychosocial adjustment, pain interference, and functional status). Unfortunately, I was unable to test this mediational hypothesis. This is not surprising given a failure to find evidence for Hypothesis 3. Without engagement acting as a predictor of training performance (i.e. no mediator (M) to dependent variable (Y) relationship), a mediation model is simply unfeasible. Likewise, the correlation matrix showed relatively small and non-significant correlations between condition and each of the outcome measures indicating no predictor (X) to Y relationship. Thus, in multiple ways the results failed to meet the basic requirements for a mediation model and so provide no evidence for Hypothesis 5 and the mediating role of engagement.

Recall that we have already discussed the reasons for the lack of relationship between engagement and the outcome variables (i.e. the M to Y relationship). However, it may still be useful to discuss a potential explanation for why there was no relationship between training condition and any of the outcome variables. Regardless of training condition, there were no pre-post differences in any of the outcome variables. Thus, one explanation for the lack of relationship between condition and the outcomes is that training was altogether ineffective. The fact that both training conditions were missing some key components of successful training may explain the complete lack of pre to post change in the outcome variables. For instance, research has shown that certain features are critical for creating an optimal learning environment including the provision of cues designed to help trainees learn and recall content (e.g. diagrams, models, key behaviors etc.), and the opportunity to receive feedback from a trainer, a video, or the task itself (Noe & Colquitt, 2002). Both of these features were largely absent from the two training conditions used in the present study. Also absent from either training condition was any form of adaptive guidance. Research by Bell and Kozlowski (2002) demonstrated that providing learners with adaptive guidance (evaluative information and individualized suggestions about what to study and practice) can be far more effective than presenting them with total control over their learning experience. In reality, had the trainees been provided with effective learning cues, rich feedback from trainers, quizzes, etc., and had been provided with adaptive guidance, the study outcomes may have been impacted.

The other potential explanations for the nonexistent relationship between training and any of the outcome variables have already been discussed above. To reiterate, it could certainly be the case that the trainees did learn from the training but failed to transfer that learning back to their daily lives resulting in no measurable differences in the outcomes of interest. Also, it is certainly possible that gamma change occurred; resulting in the redefinition of each of the outcome variables as a consequence of the training intervention. In this case, trainees may have learned from the training and changed as a result, but this change was undetected due to participant's redefinition of the outcome measures. Regardless, of which of the above explanations is correct, the findings from the present study did not find support for the effectiveness of either training intervention and as a result, was unable to shed light on the mediating role of training engagement in the proposed integrated SLE model.

Limitations and Future Directions

Both Study One and Study Two are limited in a few key ways. Study One is limited in terms of its generalizability due to its use of an undergraduate student sample. Some research has suggested that college student samples may not be representative of more general populations that one would expect to encounter in the workplace. As noted above, Sears (1986) contended that college student samples tend to have less-crystallized attitudes, stronger tendencies to comply with authority, and more unstable peer-group relationships than more mature adult samples. Having said this, many of these undergraduates will be entering the job market within the next two to three years and could represent the effects expected among a young sample of employees. Still, there may be other significant differences among a sample of college students trainees compared to a sample of employees. For instance, a manager completing a leadership skills training is probably much more invested in his/her learning experience than many of the undergraduates participating only for course credit. Future research should attempt to study the link between user judgments and user behaviors using a sample more representative of the average trainee in an organization.

Next, although a number of limitations for Study Two have already been discussed (e.g. not measuring learning, providing total control to learners etc.); there are other, more general issues that that must be considered. One of the most obvious limitations was the unusually small sample size. A small sample can have an impact on research results in two ways. First, the small sample may have limited the study's power to detect differences that may have actually been present. For instance, the study's failure to find a link between engagement and many of the outcome variables may have been an artifact of the study's unusually small sample size. In fact, a post-hoc power analysis of the regressions for hypothesis three showed power ranging from .04 to .33 – well below the general recommendation of .8.

Second, because of the small size and unique characteristics of the study's sample, caution should be used when generalizing its findings to the population of lower and/or upper limb amputee adults in the U.S. and/or the population of all working adult learners. For instance, small sample of trainees in the present study volunteered to participate which often is not the case with organizational training (e.g. compliance training etc.). In other words, although using a sample of amputees was good in the sense that they should have been naturally motivated to learn the content and transfer their skills to the work environment" (i.e. their day-to-day lives), it may not have been particularly generalizable to the overall population to which SLE research is aimed. Hence, future research should strive to replicate this study using a much larger and potentially more representative sample.

Another limitation of the present study is the SLE that was used. Although Second Life could definitely be considered a synthetic learning environment similar to a training videogame or other SLEs, the way it was used in the present study failed to offer many of the features that are considered by experts to be important in synthetic learning environments. For example, although the Second Life condition did feature very appealing aesthetics, abstractions of reality, and opportunities for “do overs,” it offered no reward/reinforcement, competition, or defined goals, and offered almost no feedback (cf., Kapp, 2012). These missing components may have adversely impacted the effectiveness of the SLE training. For, example, without any form of competition, the appealing game-like nature (which is characteristic of many SLEs) may have been seriously undermined. For instance you might consider the effort you exert when playing a sport with vs. without others to compete against. Likewise, consider the effort you would put forth in that sport with vs. without feedback (e.g. imagine not knowing if you made or missed a shot). In both cases it is reasonable to think that effort would be significantly reduced when playing without competition or feedback. Therefore, future researchers should seek to explore the questions outlined in this study using a SLE that involves competition, quality feedback, rewards, and defined goals etc.

In summary, my results indicate that there are no differences in motivation to learn, but there are differences in engagement when comparing trainees in an SLE group to a control group. It seems that motivation to learn is not the type of “motivation” that the majority of the literature is referring to when discussing the benefits of synthetic learning environments. On the other hand, results from this study lead me to believe that significant differences between SLE and control groups could be expected when “motivation” is conceptualized more fully and correctly as training engagement. Thus, it seems that training engagement can be increased through the use of SLEs. Practitioners seeking to increase training engagement might be well served to consider SLEs as a component of their overall instructional program. Furthermore, my findings indicate some support for the integrated SLE model in that user judgments (primarily meaningfulness and availability) predict user behaviors (the combination of physical, cognitive, and emotional engagement). Practitioners seeking to increase trainee engagement with SLEs are advised to consider methods of improving judgements of both meaningfulness and availability.

Alternatively, this study did not find evidence that engagement relates to important training outcomes or that it serves as a mediator between training condition and performance. However, key limitations in the study (e.g. failure to measure learning outcomes directly or ensure training transfer) may have contributed to my failure to support Hypotheses 3 and 5. Had I measured learning, I might have determined with certainty whether or not the training had any impact on the trainees and whether or not transfer was an issue. Because the results showed no change in any of the training outcomes it is difficult to make any conclusion about the role of training engagement in the integrated SLE model aside from concluding that user judgments lead to user behaviors in the form of engagement and that SLEs have the potential to increase trainee engagement.

Future research would be well-advised to study the hypotheses outlined in the present study with a few modifications. First, obtaining a larger and more representative employee sample would help overcome the limitations of power, and generalizability, and would help ensure that many of the basic assumptions for the most appropriate analyses are met (e.g. regression, MANCOVA). Second, measuring learning immediately following the training intervention, and measuring near transfer would help paint a clearer picture of the effects of each training program. For instance, with learning measures researchers would be able to determine exactly how, and if, the two training programs differ and whether transfer of training is or is not an issue. Third, different training content might reduce the potential for gamma change. For example, the chances of trainees experiencing a shift in scale interpretation could be less likely to occur when dealing with less emotionally charged material (software training vs. amputee training) (Thompson, & Hunt, 1996).

Finally, future research might consider a different synthetic learning environment. While Second Life is a particularly cost effective choice for SLE designers concerned with aesthetics, in its basic form (i.e. without significant forethought and programing of the virtual space) it lacks the structure, clear goals, rewards, competition, and feedback that is characteristic of truly high quality SLEs (Kapp, 2012). Thus, future researchers would do well to ensure these characteristics are present in whatever SLE is used.

In summary, the present study offers some much needed insight regarding the impact of SLEs on engagement and motivation to learn. It seems that SLEs do not have an impact on motivation to learn but do impact trainee engagement. Moreover, this engagement is largely influenced by user judgments of meaningfulness and availability. The contributions of this study are three fold. First, it provides some evidence that engagement may be the “motivational” concept referred to by those who commonly make the claim that videogame training or SLEs work because they are motivating. Second, it provides some useful modifications to the Garris et al. (2002) input-process-outcome game model. Specifically, it proposes the integrated SLE model which clarifies some of the internal components of the game/SLE cycle and proposes some measures of these components. Third, this study provides evidence of the relationship between user judgments and user behaviors (i.e. engagement). Fourth, this is the first study to demonstrate the advantages in trainee engagement that SLEs can have over other forms of training.

Tables

Table 1.

Comparison of the Characteristics of Multiple Synthetic Learning Environments

Defining Characteristics	Videogames	Simulations	Virtual Worlds
Computer/console mediated	X	X	X
Grounding in reality	Y	Y	Y
Goal oriented	X	X	Y
Hosts many players simultaneously	Y	Y	X
Continues after player leaves	Y	**--**	X
Interactivity	X	X	X
Feedback	X	X	Y
Rules	X	X	Y
Competition	X	Y	Y

Note. X = present in almost all cases, Y = present in some cases, **--** = indicates not present.

Table 2.

Descriptive Statistics and Correlations for Study Variables.

	M	*SD*	1	2	3	4	5
1. Engagement	3.25	.70	--				
2. Meaningfulness	2.82	.88	.70**	--			
3. Confidence	3.41	.95	.40**	.48**	--		
4. Availability	3.55	.86	.41**	.23*	.26**	--	
5. Safety	2.52	-.08	-.07	-.08	.30**	.08	--

$*p < .01$. $**p < .001$.

Table 3.

Multiple Regression of User Judgments on Trainee Engagement.

Coefficient	*B*	*SE*	β	*r*	*t*	*p*	sr^2
Constant	.88	.29	--	--	3.07	.003	--
Meaningfulness	.49	.06	.61	.69	7.71	.000	.51
Safety	.04	.05	.06	.07	.80	.428	.05
Confidence	.04	.06	.06	.40	.72	.473	.05
Availability	.21	.06	.26	.41	3.74	.000	.25

Table 4.

Means and Standard Deviations for Study Two Control Variables

	Technological Savvy	Hours per week: Computer	Hours per week: Videogames	Hours per week: Virtual Worlds
M	4.4	34.5	5.0	.9
SD	.9	18.2	7.7	4.3

Note. Technological savvy was measured on a 1-5 Likert-type scale.

Table 5.

Difference Score Means, Standard Deviations, and Correlations of Main Study Variables

	Range	*M*	*SD*	1	2	3	4	5	6	7	8
1. Condition	N/A	.37	.50	--							
2. Motivation to Learn	1-5	4.27	.52	.42	--						
3. Engagement	1-5	4.25	.63	.44	.68*	--					
4. Self-efficacy General	10-40	.32	2.21	.19	.09	.24	--				
5. Self-efficacy Specific	6-60	-.05	5.99	.03	.19	-.03	-.31	--			
6. Social Support	12-84	-4.15	9.38	-.41	-.28	-.35	-.43	-.01	--		
7. Psychosocial Adjustment	15-60	3.84	7.61	.30	.04	.24	-.02	.27	.05	--	
8. Pain interference	0-6	.02	.72	.27	.10	.03	-.06	-.03	.19	.38	--
9. Functional Status	0-112[L] 0-80[U]	.24	.33	-.02	-.03	.20	.04	.27	-.25	.48*	-.38

Note. Self-efficacy General, Self-efficacy Specific, Social Support, Psychosocial Adjustment, Pain Interference, and Functional Status are represented here as the difference between pre and post measures of the variables. Higher numbers for pain interference mean less pain. Range = scale score range. [L] = LEFS, [U] = UEFS. **p* < .05. ***p* < .01.

Table 6.

Results for Individual Regressions Examining the Predictive Relationship Between Engagement and Each of the Outcome Variables.

	F	*p*	*b*	*t*	*p*
10. Self-efficacy General	1.02	.326	.83	1.01	.326
11. Self-efficacy Specific	.02	.893	-.31	-.14	.893
12. Social Support	2.34	.144	-5.15	-1.53	.144
13. Psychosocial Adjustment	1.00	.331	2.84	1.00	.331
14. Pain interference	.01	.920	.03	.10	.920
15. Functional Status	.67	.416	.10	.83	.416

Note: Self-efficacy General, Self-efficacy Specific, Social Support, Psychosocial Adjustment, Pain Interference, and Functional Status are represented here by change scores. Thus each represents the difference between pre and post measures of the variables.

Figures

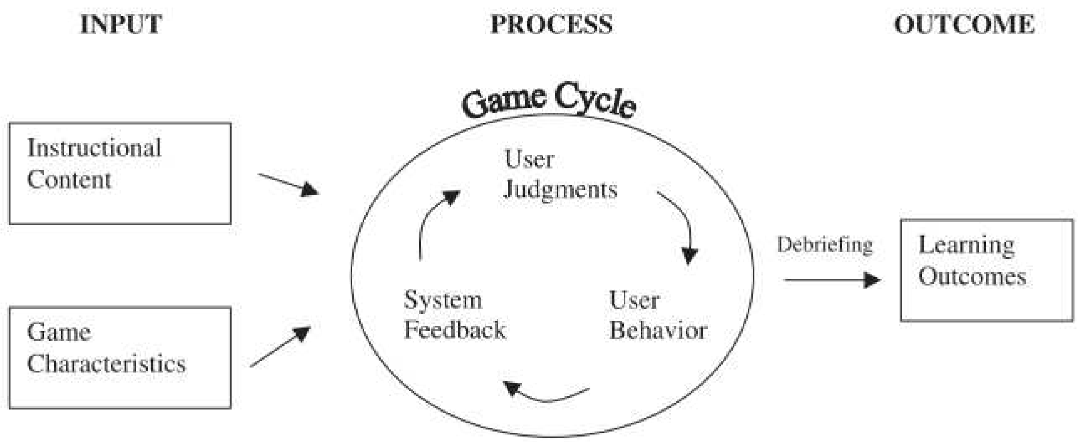

Figure 1. Input-Process-Outcome Game Model from Garris et al. (2002).

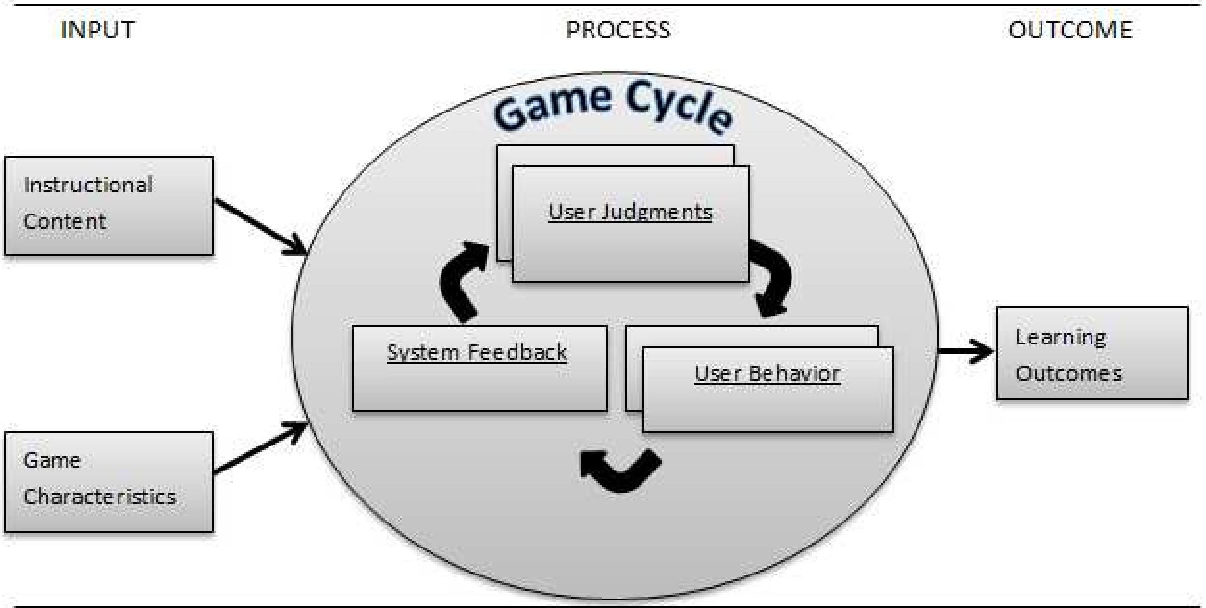

Figure 2a: Input-Process-Outcome Game Model. Taken from Garris et al. (2002)

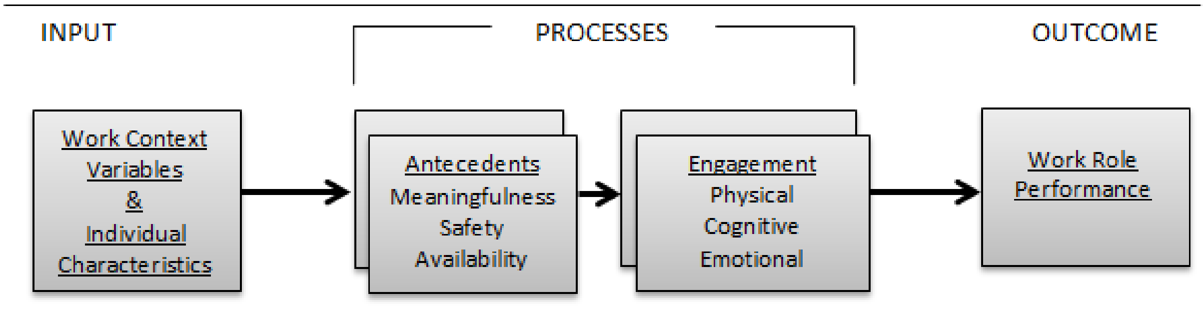

Figure 2b: Model of Engagement. Adapted from Kahn (1990).

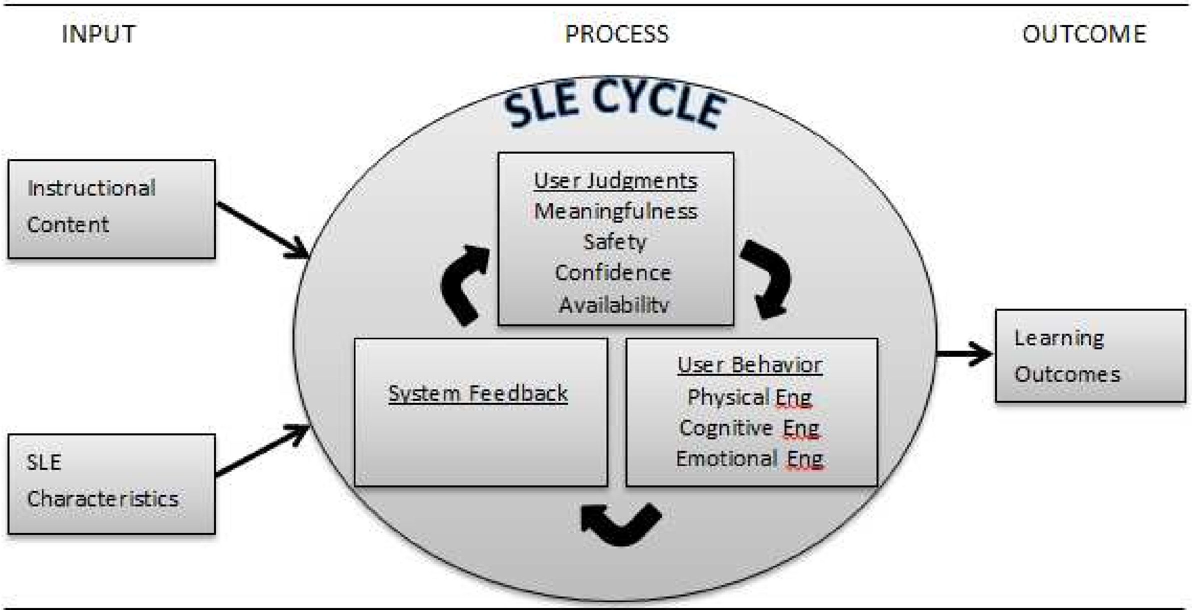

Figure 3: Integrated SLE Model. Eng = Engagement.

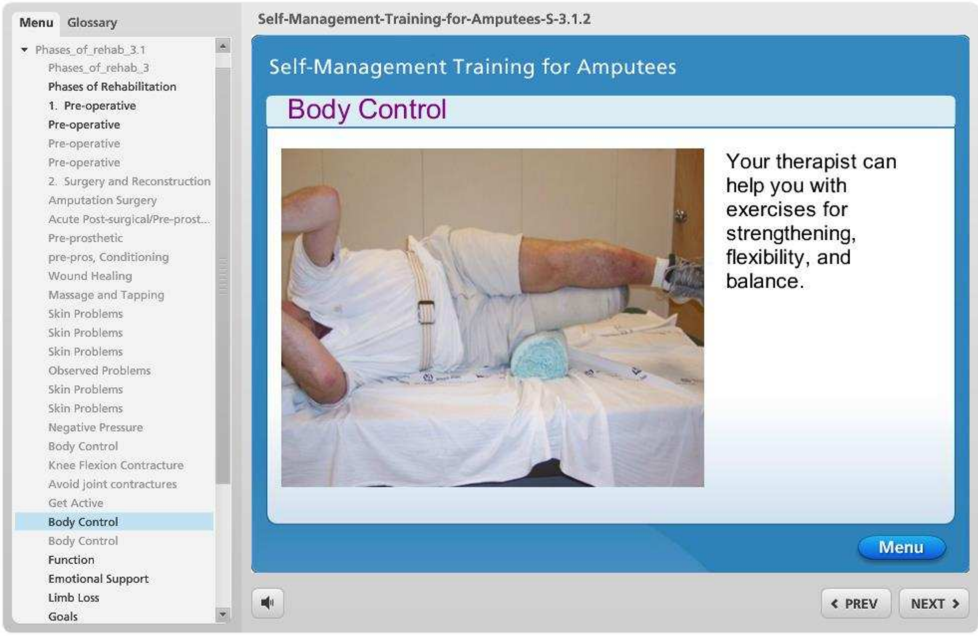

Figure 4. Screenshot from the e-learning control condition

Figure 5. Virtual Health Adventures Island in Second Life

References

Air marshaller training serious game. (2015). In *Designing Digitally*.com. Retrieved February 11, 2015 from http://www.designingdigitally.com/portfolio/serious-games/air-marshaller-training-serious-game#axzz3ROGR1wxP

Alliger G. M., Tannenbaum, S.I., Bennett, W., Traver, H.,& Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341–58.

America's Army. (n.d.). In *Goarmy.com*. Retrieved December 12, 2013, from http://www.goarmy.com/downloads/americas-army-game.html

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.

Barber, N. (2013). *Why are zombies still so popular?* Retrieved February 11, 2015 from http://www.bbc.com/culture/story/20131025-zombie-nation

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612-637.

Barrett, R. (2008, July 11). Harley works to hold its place in American culture. Journal Sentinel. Retrieved from http://www.jsonline.com/business/29463799.html

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*, 1173-1182.

Bell, B. S., & Kozlowski, S. J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology based training. *Personnel Psychology, 55*, 267-306.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, *93*, 296–316. doi:10.1037/0021-9010.93.2.296

Bikson, T. K., & Eveland, J. D. (1990). The interplay of workgroup structures and computer support. In J. Galagher, R. Kraut, & C. Egido, C. (Eds.), *Intellectual teamwork* (pp.243-290). Norwood, NJ: Erlbaum.

Biocca, F., & Levy, M. R. (1995). *Communication in the age of virtual reality*. Hillsdale, NJ: Erlbaum.

Boskic, N. (2015). *Resources on virtual worlds and games for learning.* Retrieved February 10, 2015 from http://www.tonybates.ca/resources/virtual-worlds/

Brown, S. P., & Leigh, T. W. (1996). A new look at psychological climate and its relationship to job involvement, effort, and performance. *Journal of Applied Psychology, 81*, 358-368.

Buffa, C. (2014). *Cool stuff you missed in GTA 5*. Retrieved January 10, 2015 from http://www.primagames.com/games/grand-theft-auto-v/feature/cool-stuff-you-missed-gta-5

Burger, H., Franchignoni, F., Heinemann, A. W., Kotnik, S., & Giordano, A. (2008). Validation of the orthotics and prosthetics user survey upper extremity functional status module in people with unilateral upper limb amputation. *Journal of Rehabilitation Medicine*, *40*, 393-399.

Burton, B. G., Martin, B. & Robins., J. (2013). An examination of student engagement, knowledge creation and expansive learning in a virtual world. In Childs, M. (Ed.) *Understanding learning in virtual worlds* (pp. 65-82). London: Springer.

Byers, C. (2007). Playing to learn: Game-driven comprehension of complex content. *International Journal of Teaching and Learning in Higher Education*, *19,* 33-42. Retrieved from http://www.isetl.org/ijtlhe/pdf/ijtlhe148.pdf

Cannon-Bowers, J. A., & Bowers, C. (2009). Synthetic learning environments: On developing a science of simulation, games, and virtual worlds for training. In Kozlowski, S. W., & Salas, E. (Eds.), *Learning, training, and development in organizations* (pp. 229-261). New York, NY: Routledge

Christian, M. S., Garza, A. S., & Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, *64,* 89-136.

Clark, D. (2007). *Games, motivation & learning.* Sunderland, UK: Caspian Learning.

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

Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: a meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85,* 678-707.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661-686.

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

Cortina, J. M., & Folger, R. G. (1998). When is it acceptable to accept a null hypothesis: No way, Jose? *Organizational Research Methods, 1*, 334-350.

Crawford, E. R., LePine, J. A., & Rich, B. L. (2010). Linking job demands and resources to employee engagement and burnout: a theoretical extension and meta-analytic test. *Journal of Applied Psychology, 95*, 834.

Csikszentmihalyi, M. (1997). *Finding flow: The psychology of engagement with everyday life*. New York: Basic Books.

Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*, 10-32.

Davis, M. (2005, July). *Perfect pitch for video game music*. Retrieved December 12, 2013 from http://news.bbc.co.uk/2/hi/technology/4670009.stm

Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

DeRouin-Jessen, R. (2008). *Game on: The impact of game features in computer-based training.* Unpublished Doctoral Dissertation. University of Central Florida, Orlando. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Dow, S., Mehta, M., Harmon, E., Macintyre, B., & Mateas, M. (2007). Presence and engagement in an interactive drama. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07* (pp. 1475-1484). New York, NY: ACM.

Driskell, J. E., & Johnston, J. H. (1998). Stress exposure training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 191-217).Washington, DC: American Psychological Association.

Druckman, D. (1995). The educational effectiveness of interactive games. In D. Crookall & K. Arai (Eds.) *Simulation and gaming across disciplines and cultures: ISAGA at a watershed* (pp. 178-187). Thousand Oaks, CA: Sage.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350-383.

Edwards, J. (2009). *"America's Army" recruitment video game cost taxpayers $33M.* Retrieved December 12, 2013 from http://www.cbsnews.com/news/americas-army-recruitment-video-game-cost-taxpayers-33m/

Entertainment Software Association. (2013). *Essential facts about the computer and video game industry*. Washington, D.C.: Same as author. Retrieved December 12, 2013 from http://www.theesa.com/facts/pdfs/ESA_EF_2013.pdf

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Filsecker, M., & Kerres, M. (2014). Engagement as a volitional construct: A framework for evidence based research on educational games. *Simulation & Gaming, 45*, 450-470.

Finkelstein, N. D., Adams, W. K., Keller, C. J., Kohl, P. B., Perkins, K. K., Podolefsky, N. S. & LeMaster, R. (2005). When learning about the real world is better done virtually: A study of substituting computer simulations for laboratory equipment. *Physical Review Special Topics-Physics Education Research, 1*, 010103-1–010103-8.

Flood, (2002). Read all about it: Online learning facing 80% attrition rates. *Turkish Online Journal of Distance Education, 3*, 1302-6488.

Fried, Y., & Ferris, G. R. (1987). The validity of the job characteristics model: A review and meta-analysis. *Personnel Psychology, 40*, 287-322.

Gagné, R. M., & Deci, E. L. (2005) Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*, 331–362.

Gagné, R. M., & Medsker, K. (1996). *The conditions of learning*. New York: Harcourt Brace.

Gallagher, P., & MacLachlan, M. (2004). The Trinity Amputation and Prosthesis Experience Scales and quality of life in people with lower-limb amputation. *Archives of Physical Medicine and Rehabilitation*, *85*, 730-736.

Gallagher, P., & MacLachlan, M. (2000). Development and psychometric evaluation of the Trinity Amputation and Prosthesis Experience Scales (TAPES). *Rehabilitation Psychology*, *45*, 130-154.

Garris, R. & Allers, R. (2001). *A game-based training model: Development, application, and evaluation.* Paper presented at the 2001 Interservice /Industry Training, Simulation, and Education Conference, Orlando, FL.

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

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning, 29,* 207–219. doi:10.1111/j.1365-2729.2012.00489.x

Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of Applied Psychology*, *74*, 884-891.

Goldstein, I. L., & Ford, J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4th ed.). Belmont, CA: Wadsworth.

Golembiewski, R.T., Billingsley, K., & Yeager, S. (1976). Measuring change in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science, 12,* 133-157.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis.* New York: Guilford Press.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion* (Tech. Rep. No. 2005-004). Orlando, FL: Naval Air Warfare Center, Training Systems Division.

Heinemann, A. W., Bode, R. K. & O'Reilly, C. (2003). Development and measurement properties of the Orthotics and Prosthetics Users' Survey (OPUS): A comprehensive set of clinical outcome instruments. *Prosthetics and Orthotics International, 27*, 191-206.

Heiphetz, A., & Woodill, G. (2010). *Training and collaboration with virtual worlds: How to create cost-saving, efficient and engaging programs.* NY: McGraw-Hill.

Hickley, D. T., & Anderson, K. (2007). Situative approaches to assessment for resolving problems in educational testing and transforming communities of educational practice. In P. Moss (Ed.), *Evidence and decision making*. *The 103rd NSSE Yearbook* (pp. 269-293). Chicago, IL: National Society for the Study of Education, University of Chicago Press.

Jacobson, M. J. (2006). *Empirical research into learning in 3D virtual and game environments: selected review of the literature* [working paper]. Singapore: Learning Sciences Laboratory, National Institute of Education, Nanyang Technological University.

Jana, R. (2006). On-the-job video gaming. *Business Week, 43*(3977). Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:On-the-job+video+gaming#0

Jarl, G.M., Heinemann, A.W., Norling Hermansson, L.M. (2003). Validity evidence for a modified version of the Orthotics and Prosthetics Users' Survey. *Disability and Rehabilitation Assistive Technology, 7,* 469-478.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal, 33*, 692–724.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74,* 657-690.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education.* San Francisco: Wiley.

Karweit, N. (1989). Time and learning: A review. In R. E. Slavin (Ed.), *School and classroom organization* (pp. 69-95). Hillsdale, NJ: Erlbaum.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2-10.

Kelly, H. (2014). *Meet the virtual patients training real doctors*. Retrieved October 8, 2014 from http://www.cnn.com/2014/10/08/tech/innovation/virtual-patients-medical-training/index.html

Kerns, R. D., Turk, D. C., & Rudy, T. E. (1985). The West Haven-Yale Multidimensional Pain Inventory (WHYMPI). *Pain, 23,* 345-356.

Khan, D. A. (2014). *How major league gamers train*. Retrieved April 15 from http://gearnuke.com/major-league-gamers-train/

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

Kuh, G. D., Kinzie, J., Schuh, J. H., & Whitt, E. J. (2011). *Student success in college: Creating conditions that matter.* San Francisco: Wiley.

Laker, D. R. (1990). Dual dimensionality of training transfer. *Human Resource Development Quarterly, 1,* 209-235.

Learn while you play. (n.d.). In *Learning network*. Retrieved December 12, 2013 from https://learningnetwork.cisco.com/community/learning_center/games

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

Lorig, K., Sobel, D.S., Ritter, P.L., Laurent, D., Hobbs, M. (2001). Effect of a self-management program for patients with chronic disease. *Effective Clinical Practice, 4*, 256-262.

Lorig, K., Stewart, A., Ritter, P., González, V., Laurent, D., & Lynch, J. (1996). *Outcome measures for health education and other health care interventions* (pp. 24-25, 41-45). Thousand Oaks, CA: Sage.

Machin, M. A. (2002). Planning, managing, and optimizing transfer for training. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development* (p. 263-301). San Francisco: Jossey-Bass.

Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 35,* 828-847.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, Learning, and Instruction*, *3*, 223-253.

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

May, D. R., Gilson, R. L., & Harter, L. M. (2004). The psychological conditions of meaningfulness, safety and availability and the engagement of the human spirit at work. *Journal of Occupational and Organizational Psychology, 77*, 11-37.

Mayes, D. K., & Cotton, J. E. (2001). *Measuring engagement in video games: A questionnaire*. Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting, Minneapolis, Minnesota, 8-12 October 2001, (pp. 692-696).

Meister, J. C. (2006, October). Turning online learners into engaged learners [Electronic Version]. *Chief Learning Officer*. Retrieved December 12, 2013, from http://clomedia.com/articles/view/turning_online_learners_into_engaged_learners.

Noe, R. A., & Colquitt, J. A. (2002). Planning for training impact: Principles of training effectiveness. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development* (pp. 53-79) San Francisco: Jossey-Bass.

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

North, M. M., Sessum, J. & Zakalev, A. (2003). Immersive visualization tool for pedagogical practices of computer science concepts: A pilot study. *Journal of Computing Sciences in Colleges, 19,* 207-215.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59,* 938-955.

Parchman, S., Ellis, J., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*, 73–87.

Phillips, R. S., Horstman, T., Vye, N., & Bransford, J. (2014). Engagement and games for learning expanding definitions and methodologies. *Simulation & Gaming, 45,* 548-568.

Pinder, C. C. (1998). *Work motivation in organizational behavior*. Upper Saddle River, NJ: Prentice-Hall.

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

Rastegarpour, H., & Marashi, P. (2012). The effect of card games and computer games on learning of chemistry concepts. *Procedia - Social and Behavioral Sciences, 31*, 597–601.

Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology, 8,* 343-367.

Reeves, B., Malone, T.W., & O'Driscoll, T. (2008). Leadership's online labs. *Harvard Business Review, 86,* 58-66.

Resident Evil Collection. (n.d.). In *Themoviedb.com* Retrieved February 9, 2015, from https://www.themoviedb.org/collection/17255-resident-evil-collection?language=en

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

Rich, B. L., LePine, J. a., & Crawford, E. R. (2010). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, *53,* 617–635. http://doi.org/10.5465/AMJ.2010.51468988

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67. http://doi.org/10.1006/ceps.1999.1020

Saatchi & Saatchi. (2011, June). *Engagement unleashed: Gamification for business, brands, and loyalty.* Retrieved December 12, 2013 from http://saatchi.com/en-s/news/engagement_unleashed_gamification_for_business_brands_and_loyalty

Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology*, *21,* 600-619.

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

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86,* 162-173.

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35- 37). Windsor, England: NFER-NELSON.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515-530.

Sharek, D., & Wiebe, E. (2014). Measuring video game engagement through the cognitive and affective dimensions. *Simulation & Gaming, 45*, 569-592.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64,* 489–528. doi:10.1111/j.1744-6570.2011.01190.x

Slagle, M. (2006). Poll: 4 in 10 Americans play video games [Electronic Version]. *Washington Post.* Retrieved December 12, 2013 from http://www.washington post.com/wp-dyn/content/article/2006/05/07/AR2006050700172.html.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco: Jossey-Bass.

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: electronic mail in organizational communication. *Management Science, 32*, 1492-1512.

Taylor, R. S., & Chi, M. T. H. (2006). Simulation versus text: Acquisition of implicit and explicit information. *Journal of Educational Computing Research, 35*, 289-313.

Thompson, R. C., & Hunt, J. G. J. (1996). Inside the black box of alpha, beta, and gamma change: Using a cognitive-processing model to assess attitude structure. *Academy of Management Review*, *21*, 655-690.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, G. a, Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*, 229–243.

Whitton, N. (2011a). Game engagement theory and adult learning. *Simulation & Gaming, 42,* 596-609.

Whitton, N. (2011b). Theories of motivation for adults learning with games. In P. Felicia (Ed.), *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches* (pp. 352-369). Hershey, PA: IGI Global.

Whitton, N., & Moseley, A. (2014). Deconstructing engagement: Rethinking involvement in learning. *Simulation & Gaming, 45*, 433-449.

Williams R. M., Ehde, D. M., Smith, D. G., Czerniecki, J. M., Hoffman, A. J., Robinson, L. R. (2004). A two-year longitudinal study of social support following amputation. *Disability and Rehabilitation, 26,* 862-74.

Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively-multiuser online graphical environments. *Presence: Teleoperators and Virtual Environments, 15,* 309–329.

Zigurs, I., Poole, M. S., & DeSanctis, G. L. (1988). A study of influence in computer-mediated group decision making. *MIS Quarterly, 12,* 625-644.

Zimet, G. D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988). The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment, 52,* 30-41.

Appendix

Control Variables

1. I consider myself to be technologically savvy (1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree)
2. I consider myself to be a "gamer," e.g., I spend a lot of time playing online or video games (1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree)
3. I consider myself a very social person (1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree)
4. On average, how many hours per week do you spend on the computer?

   (enter 0 if no time spent on the computer)
5. On average, how many hours per week do you spend playing videogames or online games?

   (enter 0 if no time spent on the computer)
6. On average, how many hours per week do you spend interacting in virtual worlds (e.g. Sims, Second Life)?

   (enter 0 if no time spent on interacting in virtual worlds)
7. Have you ever participated in Second Life? a. Yes b. No
8. If you answered "yes" to the previous question, about how many hours a week do you spend in Second Life? _______?.
9. On a scale from 1 to 5, how would you classify yourself in terms of your Second Life expertise? (1=Never Used; 5= Expert)

Learning Motivation

(1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree).

1. I was motivated to learn the information presented in the training program.
2. I tried to learn as much as I could from the training.
3. I got more from this training than most people.
4. The knowledge I gained in this training may advance my career and/or personal life.
5. I volunteered for this training program as soon as I could.
6. The reason I stuck with the training program was because I wanted to learn how to improve my knowledge and skills related to prosthetics.
7. I wanted to improve my knowledge and skills related to prosthetics.
8. If I didn’t understand some part of the training, I tried harder.

<u>User Judgments/Antecedents to Engagement</u>

(1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree).

Meaningfulness

1. This game was entertaining*
2. I got a lot of enjoyment out of playing this game
3. My time was well spent playing this game
4. Playing this game was a meaningful experience*
5. Playing this game heightened my physical energy
6. Playing this game heightened my mental energy
7. Playing this game heightened by emotional energy
8. My time playing this game was <u>NOT</u> well spent
9. Playing this game contributed to my happiness
10. Playing this game was worthwhile
11. This game was important to me

Safety

1. I was **<u>NOT</u>** afraid to be myself in this game
2. I was afraid to make a mistake in this game
3. I worry about looking stupid in this game
4. I know what is expected of me in this game
5. During the game, I felt worried about what others might think about me
6. I worried about making a good impression on others during the game

Availability

1. I felt confident in my ability to handle multiple demands during the game
2. I felt confident in my ability to deal with problems that arose during the game

3. I felt confident in my ability to think clearly during the game
4. I felt confident in by ability to display the appropriate emotions during the game
5. I had trouble focusing on the game
6. I felt tired during the game
7. I felt distractible during the game

Note * = Item removed from scale based on CFA results.

Engagement

(1 = strongly disagree; 2 = disagree; 3 = Undecided; 4 = agree; 5 = strongly agree).

Physical

1. I worked with intensity on this training
2. I exerted my full effort on this training
3. I devoted a lot of energy to this training
4. I tried my hardest to perform well on this training
5. I strove as hard as I could to complete this training
6. I exerted a lot of energy on this training

Emotional

7. I was enthusiastic in this training
8. I felt energetic in this training
9. I was interested in this training
10. I felt proud of my work in this training
11. I felt positive about this training
12. I felt excited about this training

Cognitive

13. During this training, my mind was focused on the material
14. During this training, I paid a lot of attention to the material
15. During this training, I focused a great deal of attention on the material
16. During this training, I was absorbed by the material

17. During this training, I concentrated on the material

18. During the training, I devote a lot of attention to the material

General Self-Efficacy (GSE) Scale

(1 = Not at all true   2 = Hardly true   3 = Moderately true   4 = Exactly true)

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.

<u>The revised Stanford University School of Medicine Chronic Disease Self-Efficacy Scale</u>

(1 = Not at all confident; 10 = Totally Confident)

1. Having an illness often means doing different tasks and activities to manage your condition. How confident are you that you can do the different tasks and activities needed to manage your health condition?
2. How confident are you that you can judge when the ***changes in your condition are affecting your daily life****?*
3. How confident are you that you can reduce the emotional distress caused by your health condition so that it does not affect your everyday life?
4. How confident are you that you can do things other than just taking medication to reduce how much your illness affects your everyday life?
5. How confident are you that you can keep the ***conditions associated with amputation*** from interfering with the things you want to do?
6. How confident are you that you can keep the physical discomfort or pain of your disease from interfering with the things you want to do?

**Note that changes to the original measure are marked with bold italics

Multidimensional Scale of Perceived Social Support

Instructions: We are interested in how you feel about the following statements. Read each statement carefully. Indicate how you feel about each statement.

(1 = Very Strongly Disagree, 2 = Strongly Disagree, 3 = Mildly Disagree, 4 = Neutral, 5 = Mildly Agree, 6 = Strongly Agree, 7 = Very Strongly Agree)

1. There is a special person who is around when I am in need. SO
2. There is a special person with whom I can share my joys and sorrows. SO
3. My family really tries to help me. Fam
4. I get the emotional help and support I need from my family. Fam
5. I have a special person who is a real source of comfort to me. SO
6. My friends really try to help me. Fri
7. I can count on my friends when things go wrong. Fri
8. I can talk about my problems with my family. Fam
9. I have friends with whom I can share my joys and sorrows. Fri
10. There is a special person in my life who cares about my feelings. SO
11. My family is willing to help me make decisions. Fam
12. I can talk about my problems with my friends. Fri

** Note that the items tended to divide into factor groups relating to the source of the social support, namely family (Fam), friends (Fri) or significant other (SO).

Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R)

(1 = Strongly Disagree, 2 = Disagree, 3 = Agree, 4 = Strongly Agree, 5 = Not Applicable)

1. I have adjusted to having a prosthesis
2. As time goes by, I accept my prosthesis more
3. I feel that I have dealt successfully with this trauma in my
4. Although I have a prosthesis, my life is full
5. I have gotten used to wearing a prosthesis
6. I don't care if somebody looks at my prosthesis
7. I find it easy to talk about my prosthesis
8. I don't mind people asking about my prosthesis
9. I find it easy to talk about my limb loss in conversation
10. I don't care if somebody notices that I am limping
11. A prosthesis interferes with the ability to do my work
12. Having a prosthesis makes me more dependent on others than I would like to
13. Having a prosthesis limits the kind of work that I
14. Being an amputee means that I can't do what I want to do
15. Having a prosthesis limits the amount of work that I can do

West Haven-Yale Multidimensional Pain Inventory

In the following 9 questions, you will be asked to describe your pain and how it affects your life. Under each question is a scale to record your answer. Read each question carefully and then circle a number on the scale under that question to indicate how that specific question applies to you.

2. In general, how much does your pain problem interfere with your day to day activities?

0 1 2 3 4 5 6

No interference Extreme interference

3. Since the time you developed a pain problem, how much has your pain changed your ability to work?

0 1 2 3 4 5 6

No change Extreme change

___ Check here, if you have retired for reasons other than your pain problem

4. How much has your pain changed the amount of satisfaction or enjoyment you get from participating in social and recreational activities?

0 1 2 3 4 5 6

No change Extreme change

8 .How much has your pain changed your ability to participate in recreational and other social activities?

0 1 2 3 4 5 6

No change Extreme change

9. How much has your pain changed the amount of satisfaction you get from family-related

activities?

0 1 2 3 4 5 6

No change Extreme change

13. How much has your pain changed your marriage and other family relationships?

0 1 2 3 4 5 6

No change Extreme change

14. How much has your pain changed the amount of satisfaction or enjoyment you get from work?

0 1 2 3 4 5 6

No change Extreme change

__ Check here, if you are not presently working.

17. How much has your pain changed your ability to do household chores?

0 1 2 3 4 5 6

No change Extreme change

19. How much has your pain changed your friendships with people other than your family?

0 1 2 3 4 5 6

No change Extreme change

Draft

Patient ID# ____ Date __ / __ / ____

OPUS UPPER EXTREMITY FUNCTIONAL STATUS

I. Please indicate your affected limb(s). ☐ Left arm ☐ Right arm ☐ Both arms

II. How many hours per day do you currently wear your prosthesis or orthosis? ____

III. Using the scale to the right, please indicate how easily you perform the following activities.	Very easy	Easy	Slightly difficult	Very difficult	Cannot perform activity	Not applicable	IV. Do you usually perform this activity using or not using your prosthesis or orthosis? Using	Not using
1. Wash face								
2. Put toothpaste on brush and brush teeth								
3. Brush/comb hair								
4. Put on and remove t-shirt								
5. Button shirt with front buttons								
6. Attach end of zipper and zip jacket								
7. Put on socks								
8. Tie shoe laces								
9. Drink from a paper cup								
10. Use fork or spoon								
11. Cut meat with knife and fork								
12. Pour from a 12 oz can								
13. Write name legibly								
14. Use scissors								

Draft

Patient ID#

OPUS UPPER EXTREMITY FUNCTIONAL STATUS

III. Using the scale to the right, please indicate how easily you perform the following activities.	Very easy	Easy	Slightly difficult	Very difficult	Cannot perform activity	Not applicable	IV. Do you usually perform this activity using or not using your prosthesis or orthosis? Using	Not using
15. Open door with knob								
16. Use a key in a lock								
17. Carry laundry basket								
18. Dial a touch tone phone								
19. Use a hammer and nail								
20. Fold bath towel								
21. Open an envelope								
22. Stir in a bowl								
23. Put on and take of prosthesis or orthosis								
24. Open a bag of chips using both hands								
25. Twist a lid off a small bottle								
26. Sharpen a pencil								
27. Peel potatoes (or fruit) with a knife/peeler								
28. Take bank note out of the wallet								

Draft

Patient ID

OPUS Functional Status Measure

Please enter today's date.

/ /

Key terms:

Orthotic - a brace that helps support a limb and improve its functioning.

Prosthetic - an artificial extension that replaces a missing body part.

How easy, or difficult, is it for you to:	Very easy	Easy	Slightly difficult	Very difficult	Cannot do this activity	Do you typically wear an orthotic or prosthetic device to perform this activity? No	Yes
1. Get into and out of the tub or shower	●	○	○	○	○	○	○
2. Dress your lower body	○	○	○	○	○	○	○
3. Get on and off the toilet	○	○	○	○	○	○	○
4. Get up from the floor	○	○	○	○	○	○	○
5. Balance while standing	○	○	○	○	○	○	○
6. Stand for one-half hour	○	○	○	○	○	○	○
7. Pick up an object from floor while standing	○	○	○	○	○	○	○
8. Get up from a chair	○	○	○	○	○	○	○
9. Get into and out of a car	○	○	○	○	○	○	○
10. Walk around indoors	○	○	○	○	○	○	○
11. Walk outside on uneven ground	○	○	○	○	○	○	○
12. Walk in bad weather (*e.g.*, rain, snow, wind)	○	○	○	○	○	○	○
13. Walk up to two hours	○	○	○	○	○	○	○
14. Walk up a steep ramp	○	○	○	○	○	○	○
15. Get on and off an escalator	○	○	○	○	○	○	○
16. Climb one flight of stairs with a rail	○	○	○	○	○	○	○
17. Climb one flight of stairs without a rail	○	○	○	○	○	○	○
18. Run one block	○	○	○	○	○	○	○
19. Carry a plate of food while walking	○	○	○	○	○	○	○
20. Put on and take off orthotic or prosthetic	○	○	○	○	○		

How many days per week do you currently wear an orthotic or prosthetic device?
(Please enter # of days) *(0 - 7)*

How many hours per day do you currently wear an orthotic or prosthetic device?
(Please enter # of hours) *(0 - 24)*

This is a statement other patients have made. ***"I should not do physical activities which (might) make my pain worse."*** **Do you:**

- ○ Completely disagree
- ○ Mostly disagree
- ○ Slightly disagree
- ○ Unsure
- ○ Slightly agree
- ○ Mostly agree
- ○ Completely agree