Feedback Source Modality Effects in Game-Based Training: A Trade-Off Analysis

Benjamin Goldberg[a]
[a)]U.S. Army Research Laboratory-Human Research and Engineering Directorate-
Simulation and Training Technology Center

In this paper we explore available tools for integrating intelligent tutoring communications in game-based learning platforms and examine theory-based techniques for delivering explicit feedback in such environments. The primary tool influencing the design of this research is the open-source Generalized Intelligent Framework for Tutoring (GIFT), a modular domain-independent architecture that provides the tools to author, deliver, and evaluate intelligent tutoring technologies. Influenced by social cognitive theory and cognitive load theory based research, the resulting experiment looks at varying approaches for using an Embodied Pedagogical Agent (EPA) defined as a tutor in a game-based training environment. Treatments were authored to examine tradeoffs between embedding an EPA directly in a game, embedding an EPA in GIFT's browser-based Tutor-User Interface (TUI), or using audio prompts alone with no social grounding. A condition effectiveness evaluation is presented examining treatment effects on performance, reported mental demand, and perceived usefulness of feedback. Outcomes support the use of an EPA present in GIFT's TUI when compared to effectiveness scores on audio prompts alone and the game embedded EPA.

INTRODUCTION

Game-based training designed within synthetic virtual worlds provide the environments for 'practicing' the application of acquired skills, but often lack instructional guidance essential for effective training to occur (Nicholson et al., 2007). However, through proper integration of artificial intelligence and Intelligent Tutoring System (ITS) technologies, game-based training applications now have the ability to support pedagogical interventions intended to maintain training progression in the absence of live instruction (Sottilare et. al., 2013). It is the goal of this research to address a fundamental gap identified in the adaptive game-based training literature; specifically, what is the best method to embed pedagogy and feedback within game-based instructional environments that optimizes outcomes and reduces implementation complexities.

A fundamental problem in this area is a lack of empirical evidence supporting the usefulness of instructional components and explicit feedback mechanisms in game-based learning events. A common trend is incorporating ITS functions in game environments just because they are now possible rather than because there is evidence of their effectiveness (Sweller, 2008). Empirical analysis is required to identify optimal approaches for delivering training relevant feedback in such environments.

The intent of this work is to support a tradeoff assessment of varying communication methodologies to determine the most favorable implementation. Selected methodologies are grounded in research linked to social cognitive theory and cognitive load theory. From the perspective of social cognitive theory, this study assesses whether explicit feedback delivered by Embodied Pedagogical Agents (EPAs) present in a scenario has a significant effect on performance when compared to external feedback source modalities. This tests prior findings in the learning sciences community that posits using virtual entities in computer-based learning events improves performance outcomes and improves affective responses towards interaction (Bandura, 2011; Moreno et al., 2001; Graesser & McNamara, 2010).

From the perspective of cognitive load theory, this research examines different feedback interfacing modalities and tests principles highlighted in Wicken's (2002) multiple resource theory. Specifically, the experiment is designed to investigate approaches for housing an EPA, whether it is embedded directly in the game environment or if it is located in an external interface separate from the game world itself. This will offer insight into whether there is a direct benefit associated with embedding an agent acting as a socialized tutor directly in the game world or if other technologies can be leveraged, such as using a separate interface component to house a virtual entity, assuming this new component does not strain cognitive resources, thus effecting performance.

To inform future design considerations, a comparative evaluation is presented. Metrics linked to performance, mental demand (MD), and usefulness of feedback were collected to fuel an analysis technique that examines the effectiveness of a condition in relation to a defined control. The technique being applied is derived from Kalyuga et. al.'s (1999) methodology to produce an instructional effectiveness metric. The benefit is that it allows you to observe the effect experimental conditions have on outcomes for two defined dependent variables and their relationship when compared against a designated control. The analysis presented was administered similarly to Kalyuga et al.'s (1999) implementation in that it is applied only to the experiment's associated transfer and retention tests.

METHODS

Subjects

One-hundred thirty-one cadets enrolled at the United States Military Academy (USMA) at West Point were recruited as volunteer subjects for the experiment. USMA cadets were selected because they represent an Army relevant population of future Officers who will potentially interact with

Not subject to U.S. copyright restrictions. DOI 10.1177/1541931215591401.

ITS integrated training platforms. They also represent an ideal sample for a university student population who lack specific skill sets, which is a key focus for development of such technology. Of the 131 subjects, 105 were male and 26 were female. Participant age ranged between 18 and 23 years of age and all were registered at the time in the PL100 General Psychology course. Of the 131 subjects, six data sets were removed due to complications, leaving 125 usable sets.

Apparatus

TC3Sim. The serious game selected for this study was the Tactical Combat Casualty Care Simulation (TC3Sim), also known as vMedic. The game was developed to instruct the tactics, techniques, and procedures associated with operating as an Army Combat Medic and Combat Lifesaver (CLS). Interaction is based around story-driven scenarios within a game-engine based simulation and uses goal-oriented exercises to provide a means to train a closely grouped set of related tasks (Fowler, Smith, & Litteral, 2005). TC3Sim tasks include assessing casualties, performing triage, providing initial treatments, and preparing a casualty for evacuation under conditions of conflict.

Generalized Intelligent Framework for Tutoring. The Generalized Intelligent Framework for Tutoring (GIFT) is an open-source domain-independent Intelligent Tutoring System (ITS) authoring environment (Goldberg et al., 2012; Sottilare et al., 2013). For the purpose of this experiment, GIFT was applied to conduct real-time assessment on interaction within the serious game environment TC3Sim. Assessment capabilities were provided through a tool within GIFT called SIMILE (Student Information Models for Intelligent Learning Environments; Mall & Goldberg, 2015). SIMILE serves as a run-time assessment engine by examining user data generated during gameplay, and compares specific message types against pre-defined rule sets.

This real-time assessment enables GIFT to detect errors in task performance, which in turn triggers pedagogical interventions intended to influence subsequent behaviors. In the context of this study, GIFT serves as the testbed architecture for managing both real-time assessment, as well as directing what feedback is delivered to a participant during gameplay and how that feedback is delivered (i.e., the modality from which the information is communicated). A functional component in GIFT relevant to this study is the Tutor-User Interface (TUI). The TUI is a browser-based communication layer built to collect user inputs and to relay information back to the user. For real-time guided instruction, the TUI can be used for delivering explicit feedback content in real-time. It supports multimedia applications and the presence of virtual entities acting as defined tutors.

The TUI is an interesting component because it enables the inclusion of EPAs with no programming required. It utilizes open-source technologies and does not require any modifications to a game environment to support the presence of a virtual tutor. In the context of feedback, this requires the evaluation of its function to determine if it supports or hinders performance outcomes. Hence, two subordinate questions will be specifically addressed: 1) what effect does the source modality of explicit feedback have on performance and workload; and 2) what effect does the source modality of explicit feedback have on subsequent interaction related to retention and transfer?

Experimental Design

The design for this experiment is a counter-balanced mixed design with two independent variables (IV), (1) source of feedback and (2) character profile. Source of feedback refers to the communication component that relays feedback information to the user. For this experiment, source conditions are described as being internal or external to the training environment. These conditions incorporate EPAs as interfacing characters, which are present either in the game environment as an agent of the scenario or located externally from the scenario in the GIFT TUI (see Figure 1). The second IV, character profile, was based on a description of the EPA's background and role within the training event, and was based around research on social cognitive theory's persona effect (Lester et al., 1997; Moundridou & Virvou, 2002).

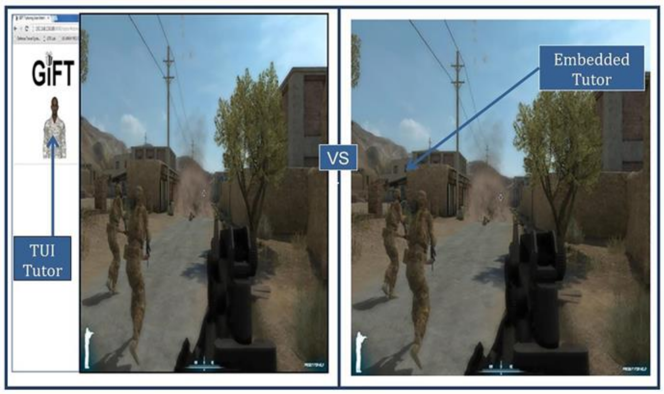

Figure 1. Variable Source Modality Conditions

For assessing the effect manipulated variables have on associated dependent measures, there are two control conditions. The first involves the initial TC3Sim guided scenario without any tutor interaction or explicit feedback. This is how TC3Sim is currently implemented. The second control incorporates the initial TC3Sim guided scenario with feedback provided solely as an audio message. This condition is being termed 'Voice of God' (VoG) as there is no direct visual component accompanying the voice message; as if it comes from nowhere. This condition enables the ability to determine if the presence of an EPA effects participant outcomes on dependent variables of interest, as well as if the feedback presented solely as an audio file improves performance when compared to the baseline condition.

Metrics

For the purpose of this paper, we are interested in the dependent metrics that can be used for determining condition effectiveness scores as they relate to the IVs of interest in the experimental design. As such, the metrics that were found to have consistently reliable differences between groups were selected for further analysis presented here. These included two sources of performance-based outcomes, a self-reported value of exerted MD as captured within the NASA-TLX (Hart & Staveland, 1988), and the usefulness of feedback provided during interaction as captured within the RETRO Flow Scale (Procci et. al., 2012).

Procedure

Upon arrival participants signed an informed consent and were assigned to an experimental condition. Next, they began interaction with GIFT by logging in. Once started, a participant completed a demographics survey and a videogame experience metric. Following, a pre-test assessing initial knowledge levels was administered. This initial metric was used to determine learning gains following interaction with the training materials. Next, GIFT directed a set of slides developed to deliver TC3 associated content. The course materials were self-guided and included interactive multimedia selected across multiple source applications. All participants interacted with the same courseware, with subjects spending an average of 10-12 minutes.

Following courseware presentation, GIFT initialized the first interaction with the TC3Sim environment. Participants performed a short scenario designed to introduce the interfaces and inputs associated with the game. Next, GIFT prepped the subject for the first of two scenarios in TC3Sim. This is where manipulations to the IVs were introduced. All conditions presented a mission overview highlighting the objectives of the game session. Within this overview was an introduction to the EPA the participant would interact with. A background description associated with the EPA was provided for the purpose of defining the agent's perceived role. Participants in the two assigned control conditions only received a mission overview before progressing into the game.

The mission introductory materials led directly into the first of two scenarios in TC3Sim. The first scenario incorporated real-time feedback presented through the assigned condition source. During task interaction, GIFT interpreted user inputs for determining performance and communicated results for executing feedback scripts. Based on the assigned modality condition, feedback was delivered either as audio only (VoG), through an EPA present in GIFT's TUI, through a character present in the game environment, or no feedback at all. Upon completion, participants answered survey instruments providing data on MD experienced during interaction and the usefulness of feedback provided by GIFT. This led into the second of two scenarios involving similar events to the first, minus the real-time feedback element. GIFT monitored interaction and provided outcome results as a source of performance for determining skill at executing trained procedures with no assistance. Following, a post-test was administered in similar fashion to the initial pre-test.

RESULTS

Prior analyses on the resulting data set yielded interesting findings (see Goldberg & Cannon-Bowers, 2015 for a full breakdown of condition comparisons). The dependent measures found to be reliably effected were (1) performance marks on both game and test outcomes, (2) reported MD levels experienced during the training event as collected from the NASA-TLX, and (3) the perceived usefulness of the feedback provided during game interaction as collected from the RETRO Flow Scale. While the dependent variables were assessed against both IVs, the variable of Feedback Source Modality was the only one to produce reliable differences. As such, we will first present results of comparative analyses that were found to be significantly different (see Table 1), followed by the application of those results in the context of a condition effectiveness evaluation. For this paper, we focus on the condition effectiveness evaluations, as it will provide valuable insight into tradeoff considerations associated with the design and development of game-based training applications embedded with intelligent tutoring capabilities.

Table 1. Bonferroni Post-hoc Condition Comparisons found to be Significantly Different across Dependent Metrics

Source Modality Conditions		***Knowledge Post-Test Scores*** Mean	Standard Error	***Significance***
TC3Sim Embedded VoG		68.20 60.00	.021 .030	$p = .029$
TUI Embedded VoG		69.80 60.00	.021 .030	$p = .009$
TUI Embedded No Feedback		69.80 61.90	.021 .031	$p = .036$
		Game Capstone Scenario % ***Mean***	***Standard Error***	
TUI Embedded No Feedback		40.60 35.40	.013 .018	$p = .024$
TC3Sim Embedded No Feedback		39.40 35.40	.013 .018	$p = .080^*$
		Flow Feedback Usefulness ***Mean***	***Standard Error***	
TC3Sim Embedded No Feedback		3.33 2.63	.103 .149	$p = .001$
TUI Embedded No Feedback		3.31 2.63	.104 .149	$p = .002$
VoG No Feedback		3.53 2.63	.145 .149	$p < .001$
		NASA-TLX Mental Demand ***Mean***	***Standard Error***	
TC3Sim Embedded VoG		80.81 73.86	1.95 2.96	$p = .049$
TUI Embedded VoG		84.55 73.86	2.84 2.96	$p = .022$
No Feedback VoG		85.33 73.86	3.19 2.96	$p = .012$

Condition Effectiveness Scores

The effectiveness metric is derived from calculated *Z*-scores across two variables as they relate to a control treatment. These generated *Z*-scores are then represented as a coordinate system to provide a visual representation of the experimental condition's efficiency (see Figure 2). For this tradeoff analysis, the variables examined are those that were found to have significant differences across experimental treatments as reported above. This approach shows how individuals mentally engage in game-based training events with real-time feedback and how communication modalities affect the way resources are managed and how that content is ultimately perceived. These factors will be compared to performance outcomes, thus potentially providing empirical evidence to support a data informed source modality solution.

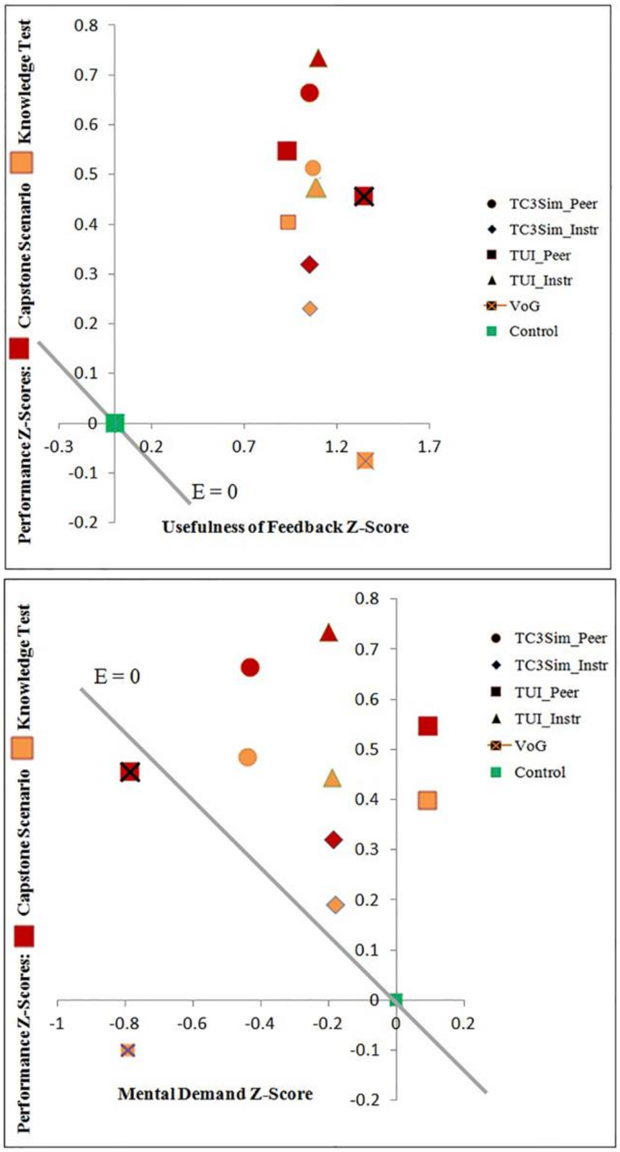

Figure2. Condition Effectiveness Plots Comparing Game and Test Outcomes with Feedback Usefulness and Mental Demand

With *Z*-score values calculated for each variable, the following formula is applied:

$$E = (P + R) / /SQRT(2)/ \quad (1)$$

where performance *Z*-score (P) and MD/Feedback Usefulness *Z*-score (R) produce a value to determine training effectiveness (E) (Kalyuga et al., 2009; see Table 2). The resulting coordinate points determined above are measured against the line of zero effectiveness ($E = 0$), as represented in Figure 2. It is important to note that the resulting effectiveness scores for MD are based on positive associations with higher reported outcomes. In other words, the condition effectiveness outcomes associate higher mental exertion with better overall scores. While many systems are designed to control, and oftentimes minimize MD for operating a system, in the context of a learning event higher MD is believed to be a desired state as this associates with attentive interaction.

Table 2. Condition Effectiveness Scores Comparing Performance Metrics with MD and Feedback Usefulness

	Game/ Feedback	Game/ MD	PostTest/ Feedback	PostTest/ MD
TC3SimPeer	1.211	0.162	1.083	0.035
TC3SimInstr	0.967	0.095	0.876	0.004
TUIPeer	1.043	0.451	0.941	0.349
TUIInstr	1.293	0.378	1.092	0.177
VoG	1.273	-0.232	0.878	-0.627
Control	0	0	0	0

DISCUSSION

The discussion of results is conducted from two perspectives. The first is looking at in-game interaction where the MD and Feedback Usefulness metrics from the TC3Sim training scenario are compared against the subject's performance on the subsequent capstone game scenario. This shows if the level of perceived MD linked to the training scenario correlates with performance outcomes on the following assessment. This technique also illustrates the relationship between the perceived feedback value delivered in the training scenario and its effect on performance in a transfer setting. The second perspective is examining the same DVs in conjunction with outcomes from the knowledge post-test. This examines if the feedback provided is attributable to increases in knowledge acquisition. According to Schmidt & Bjork (1992) it is critical to add transfer and retention phases when comparing treatment conditions on learning effect, as these subsequent measures are often better indicators of the IVs influence on performance differences across groups.

In examining the condition effectiveness scores in relation to capstone game performance, outcomes show the VoG condition to rate similarly with all EPA conditions for Feedback Usefulness. However, the VoG group ranks last in terms of reported MD, even when compared against the control. This relationship is interesting, as the individuals in the VoG condition were cognizant enough to understand that information received during gameplay was helpful, yet their level of mental effort was reported significantly lower when compared against all conditions. The question is why do subjects in the VoG condition reliably report interaction in the training scenario to be less cognitively demanding when they are receiving the same feedback information?

Based on the authors' opinion, there are two possible explanations for this relationship. First, the variance of MD between the VoG and EPA groups is attributable to the presence of a virtual entity who manages feedback delivery. This holds in line with tenets of social cognitive theory, in that simulated agents assist individuals in processing and remembering information when compared to more simplistic communication modalities (Graesser & McNamara, 2010; Gulz, 2004). In line with this thought, the inclusion of an EPA requires more mental exertion to maintain awareness of actions in the environment while also perceiving and processing information from the tutor. Yet, participants in the control condition with no feedback reported similar levels of MD when compared against all EPA treatments. The possible difference is while the EPA participants used additional cognitive resources to monitor tutor interaction, the no feedback participants increased their cognitive attention to gauge scenario progress by monitoring information implicitly delivered in the game environment alone.

The other feasible explanation is what is believed to be a cognitive prompting effect linked with EPA treatments. For the context of this research, cognitive prompting is defined as any set of interactions experienced by a participant prior to a training event that implicitly influences the processing of material and communications (Cleeremans, 2001). For this study, participants in the EPA related conditions were presented a tutor profile notifying them that their performance would be monitored and that feedback would be provided based on real-time assessment; thus notifying the participant that feedback would be linked to objectives the game instills. It also primes the individual to use additional cognitive resources to efficiently perceive these additional channels of information. This was not the case for the VoG condition, as no form of cognitive prompting was administered. These individuals received a scenario background and were launched directly into the environment with no notification that performance-based feedback would be delivered. As a result, these participants were not expecting feedback and most likely viewed the reflective prompts triggered by GIFT as elements associated with the game environment itself.

To follow-up effectiveness interpretations based on in-game performance, the same effectiveness scores were produced in relation to outcomes on the knowledge post-test. While the VoG reported the lowest MD and the highest in perceived Feedback Usefulness during TC3Sim training, these subjects showed the poorest transfer of knowledge. In fact, this group was the only one to produce negative performance gains between the pre- and post-test, while also producing lower scores than the group who received no feedback at all. This evidence may support the cognitive prompting effect referenced above. It seems as if participants in the VoG condition disregard the explicit feedback provided as if it is not grounded to a pedagogical function.

Overall, the outcomes of this analysis show the presence of an EPA to benefit performance outcomes, yet determining where best to situate that EPA is still up for debate. In this instance, embedding a tutor in the game world rather than using GIFT's TUI shows no benefit on performance or across any of the DVs. Because of this, using GIFT's TUI can provide a large advantage for incorporating an EPA element in a game-based application. It drastically reduces the amount of time, effort, and money to modify a game to support character interaction requests from the tutor's pedagogical model.

In terms of identifying tradeoffs between EPAs and the VoG approach, the data supports the inclusion of an EPA in the context of this training application. Though the VoG was effective in aiding a learner to perform, the reduction in performance on transfer and retention tests creates concerns on its effect in aiding a learner to commit feedback content to memory for future application.

With respect to future research, one key element that needs further investigation is the effect of the aforementioned cognitive prompting on MD. For game-based learning, this involves notifying users that performance relevant feedback will be provided. A new study is required to test this effect across the same experimental procedure, where the VoG participants receive a 'feedback will be present' prompt.

REFERENCES

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

Cleeremans, A. (2001). Conscious and unconscious processes in cognition. *International Encyclopedia of the Social and Behavioral Sciences*, 2584-2589.

Fowler, S., Smith, B., & Litteral, C. S. M. D. J. (2005). *A TC3 game-based simulation for combat medic training.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Goldberg, B., Brawner, K. W., Holden, H., & Sottilare, R. (2012). *Adaptive Game-Based Tutoring: Mechanisms for Real-Time Feedback and Adaptation.* Paper presented at the DHSS 2012, Vienna, Austria.

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior, 52*, 1-11.

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

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

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

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

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

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

Mall, H., & Goldberg, B. (2015). *SIMILE: An Authoring and Reasoning System for GIFT.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2).

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

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

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

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

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

Sottilare, R., Brawner, K. W., Goldberg, B., & Holden, H. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In C. Best, G. Galanis, J. Kerry & R. Sottilare (Eds.), *Fundamental Issues in Defense Training and Simulation* (pp. 223-234). Burlington, VT: Ashgate Publishing Company.

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

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science, 3*(2), 159-177.