Computers in Human Behavior 52 (2015) 1–11

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference

Benjamin Goldberg [a,*], Janis Cannon-Bowers [b]

[a] U.S. Army Research Laboratory, 12423 Research Parkway, Orlando, FL 32826, United States
[b] Institute for Simulation and Training, University of Central Florida, 3100 Technology Parkway, Orlando, FL 32826, United States

ARTICLE INFO

Article history:

Keywords:
Intelligent tutoring systems
Explicit feedback
Pedagogical agents
Generalized intelligent framework for tutoring
Game-based training
Cognitive load

ABSTRACT

The aim of this research is to enhance game-based training applications to support educational events in the absence of live instruction. The overarching purpose of the presented study was to explore available tools for integrating intelligent tutoring communications in game-based learning platforms and to examine theory-based techniques for delivering explicit feedback in such environments. The primary tool influencing the design of this research was the open-source Generalized Intelligent Framework for Tutoring (GIFT), a modular domain-independent architecture that provides the tools and methods to author, deliver, and evaluate intelligent tutoring technologies within any instructional domain. Influenced by research surrounding social cognitive theory and cognitive load theory, the resulting experiment tested varying approaches for utilizing an Embodied Pedagogical Agent (EPA) to function as a tutor during interaction in a game-based training environment. Conditions were authored to assess the trade-offs between embedding an EPA directly in a game, embedding an EPA in GIFT's browser-based Tutor–User Interface (TUI), or using audio prompts alone with no social grounding. The resulting data supported the application of using an EPA embedded in GIFT's TUI to provide explicit feedback during a game-based learning event. Analyses revealed conditions with an EPA situated in the TUI to be as effective as embedding the agent directly in the game environment.

1. Introduction

Across education and training communities the use of virtual environments and dynamic game-based training applications is on the rise. When implemented correctly, these programs replicate task features of a domain that elicit realistic human behavior during scenario interactions (Salas, Rosen, Held, & Weissmuller, 2009). This is often represented behaviorally as cognitive decision making, where inputs into a system designate conceptual understanding of task procedures as they relate to the context of a given event. From an educational perspective, these environments are historically effective under two conditions: (1) the learner is fully aware of what is required to meet task standards and adapts his/her approach based on scenario stimuli and implicit feedback occurring naturally in the environment, or (2) a learner's behavior is observed by an instructor and guidance and explicit feedback is provided to assist the learner in overcoming identified misconceptions and impasses (Kluger & DeNisi, 1996; Narciss, 2008). As the former implies expert understanding of all task characteristics, the latter condition is the more prevalent use case of learning in virtual environments.

For years researchers have investigated the use of Embodied Pedagogical Agents (EPAs) in computer-based learning environments to facilitate this more prevalent use case in the absence of live instruction (Veletsianos & Russell, 2014; Yee, Bailenson, & Rickertsen, 2007). EPAs are used as the communication layer within an Intelligent Tutoring System (ITS), where human intervention is replaced by Artificial Intelligence methods (Soliman & Guetl, 2010). The intent is for an agent–learner relationship to mimic Vygotsky's (1987) social-derived theory in that more capable others facilitate the development of an individual's knowledge and skill through active guidance and feedback (Moreno, Mayer, Spires, & Lester, 2001). This capability is important within game-based instructional environments, as a learner is dependent on guidance and feedback to identify/recognize mistakes, confirm correct actions, and to build appropriate schema representations to promote effective transfer of knowledge and skills into a real-world task setting (Narciss, 2008; Shute, 2007).

With advancements in ITSs being able to operate in unison with open-ended virtual environments applied for instructional

* Corresponding author.
E-mail addresses: benjamin.s.goldberg.civ@mail.mil (B. Goldberg), janis.canon-bowers@ucf.edu (J. Cannon-Bowers).

http://dx.doi.org/10.1016/j.chb.2015.05.008
0747-5632/Published by Elsevier Ltd.

purposes (Mall & Goldberg, 2015; Shute, Ventura, Small, & Goldberg, 2013), we are interested in modality effects associated with EPA integration and to what effect mode variations have on performance, perception, and cognitive load. With modeling techniques in place to compare user inputs against a set of specified expert models, the missing pieces are how autonomously generated explicit feedback within game-based training environments influences performance, and how best to interface these feedback channels with the learner. As such, virtual environments integrated with ITS technologies provide an excellent test-bed for studying how explicit feedback can influence human behavior, and what effect that influence has on performance outcomes linked with retention and transfer. This is of relevance as more education and training organizations are using flexible sandbox type virtual environments, such as Unity 3D and Virtual Battle Space 3 (VBS3), to build scenarios for instructing and practicing across multiple domains and skills.

With social learning theory driving the use of EPAs, there have been numerous empirical investigations over the years looking at variables associated with agent design and the resulting effect on metrics of performance, learning, motivation, and perception (Veletsianos & Russell, 2014). This involves examining the type of communications and interactions these agents support (i.e., timing of feedback, specificity of feedback, etc.), and how their appearance and social agency effects outcomes of interest (Shiban et al., 2015). This aligns with Baylor's (2011) description of a genuine agent interaction, which consists of three factors: (1) how the agent appears to the learner, (2) how the agent communicates non-verbally with the learner, and (3) the type of content and discourse supported by the agent. In turn, prior research can be categorized around three themes (Shiban et al., 2015): (1) dialogue design (Johnson et al., 2004; Lin, Atkinson, Christopherson, Joseph, & Harrison, 2013); (2) advanced communication features, like displays of emotion and gestures (Allmendinger, 2010; Olney et al., 2012; Visschedijk, Lazonder, van der Hulst, Vink, & Leemkuil, 2013); and (3) agent appearance (Bailenson, Blascovich, & Guadagno, 2008; Baylor, 2011; Baylor & Kim, 2009; Veletsianos, 2010).

For this study we focused primarily on agent appearance within a highly dynamic open-ended game-based training environment, which is a recognized gap in EPA related research (Gulz & Haake, 2006; Shiban et al., 2015; Veletsianos & Russell, 2014). Existing work on agent appearance primarily associates with what is termed the *persona effect*. In its simplest form, the persona effect posits that the mere presence of a lifelike character in an interactive learning environment can have a significant positive effect on the perception of the learning experience. This is supported in the Baylor and Kim (2009) study that demonstrated a visible agent in a learning environment significantly impacted motivation outcomes when compared to voice and text only feedback conditions. Specifically, the incorporation of social agents based on the persona effect have been found to increase motivation for using a system, as well as stimulate interest in topics across multiple subjects and learning environments (Gulz, 2004; Heidig & Clarebout, 2011; Veletsianos & Russell, 2014). In terms of motivation, a common conclusion from research shows character enhanced systems to report as more entertaining, lively, likeable, or engaging (André & Rist, 2001; Johnson, Rickel, & Lester, 2000; Lester, 2011; Lester et al., 1997a; Shiban et al., 2015). Research continues to examine elements of the persona effect to determine agent characteristics that optimize learning outcomes. This involves both physical elements of appearance (i.e., voice inflection, hairstyle, clothing, ethnicity, gender, etc.) and stereotype perceptions associated with appearance (i.e., usefulness, credibility, and intelligence) (Baylor & Kim, 2005; Liew, Tan, & Jayothisa, 2013; Shiban et al., 2015; Veletsianos & Russell, 2014).

While the persona effect continues to be investigated, what we are most interested in is the mode by which an EPA appears to the learner, and what effect that source modality has on learning outcomes and learner perceptions. To address this approach, principles of cognitive load theory (Oviatt, 2006; Van Merriënboer & Sweller, 2005) and multiple resource theory (Wickens, 2002) are applied to examine how an EPA can be situated during a learning event, in this case a dynamic game-based environment, and what effect that modality has on measures of performance and workload. This is attributable to defining game-based training with instructional ITS supports as a dual task environment; ultimately requiring an individual to maintain awareness and understanding of the task environment while also perceiving, processing, and acting upon feedback information provided by an integrated EPA. This can create competition among available cognitive resources to process information in the learning environment (Craig, Gholson, & Driscoll, 2002).

As this described dual task requires split-attention (Sweller & Chandler, 1994), we apply foundations associated with the modality principle to guide EPA source implementations. The modality principle describes an effect derived from Wicken's (2002) multiple resource theory and Mayer and Moreno's (1998) theory of multimedia learning. It implies that learners can process information more efficiently when material and feedback is presented as a mix of visual and auditory stimuli. The notion is to exploit alternative modes of feedback presentation (e.g., acoustic, visual, etc.) to avoid cognitive overload due to modality effects encountered when presenting guidance information in the same format as other elements in the training environment (Mayer & Moreno, 2002; Shute, 2007).

Empirical evidence supports the modality effect as it relates to processing information during task execution in both multimedia and game-based learning environments (Low & Sweller, 2005). Fiorella, Vogel-Walcutt, and Schatz (2012) conducted a study investigating the modality effect as it applied to real-time feedback within a simulation-based training scenario. Their procedure tested the source modality associated with feedback delivery, where it was presented either as an auditory message or as printed text overlay on the computer screen. Results showed those receiving feedback as an audio source to demonstrate better decision-making performance when compared to the printed text group. But what happens when an EPA is added to this context? In an earlier study Craig et al. (2002) found no split-attention effect when integrating an EPA into a multimedia learning environment. His analysis showed no significant differences on learning outcomes when comparing two conditions, one with spoken feedback and the second with spoken feedback accompanied by an agent. This demonstrates that the agent did not induce additional cognitive load on available resources that would affect performance outcomes.

In addition, Moreno, Mayer, and Lester (2000) ran an experiment looking at the role of an EPA's visual and auditory presence in a discovery learning environment. They based hypotheses on cognitive load theory's modality effect and social cognitive theory's persona effect, predicting students who learn with the voice and image of an agent to remember materials of the lesson better and are more likely to use what they learned to solve problems. Their analysis showed no positive or negative effect on performance as a result from the visual presence or absence of an EPA, but they found students in the agent conditions to consistently report the lesson more favorably, they recalled more information, and reported being more motivated and interested in the program (Moreno et al., 2000). However, both studies were based in static multimedia learning environments, much like other studies that support the modality effect (Atkinson, 2002; Moreno et al., 2001). An additional question relevant to this work is: does this outcome apply outside of multimedia systems and into interactive game-based platforms where visual resources are more strained and have more perceptual cues to attend to?

With new technologies being developed that enable ITS functions within game-based learning platforms, new modalities are available for relaying information to the user. Through a Tutor–User Interface (TUI), an independent visual field that can support agent visualization, content can be presented from an external channel to the training environment. Embedding an EPA in a designated TUI can have one of two effects on game interaction: (1) it provides a grounded base for the visual presence of an EPA, requiring only ambient visual scanning (Wickens, 2002) and reducing load for focused attention on the task environment, or (2) the extra interface creates an associated dual-task in the learning environment requiring a user to monitor both the game and TUI equally to maintain appropriate awareness of the interacting elements, thus introducing additional extraneous cognitive load elements.

Based on this foundation, we present findings from an experiment that examined the effect different source modalities of feedback had on performance within a virtual game-based environment. This research was motivated by consideration of the effort to develop a feedback delivery method versus the impact it had on performance. That is, based on the cost and time required to apply a feedback modality, this study determined the feedback source that produced the greatest impact on retention and transfer. To execute this research, the U.S. Army Research Laboratory's Generalized Intelligent Framework for Tutoring (GIFT) was used as the ITS authoring environment. GIFT is a community driven opensource project aimed at promoting standardized methods for building ITSs and focuses on domain-independency, ease of authoring, and reuse. GIFT is a modular architecture and consists of all working parts common to intelligent tutors, with additional functions built to accommodate its application across multiple training systems. (Goldberg, Brawner, Holden, & Sottilare, 2012; Sottilare, Brawner, Goldberg, & Holden, 2013). When used in a research setting, GIFT provides a stable adaptive training test-bed that supports easy development of experimental conditions for studying modeling techniques and instructional management practices.

The goal of the current effort was to investigate approaches for enhancing game-based training applications through the incorporation of performance-driven feedback functions delivered via GIFT. Specifically, this work examined methods for embedding feedback delivery mechanisms within game environments using GIFT interfacing methods and assessed the influence variations in the source and delivery of feedback had on behavior and performance outcomes.

2. Materials and methods

This work investigated the effect variations in the source of real-time feedback within a game-based training event had on subsequent task performance; the effect the source of feedback had on post-training learning outcomes; and whether variations in feedback source produced reliable differences in trainee self-reported measures of cognitive load. The study went deeper by exploring the impact of delivering feedback through game characters defined as EPAs, and to assess the effect varying agent delivery modalities had on trainee performance. Specifically, this research examined whether there is a significant benefit to embedding EPAs directly into the task environment versus an EPA interacting with the user from an interface external to the game world. In this section, we review the experimental methodology and all associated materials used to investigate these research questions.

2.1. Experimental test-bed

The serious game selected for this study was the Tactical Combat Casualty Care Simulation (TC3Sim), also known as vMedic. TC3Sim utilizes a game-based virtual environment to teach and reinforce the tactics, techniques, and procedures required to successfully perform as an Army Combat Medic and Combat Lifesaver (ECS, 2012). Tasks simulated within TC3Sim include assessing casualties, performing triage, providing initial treatments, and preparing a casualty for evacuation under conditions of conflict. The two scenarios developed for this experiment involved handling injuries in a hostile environment following the detonation of an improvised explosive device.

This virtual environment provides an excellent test-bed as it simulates real-world scenarios that involve on-the-spot decision making. To test the effect varying feedback modalities have on performance outcomes, TC3Sim was integrated with GIFT, enabling real-time performance assessment across a defined set of concepts linked to the simulated tasks performed in the game. This assessment is then used to trigger the delivery of explicit feedback when called for by the intelligent tutor. During gameplay, feedback was provided to participants when their actions met specified threshold conditions for all critical competency measures. These conditions were based on input from combat medic subject matter experts. These thresholds were defined around time, location, and entity data provided directly from the game environment. If specific actions or procedures were violated, a feedback string was delivered as a reflective prompt to notify the learner of actions being ignored or actions incorrectly administered.

2.1.1. The Generalized Intelligent Framework for Tutoring (GIFT)

As described above in the introduction, GIFT is a modular approach to a domain-independent ITS authoring environment (Goldberg et al., 2012; Sottilare et al., 2013). For the purpose of this experiment, GIFT was used to perform real-time assessment on interaction within the serious game environment TC3Sim. This capability was 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. An example is defining a rule for a concept titled 'stay with unit'. The concept requires a user to stay within a defined proximity of their squad leader, which is influenced by entity location data provided by the simulation. If a user is reported as being beyond the defined threshold, then the concept of 'stay with unit' is assessed as below expectation. This performance state is then communicated to GIFT for determining if a feedback intervention is required.

This real-time assessment enables GIFT to detect errors in task performance, which in turn triggers pedagogical interventions intended to influence subsequent behaviors (see Appendix A for a list of the 21 concepts being assessed in GIFT). In the context of this study, GIFT serves as the test-bed 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 unique to GIFT is the Tutor–User Interface (TUI). The TUI is a browser-based communication layer designed to collect user inputs and to relay information back to the user. In terms of providing real-time guided instruction, the TUI can be used as a tool for delivering explicit feedback content. It supports multimedia applications and the presence of virtual entities acting as defined tutors.

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.

2.2. Metrics

2.2.1. Knowledge and skill assessments

Two forms of performance measures were collected. The initial metric, learning gains, was based on performance generated on the administered post-test assessing knowledge levels in TC3 concepts, with a subject's pre-test score being defined as a covariate. Items were based on the instructional categories of technical skills (e.g., basic anatomy, physiology, pathology), tactical skills (e.g., move, shoot, communicate), and clinical skills (e.g., assess, diagnose, treat, evacuate). Each test included 15 multiple choice questions to assess the various knowledge components.

The second performance metric came directly from the TC3Sim game environment. This includes performance within a guided training scenario with a GIFT EPA and performance within an assessment capstone scenario. Interaction was monitored and logged via GIFT, and player actions were measured against defined expert models. Performance was based on observed procedures during game play, and 'go'/'no-go' determinations were marked across all defined critical competency measures (e.g., security sweep, tourniquet application, dress bleed, etc.). The metric output consisted of the number of correct actions taken within the scenario in relation to the full set of competencies being monitored. In accordance with the analysis proposed for the knowledge post-test, the in-game capstone performance analysis designated training scenario scores as a covariate.

2.2.2. Workload and mental demand metrics

Measures of an individual's subjective workload and mental demand were recorded following interaction with the guided TC3Sim scenario. For this purpose, each participant completed the NASA-TLX. A participant's overall workload was determined by a weighted average of responses across six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). Definitions of each subscale were provided to participants to reduce uncertainty associated with scale meanings. The instrument was selected because it shows good face and construct validity (Cao, Chintamani, Pandya, & Ellis, 2009), and has been found to meet criteria associated with effective workload assessment techniques: sensitivity, diagnostic capabilities, selectivity, low intrusiveness, reliability, and ease of administration (Rubio, Díaz, Martín, & Puente, 2004).

2.3. Experimental design

The design for this experiment was a counter-balanced mixed design with two independent variables (IV), source of feedback and character profile. Source of feedback had two levels; it refers to the interfacing component that relays feedback information to the user. In the context of this experiment, source conditions were described as being internal or external to the training environment. These conditions incorporated EPAs as interfacing characters, which were present either in the game environment as an entity part of the scenario or housed externally from the game in the GIFT TUI (see Fig. 1). The second IV, character profile, was based on an associated description of the EPA's background and role within the scenario, and was centered around research on social cognitive theory's persona effect (Baylor & Kim, 2005; Lester et al., 1997b; Veletsianos, 2010). The results presented will address analyses examining the first IV alone.

For the purpose of assessing the effect manipulated variables had on associated dependent measures, there was the need for baseline conditions to determine effect size. To achieve this, there were two control conditions. The first control involved the initial TC3Sim guided scenario without any tutor interaction or explicit feedback. This is how TC3Sim is currently implemented, with no real-time interpretation of results and performance; feedback is provided within an After-Action Review (AAR) following scenario completion. The second control incorporated the initial TC3Sim guided scenario with feedback provided solely as an audio message. This condition is being termed 'Voice of God' (VoG) as there

Fig. 1. Variable source modality conditions.

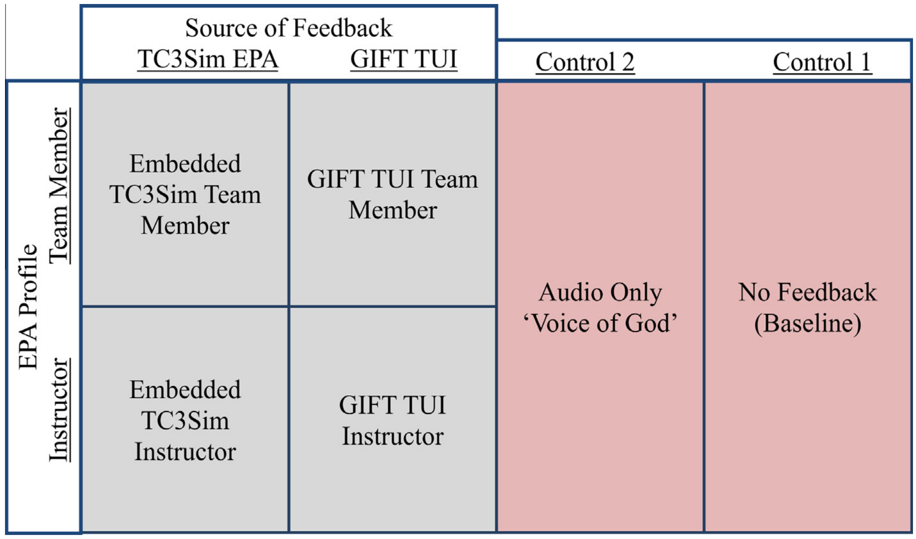

Fig. 2. Experimental conditions.

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. It is important to note that assessments and feedback scripts were consistent across all treatments. This resulted in six total conditions (see Fig. 2).

2.3.1. Hypotheses

In terms of performance related measures, it was hypothesized that the five conditions including real-time explicit feedback (i.e., participants who receive feedback during interaction with TC3Sim) would produce greater knowledge and skill outcomes in comparison to the baseline condition with only implicit environmental feedback. In addition, it was hypothesized that all conditions with interactive EPAs would produce significantly higher performance metrics when compared to both defined control conditions. It was expected that participants receiving explicit feedback from an EPA would show greater performance during the training scenario and larger learning gains as demonstrated by transfer assessments of game performance and pre-/post-test scores.

For cognitive load and mental demand based measures, it was hypothesized there would be significant differences in reported Workload (WL) and Mental Demand (MD) during TC3Sim interaction across EPA source conditions. Variations in feedback source modality were believed to affect the allocation of cognitive resources based on where the EPA was situated in the learning environment. In terms of this hypothesis, we pose two contradicting predictions based on tenets of cognitive load theory. For Prediction 1, we reasoned that reported MD and overall WL would be greatest in conditions where the EPA is present in GIFT's TUI. This was based on users having to allocate visual resources to maintain awareness of the EPAs presence, while managing complex game events. It was also expected that conditions including an EPA would score higher on MD and WL when compared to the VoG treatment, as these subjects would not have the additional visual resources for which they needed to maintain awareness.

In contrast, Prediction 2 was defined around associations linked with multiple resource theory. Based on Wickens (2002) description of ambient vision, information perceived through peripheral visual channels allows individuals to maintain a sense of orientation with that source while maintaining focus on the primary task; as seen in the Liggett, Reising, and Hartsock (1999) study. Hence, an additional question is whether an EPA situated directly in the game environment requires extra focal attention to locate among other objects in the scenario. Because the EPA is not in a static location like the TUI, load on the visual resources may increase to maintain orientation of where the agent is. If this is the case, then the prediction is reversed from number one, with expectations of WL and MD scores reporting higher in the internal feedback source condition when compared to the external TUI scores.

2.4. Participants

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

Data collection was conducted over a five-day period at USMA where a total of 131 subjects participated. This resulted in 22 participants for each experimental condition minus the control, which totaled at 21 subjects. Across all participants, 105 were male and 26 were female, and 108 were Plebes (i.e., freshmen) and 23 were Yearlings (i.e., sophomores). In addition, questions were administered to gauge an individual's videogame experience (VGE), with majority ranking (95 participants) themselves as having moderately low to no experience, and the remaining subjects (36 participants) ranking themselves as having moderately high to high experience. Based on the variability across this metric, VGE was be considered as a covariate within statistical analyses linked around game interaction.

2.5. Procedure

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

A participant was first prompted to complete a battery of surveys. Instruments included a demographics survey and a videogame experience metric. When complete, the pre-test assessing initial knowledge levels was administered. The test included questions across all associated training objectives covered in the game. This initial performance metric was used to determine learning gains following interaction with the training materials. Next, GIFT directed the participant to interact with 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 min with the materials.

Following training, GIFT initialized the first interaction with the TC3Sim interface environment. Participants performed a short scenario designed to introduce the interfaces and inputs associated with the game. This tutorial session lasted an average of 3 min and took no longer than 5 min. 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. Incorporated with this overview was an introduction to the EPA with which the participant would interact. 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 overview and EPA background narrative led directly into the first of two scenarios used to train and test TC3 procedures. 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 the results for executing feedback scripts. Based on the condition, feedback was delivered either as audio only (VoG condition), through an EPA present in GIFT's TUI, through a character present in the virtual game environment, or not at all. Upon completion, participants answered survey instruments on cognitive load (NASA-TLX). This led into the second of two scenarios in TC3Sim, which involved similar events to the first session, 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.

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

3. Results

Statistical analyses were performed using IBM SPSS Statistics 19. For indication of statistical significance, an alpha value of .05 was used for all tests, unless explicitly stated otherwise. For a list of descriptive statistics on associated performance metrics across conditions of interest, see Table 1.

3.1. Performance-based analysis

The first hypothesis examined what effect the inclusion of feedback within a game-based training event had on performance outcomes in both knowledge- and skill-based assessments. It was hypothesized that individuals receiving explicit feedback aimed at improving performance during game-play would produce higher performance scores for all game interaction as well as achievement on post-test scores. Statistical tests were conducted looking at the source of feedback variable to determine if they had an effect on any associated performance outcomes. Results will be individually reported across game related data and pre-/post-test calculated scores. For a visual representation of performance outcomes across game-related metrics and knowledge test scores, see Fig. 3.

Table 1
Descriptive statistics comparing conditions with and without an EPA on performance across game interactions and administered knowledge tests.

EPA, VoG, or no feedback		TC3Sim scenario%		Knowledge pre-test	Knowledge post-test
		Training	Capstone		
EPA-TC3Sim embedded (*N* = 44)	*M*	37.56	39.41	63.09	68.29
	SD	6.92	8.27	12.57	12.89
EPA-GIFT tutor user interface (*N* = 44)	*M*	37.49	40.56	66.20	69.92
	SD	6.59	9.05	9.13	13.12
VoG (*N* = 22)	*M*	40.91	39.09	63.03	60.00
	SD	4.61	7.60	11.77	13.80
No feedback (*N* = 21)	*M*	32.19	35.43	58.73	61.90
	SD	6.98	8.03	10.25	18.64

3.1.1. TC3Sim training and capstone scenarios

As an initial starting point, analysis was conducted to confirm the explicit feedback provided by GIFT actually impacted performance scores. This was addressed by examining performance outcomes within the TC3Sim training scenario, and grouping individuals in the analysis as whether they received or did not receive explicit feedback during gameplay. To test this, a Univariate Analysis of Co-Variance (ANCOVA) was run comparing the two groups. For this analysis VideoGame Experience (VGE) was defined as a covariate. Results showed the inclusion of explicit feedback, regardless of the source, to have a significant main effect on training scenario performance, (F (1, 129) = 11.749, p = .001, η_p^2 = .05, power = 0.925), with VGE reporting as a significant covariate, (F (1, 122) = 5.312, p < .025, η_p^2 = .040, power = 0.628). This relationship shows those scoring higher on VGE produced higher performance during training scenario interaction (Pearson r = .218).

Next, analyses were conducted examining the influence an EPA had on performance and retention scores across both game and knowledge assessments. The first test performed was to examine the effect an EPA had on performance within the training scenario alone. This is differentiated from the analysis above in that it takes into account the VoG condition to determine if performance between these two design treatments is significantly different. A Univariate ANCOVA was run across the three defined groups, with VGE defined as the covariate. The output shows the conditions relating to interaction with an EPA, VoG, or No Feedback to produce significant differences in performance outcomes, (F (2, 129) = 8.28, p < .001, η_p^2 = .117, power = 0.958), along with VGE reporting as a significant covariate, (F (1, 129) = 4.356, p < .05, η_p^2 = .034, power = 0.544). To examine further, post hoc analysis was performed using the Bonferroni test, with results showing both the EPA and VoG groups to score significantly higher than the No Feedback condition (see Table 2). However, no significant difference was found between the EPA and VoG groupings.

Next, analyses were conducted to examine participants' subsequent performance within the capstone scenario. A mixed

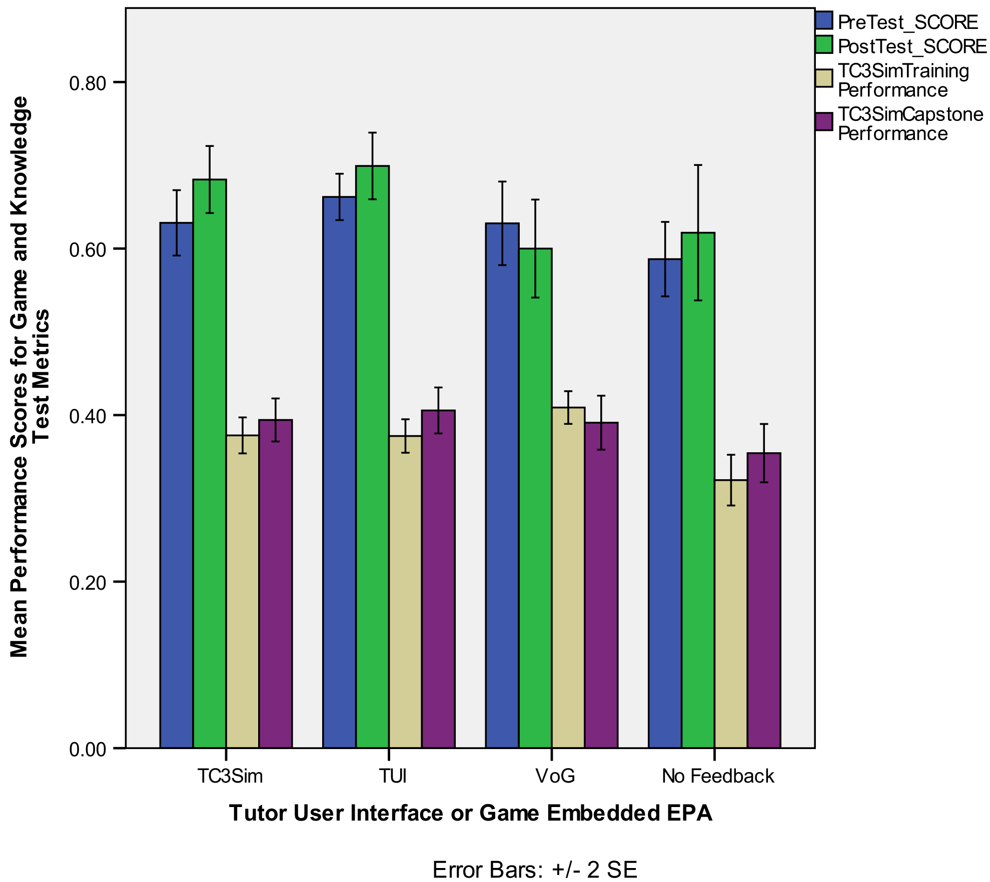

Fig. 3. Performance outcomes on game and test oriented metrics.

Table 2
Post-hoc analysis of training scenario performance across EPA treatments.

EPA, VoG, or no feedback	TC3Sim scenario%		Significance
	Mean	Standard error	
EPA vs. no feedback	37.2	.007	$p = .01$
	32.4	.014	
VoG vs. no feedback	40.6	.014	$p < .001$
	32.4	.014	

between-within subjects ANOVA was run examining differences in performance gains across the two game scenarios and to determine if the feedback source had an influence on the associated outcomes. Results showed no significant differences for within-subject interaction between scenario and experimental condition (F (1, 125) = 2.572, $p = .080$, $\eta_p{}^2 = .040$, power = 0.505). However, the test revealed a significant between subjects main effect across conditions in terms of TC3Sim performance as deemed by the scores across the two scenarios (F (2, 128) = 4.520, $p < .025$, $\eta_p{}^2 = .066$, power = 0.762).

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

The final test looked at the location of the EPA during gameplay (TUI vs. Game-Embedded) and to determine if there was an effect on resulting performance outcomes. Because all of the EPA conditions incorporated explicit feedback, it is predicted that there would be no significant differences in outcomes as a result of where the EPA was positioned. As this is the only aspect of the experimental procedure where a tutor was present, this analysis focused solely on training scenario outcomes. A Univariate ANCOVA was performed based around the TUI-embedded and TC3Sim-embedded EPA groupings, with VGE defined as a covariate. As predicted, the results showed no significant differences in training performance when comparing a tutor in the TUI ($M = 37.3$. $SE = .011$) versus being embedded in the game environment ($M = 37.1$, $SE = .011$; F (1, 86) = .023, $p = .879$, $\eta_p{}^2 = .000$, power = 0.053). As seen in the groups associated means, there was minimal variance in performance outcomes as a result of where the EPA was located during game interaction.

3.1.2. Knowledge post-test outcomes

Following examination of game-based performance metrics, analyses were performed on outcomes from the two knowledge tests administered at the beginning and end of the experimental session. First, a mixed between/within subjects ANOVA was run looking at the differences in performance across the pre- and post-test knowledge scores. A visual graphic of these performance metrics can be seen in Fig. 3. In examining the statistical outputs, results show no significant within subject interaction between

Pre-/Post-Test administration and the source conditions, (F (2, 128) = 2.413, $p < .094$, $\eta_p^2 = .036$, power = 0.479). However, a significant between subject main effect for Experimental Condition was identified (F (2, 128) = 4.520, $p < .025$, $\eta_p^2 = .066$, power = 0.7626) based on a transformed variable computed by averaging an individual's two test scores.

Because of the identified significant between subjects main effect, post hoc analysis was conducted to identify the conditions to produce reliable differences associated with knowledge learning gains. To account for performance scored on the pre-test, a Univariate ANCOVA was performed, with the pre-test score being defined as a covariate. Results showed the source condition to have a significant main effect on the knowledge post-test scores (F (2, 127) = 4.028, $p < .025$, $\eta_p^2 = .060$, power = 0.710), with an individual's pre-test score showing as a significant covariate (F (1, 127) = 12.975, $p < .001$, $\eta_p^2 = .093$, power = 0.947). As found above in game performance, an individual's score on the knowledge pre-test was found to strongly predict their performance on the subsequent post-test, regardless of the condition (Pearson's $r = .321$, $p < .001$). To examine further, post hoc analysis was performed with the Bonferroni test, resulting in an identified significant difference on post-test performance between those interacting with an EPA ($M = 68.86$, $SE = .014$) and those in the VoG condition ($M = 60.00$, $SE = .029$; $p = .026$). While the EPA conditions outperformed the No Feedback by more than five percentage points, there was no significant difference found as a result of the ANCOVA.

Table 3
Experimental workload and mental demand metrics across conditions.

Feedback modality condition		NASA-TLX results	
		Workload	Mental demand
EPA-TC3Sim embedded (*N* = 44)	*M*	56.77	80.82
	SD	8.72	12.97
EPA-GIFT tutor user interface (*N* = 44)	*M*	55.27	84.55
	SD	12.45	18.85
VoG (*N* = 22)	*M*	55.65	73.86
	SD	9.53	13.89
No feedback (*N* = 21)	*M*	52.37	85.33
	SD	13.24	15.89

3.2. Cognitive load and mental demand based analyses

Due to experimental conditions involving variations in the game-tutor interface design, analyses were conducted examining an individual's Mental Demand (MD) and associated WorkLoad (WL). The theoretical foundation associated with analysis is based on research surrounding cognitive load theory and multiple resource theory (Oviatt, 2006; Wickens, 2002). The results presented are based on self-reported WL and MD metrics collected from the NASA-TLX directly following the TC3Sim training scenario. Due to time limitations with the subject pool, we were unable to re-administer the NASA-TLX following the capstone scenario to determine if further exposure to the game reduces the perceived amount of effort to perform effectively. As a result, statistical tests examine the relationship between the IV of feedback source modality and its impact on WL and MD within only one of the two scenarios. For a list of descriptive statistics on associated WL and MD metrics across each source modality condition, see Table 3.

To establish if there were reliable differences in reported WL and MD scores across treatments, two separate Univariate ANOVAs were performed on each of the cognitive load metrics (see Fig. 4 for visual representation). Results show the overall WL metric (i.e., metric computed from all six dimensions of

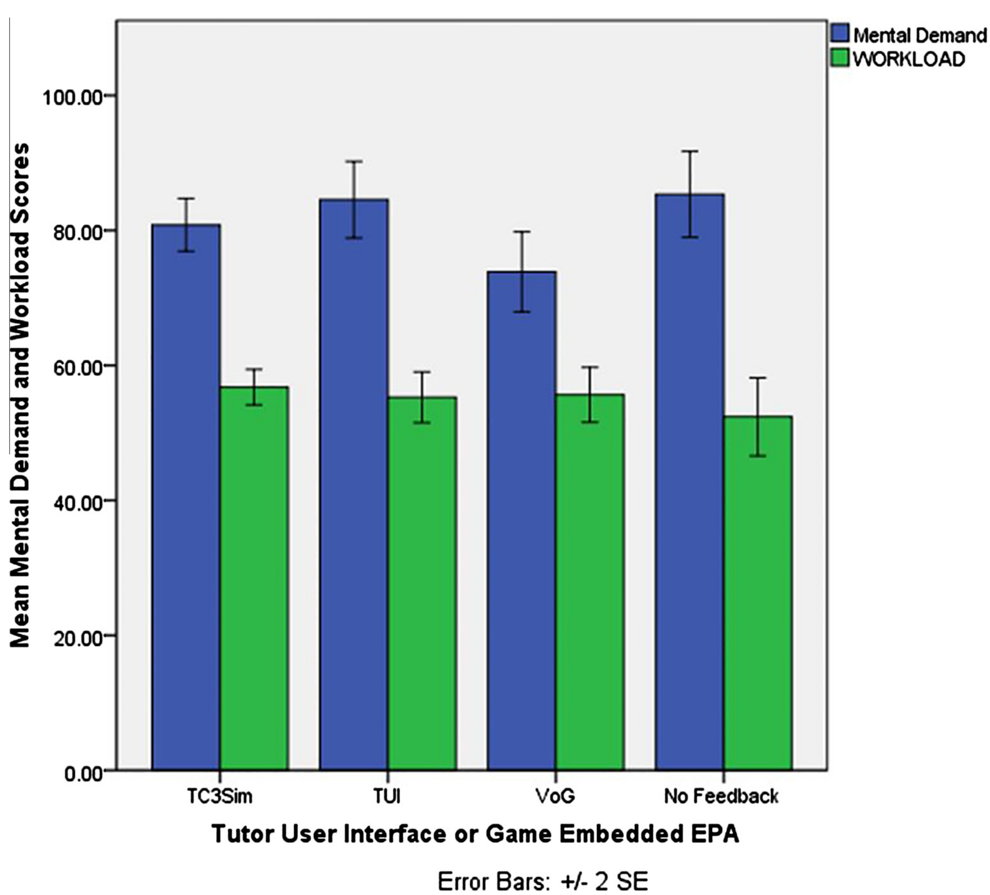

Fig. 4. Mental demand and workload scores across source modality conditions.

NASA-TLX) to reveal no significant differences between conditions ($F\,(2, 107) = .235$, $p = .791$, $\eta_p^2 = .004$, power = 0.086), while the MD metric showed reliable differences as a result of whether a participant interacted with the TUI-Embedded tutor, the TC3Sim-Embedded tutor, or the VoG condition ($F\,(2, 107) = 3.373$, $p < .05$, $\eta_p^2 = .059$, power = 0.625). To examine further, planned comparisons were performed to determine the specific treatments contributing to this statistical finding. Outcomes from these tests showed both the TUI-Embedded tutor ($M = 84.54$, $SD = 18.85$) conditions and TC3Sim-Embedded tutor ($M = 80.82$, $SD = 12.97$) conditions to report significantly higher MD scores when compared to the VoG condition ($M = 73.86$, $SD = 13.89$), while no reliable differences were found between the varying EPA source modalities.

With a baseline condition providing no explicit feedback during the TC3Sim training scenario, the next set of analyses focused on examining if those relying solely on implicit information from the game to gauge performance would report significantly higher WL and MD scores when compared to those receiving performance-driven feedback. Two analyses were conducted to test this hypothesis. The first was a Multivariate ANOVA (MANOVA) looking at both WL and MD against the two defined groups of Feedback and No Feedback. Results from the MANOVA show no significant differences between the two groups for both metrics (MD: $F\,(1, 129) = 1.364$, $p = .245$, $\eta_p^2 = .010$, power = 0.213; WL: $F\,(1, 129) = 1.886$, $p = .172$, $\eta_p^2 = .014$, power = 0.245). The next analysis looked at each of the EPA and VOG source conditions against those receiving no feedback through a MANOVA. As seen in all results associated with NASA-TLX data, the metric of WL showed no significant differences between the individual conditions ($F\,(3, 127) = .765$, $p = .516$, $\eta_p^2 = .018$, power = 0.210). However, results from the MANOVA on MD show significantly reliable differences between conditions ($F\,(3, 127) = 2.771$, $p = .044$, $\eta_p^2 = .061$, power = 0.658), yet post hoc tests showed contradicting condition comparisons. Bonferroni post hoc analysis revealed no conditions to report reliable differences, while Fisher's Least Significance Difference reported reliable differences when comparing the VoG ($M = 73.86$, $SE = 3.33$) condition to both the No Feedback ($M = 85.33$, $SE = 3.41$; $p < .025$) and the EPA TUI conditions ($M = 84.55$, $SE = 3.32$; $p < .025$).

4. Discussion

With results supporting the application of a real-time feedback function in the game TC3Sim, the primary focus of the discussion is to address the impact of EPA feedback modalities on dependent measures of interest. We identify causal relationships that will influence future application of ITS features in game-based training events, as well as direct follow-on research efforts. A goal is to better understand the influence of ITS feedback in game-based environments, and to identify tradeoffs between the various EPA interfacing modalities and their influence on performance outcomes.

In examining the effect an EPA has on performance within the game TC3Sim, it was found that individuals within the VoG condition scored highest in the TC3Sim training scenario when compared against all EPA related treatments. From this perspective, the inclusion of an EPA shows no true benefit. Individuals who received feedback prompts as audio alone performed the best, but results were not significantly better than those with EPA treatments. This outcome aligns with similar studies investigating the impact an EPA has on real-time training performance outcomes (Höök, Persson, & Sjölinder, 2000; Moundridou & Virvou, 2002; Van Mulken, André, & Müller, 1998).

The real insight of an EPA's effect on performance is observed in examining outcomes on subsequent assessments (i.e., capstone scenario and knowledge post-test). According to Schmidt and Bjork (1992) it is critical to add transfer and retention phases when comparing treatment conditions on learning effect, as these subsequent measures are often better indicators of an independent variables influence on performance differences between groups. In these analyses, the EPA conditions were found to perform significantly better than the VoG. The results from this analysis indicate that the presence of an EPA during game interaction led to better outcomes on subsequent interaction within similar problem spaces, leaving the VoG condition as the only treatment to produce negative learning gains and transfer across both the game and knowledge-test metrics. Hence, while VoG was shown to result in the highest performance outcomes in the TC3Sim training scenario, this treatment was shown to have the weakest transfer to alternate problems and retention of domain related facts.

This finding supports the persona effect, demonstrating that grounding feedback through a social source aids in perception of information and management of short- and long-term memory, resulting in better conceptual understanding of the material (Gulz, 2004; Veletsianos & Russell, 2014). This observed outcome of an EPA's effect on transfer assessments is consistent with prior research in the computer-based instruction field (Graesser, VanLehn, Rosé, Jordan, & Harter, 2001; Moreno et al., 2001). A remaining question is, what is the causal effect associated with these variations in performance?

While the EPA and VoG conditions incorporated the same assessment and feedback techniques, what is the underlying cause for significant differences in performance outcomes as seen in the capstone scenario and knowledge post-test measures? In analyzing the causal effects related to this question, very little prior research was found that could lend insight into the observed trends. A distinct difference between these conditions is the priming interaction delivered to participants in the EPA groupings. For the context of this research, priming is defined as any set of interactions experienced prior to a training event that implicitly influences the processing of material and communications (Cleeremans, 2001). In the EPA instances the participant is introduced to the virtual tutor agent with a narrative describing their purpose. This direct priming informs the learner that information would be provided in real-time by a virtual tutor that ties their performance directly to a domain concept being assessed. In the VoG condition, no such introduction is ever experienced. In fact, the participant is not primed at all with respect to an ITS running assessment and managing feedback functions. From this perspective, the performance-based feedback given in the VoG group can be perceived as if it is directly part of the scenario, being implicitly delivered as natural component of the game environment (Narciss, 2008).

This finding may assist in explaining why individuals in the VoG condition scored the highest during the training scenario, while producing the worst transfer and retention results on subsequent assessments. In the VoG treatment, participants are reacting to feedback provided by GIFT as if it is part of the game, due to removal of the EPA introduction that notifies the subject explicit information will be provided. As the feedback is frequent and performance-driven, the content is intended to guide the trainee toward a correct behavior. However, frequent feedback in this context can also have negative consequences in terms of learning and retention (Schmidt, 1991; Schmidt & Bjork, 1992). If feedback comes to be an inherent part of the task that is implicitly perceived, as is evident in the VoG condition, then performance is disrupted in retention when the feedback is ultimately removed from the execution environment. Schmidt and Bjork (1992) also mention that frequent feedback can block information processing activities that are necessary to the knowledge acquisition phase for producing effective response when assessed on retention. This highlights a dependency on feedback to successfully execute a task rather than use it as a learning aid to better formulate mental models of acceptable behavior.

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

Based on this association, it appears to be beneficial to provide upfront information to the learner that feedback will be provided linking game interaction to overall learning objectives the system is designed to train. This may assist the learner in associating formative feedback information with knowledge schemas in memory for correcting or reaffirming knowledge components (Shute, 2007). An additional prediction posed to Hypothesis 2 was that subjects in the baseline No Feedback treatment would report the highest WL and MD scores due to relying on implicit information from the game alone to gauge performance toward meeting objectives. Similarly to all EPA conditions, the No Feedback condition reported higher MD scores when compared against the VoG condition, with no significant differences seen between the control and the feedback source modality treatments.

4.1. Future work

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

In addition, further investigation is required looking at the impact of upfront communications priming the learner for explicit feedback functions in the training environment. While the VoG condition received the same audio prompts without the presence of an EPA, we need further investigation to determine if the EPA truly influenced the perception of feedback information. As game-based learning applications evolve to incorporate automated feedback functions, further research is required to determine how best to prepare a learner cognitively so as to efficiently process information both implicitly in the game and through explicit information that associated action with intended objective outcomes. Lastly, analyses revealed a relationship between videogame experience and recorded performance outcomes across the TC3Sim scenarios. This warrants further investigation to better understand how videogame experience influences learning in such environments. A direct question of interest is determining if an individual's experience should influence the design of game-based training and what type of feedback should be provided.

5. Conclusion

The aim of this research was to explore available tools for integrating intelligent tutoring communications in game-based learning platforms and to examine theory-based techniques for delivering explicit feedback based on individualized performance. Influenced by research surrounding social cognitive theory and cognitive load theory, the experiment presented tested varying approaches for utilizing an EPA to function as a tutor during interaction in a game-based environment. Conditions were authored to assess the tradeoffs between embedding an EPA directly in the game environment, embedding an EPA in GIFT's browser-based TUI, or using audio prompts alone with no social grounding. Although not all predictions were supported by the resulting data, the application of using an EPA in the TUI to provide feedback during learning was found to be as effective as embedding the agent directly in the game environment.

This inference is based on evidence showing reliable differences across conditions on the metrics of performance and self-reported mental demand. The overarching finding is that feedback, regardless of being delivered by an EPA, significantly improved performance in the training scenario. However, those assigned to an EPA condition were found to perform significantly better on transfer assessments when compared against subjects assigned to the audio alone condition (e.g. VoG). This finding supports previous research concerning the application of social agents in technology-based learning platforms.

Acknowledgments

This research was funded by the U.S. Army Research Laboratory – Human Research and Engineering Directorate's (ARL-HRED) Simulation and Training Technology Center (STTC) under the Adaptive Training research program. We would like to thank the Engineering Psychology department at the United States Military Academy at West Point for supporting this work and assisting with data collection and analysis.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chb.2015.05.008.

References

Allmendinger, K. (2010). Social presence in synchronous virtual learning situations: The role of nonverbal signals displayed by avatars. *Educational Psychology Review, 22*(1), 41–56.

André, E., & Rist, T. (2001). Presenting through performing: On the use of multiple lifelike characters in knowledge-based presentation systems. *Knowledge-Based Systems, 14*(1), 3–13.
Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94*(2), 416.
Bailenson, J. N., Blascovich, J., & Guadagno, R. E. (2008). Self-representations in immersive virtual environments1. *Journal of Applied Social Psychology, 38*(11), 2673–2690.
Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development, 59*(2), 291–300.
Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education, 15*(1), 95–115.
Baylor, A. L., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25*(2), 450–457.
Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*(1), 113–117.
Cleeremans, A. (2001). Conscious and unconscious processes in cognition. *International Encyclopedia of the Social and Behavioral Sciences*, 2584–2589.
Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology, 94*(2), 428.
ECS (2012). vMedic Retrieved August 5, 2012, from <http://www.ecsorl.com/products/vmedic>.
Fiorella, L., Vogel-Walcutt, J. J., & Schatz, S. (2012). Applying the modality principle to real-time feedback and the acquisition of higher-order cognitive skills. *Educational Technology Research and Development, 60*(2), 223–238.
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.
Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine, 22*(4), 39.
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.
Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents—A look at their look. *International Journal of Human-Computer Studies, 64*(4), 322–339.
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.
Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6*(1), 27–54.
Höök, K., Persson, P., & Sjölinder, M. (2000). Evaluating users' experience of a character-enhanced information space. *AI Communications, 13*(3), 195–212.
Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*(1), 47–78.
Johnson, W. L., Rizzo, P., Bosma, W., Kole, S., Ghijsen, M., & van Welbergen, H. . *Generating socially appropriate tutorial dialog. Affective Dialogue Systems*. Springer.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.
Lester, J. C. (2011). Affect, learning, and delight. In S. D'Mello, A. Graesser, B. Schuller & J.-C. Martin (Eds.), *Proceedings of the 4th international conference on Affective Computing and Intelligent Interaction (ACII 2011)*, LNCS (vol. 6974, pp. 2). Berlin Heidelberg: Springer-Verlag.
Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997a). *The persona effect: affective impact of animated pedagogical agents*. Paper presented at the SIGCHI conference on human factors in computing systems.
Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997b). *The persona effect: affective impact of animated pedagogical agents*.
Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Educational Technology & Society, 16*(4), 275–286.
Liggett, K. K., Reising, J. M., & Hartsock, D. C. (1999). Development and evaluation of a background attitude indicator. *The International Journal of Aviation Psychology, 9*(1), 49–71.
Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education, 67*, 239–249.
Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. *The Cambridge Handbook of Multimedia Learning, 147*, 158.
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).
Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312.
Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*(1), 107–119.
Moreno, R., Mayer, R. E., & Lester, J. C. (2000). *Life-Like Pedagogical Agents in Constructivist Multimedia Environments: Cognitive Consequences of their Interaction*. Paper presented at the world conference on educational mulitmedia, hypermedia and telecommunications, Chesapeake, VA.
Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177–213.
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.
Narciss, S. (2008). Feedback strategies for interactive learning tasks. In J. M. Spector, M. D. Merrill, J. J. G. v. Merrienboer & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 125–144). New York: Lawrence Erlbaum Associates Inc.
Olney, A. M., D'Mello, S., Person, N., Cade, W., Hays, P., Williams, C., Graesser, A. (2012). *Guru: A computer tutor that models expert human tutors*. Paper presented at the Intelligent Tutoring Systems.
Oviatt, S. (2006). *Human-centered design meets cognitive load theory: designing interfaces that help people think*. Paper presented at the 14th annual ACM international conference on multimedia.
Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology, 53*(1), 61–86.
Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009). Performance measurement in simulation-based training. *Simulation & Gaming, 40*(3), 328–376.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Bulletin, 3*(4), 207–217.
Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor neuroscience* (pp. 59–75). Dordrecht, The Netherlands: Kluwer.
Shiban, Y., Schelhorn, I., Jobst, V., Hörnlein, A., Puppe, F., Pauli, P., et al. (2015). The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior, 49*, 5–11.
Shute, V. J. (2007). *Focus on formative feedback*. Princeton, NJ: Educational Testing Service.
Shute, V., Ventura, M., Small, M., & Goldberg, B. (2013). Modeling student competencies in video games using stealth assessment. In R. Sottilare, A. Graesser, X. Hu & H. Holden (Eds.), *Design Recommendations for Intelligent Tutoring Systems*, *Vol. 1: Learner Modeling* (pp. 141–152).
Soliman, M., & Guetl, C. (2010). *Intelligent pedagogical agents in immersive virtual learning environments: A review*. Paper presented at the MIPRO, 2010 proceedings of the 33rd international convention.
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., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.
Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147–177.
Van Mulken, S., André, E., & Müller, J. (1998). *The persona effect: how substantial is it People and computers XIII* (pp. 53–66), Springer.
Veletsianos, G. (2010). Contextually relevant pedagogical agents: Visual appearance, stereotypes, and first impressions and their impact on learning. *Computers & Education, 55*(2), 576–585.
Veletsianos, G., & Russell, G. S. (2014). Pedagogical agents. *Handbook of research on educational communications and technology* (pp. 759–769). Springer.
Visschedijk, G. C., Lazonder, A. W., van der Hulst, A., Vink, N., & Leemkuil, H. (2013). Modelling human emotions for tactical decision-making games. *British Journal of Educational Technology, 44*(2), 197–207.
Vygotsky, L. (1987). Zone of proximal development. *Mind in society: The development of higher psychological processes*, pp. 52–91.
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177.
Yee, N., Bailenson, J. N., & Rickertsen, K. (2007). *A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces*. Paper presented at the SIGCHI conference on human factors in computing systems.