Experimentation with the Generalized Intelligent Framework for Tutoring (GIFT): A Testbed Use Case

Benjamin Goldberg[1] and Jan Cannon-Bowers[2]

[1] *United States Army Research Laboratory-Human Research & Engineering Directorate-Simulation and Training Technology Center, Orlando, FL 32826*
{benjamin.s.goldberg@us.army.mil}
[2] *Center for Advanced Medical Learning and Simulation (CAMLS) at the University of South Florida, Tampa, FL 33602*
{jcannonb@health.usf.edu}

Abstract. Computer-Based Tutoring Systems (CBTS) are grounded in instructional theory, utilizing tailored pedagogical approaches at the individual level to assist in learning and retention of associated materials. As a result, the effectiveness of such systems is dependent on the application of sound instructional tactics that take into account the strengths and weaknesses of a given learner. Researchers continue to investigate this challenge by identifying factors associated with content and guidance as it pertains to the learning process and the level of understanding an individual has for a particular domain. In terms of experimentation, a major goal is to identify specific tactics that impact an individual's performance and the information that manages their implementation. The Generalized Intelligent Framework for Tutoring (GIFT) is a valuable tool for this avenue of research, as it is a modular, domain-independent framework that enables the authoring of congruent systems that vary in terms of the research questions being addressed. This paper will present GIFT's design considerations for use as an experimental testbed, followed by the description of a use case applied to examine the modality effect of feedback during game-based training events.

Keywords: generalized intelligent framework for tutoring, instructional strategies, testbed, feedback, experimentation

1 Introduction

The overarching goal of Computer-Based Tutoring Systems (CBTSs) is to enable computer-based training applications to better serve a leaner's needs by tailoring and personalizing instruction [1]. Specifically, the goal is to achieve performance benefits within computer-based instruction as seen in Bloom's 1984 study "the 2-Sigma Problem". Though there is recent debate on the validity of these results [2], this classic experiment showed that individuals receiving one-on-one instruction with an expert tutor outperformed their fellow classmates in a traditional one-to-many condition by an average of two standard deviations. The success of this interaction is in the ability of the instructor to tailor the learning experience to the needs of the individual. Interaction is based on the knowledge level of the learner as well as their performance and reaction (i.e., cognitive and affective response) to subsequent problems and communications [3].

With the recent development of the Generalized Intelligent Framework for Tutoring (GIFT; see Figure 1), a fundamental goal is to develop a domain-independent pedagogical model that applies broad instructional strategies identified in the literature. This framework would then be used to author adaptive environments across learning tasks to produce benefits accrued through one-on-one instruction. At the core of GIFT is pedagogical modeling, which is associated with the application of learning theory based on variables empirically proven to influence performance outcomes [4]. According to Beal and Lee [5] the role of a pedagogical model is to balance the level of guidance and challenge during a learning event so as to maintain engagement and motivation. The notion for GIFT is to identify generalized strategies on both a macro- and micro-adaptive level that can be used to author specific instructional tactics for execution in a managed ITS application. The pedagogical model uses data on '*Who*' is being instructed, '*What*' is being instructed, and the '*Content*' available from which to instruct. In an ideal case, GIFT can identify recommended strategies based on this information, and also provide tools to convert those strategies into specific instructional tactics for implementation.

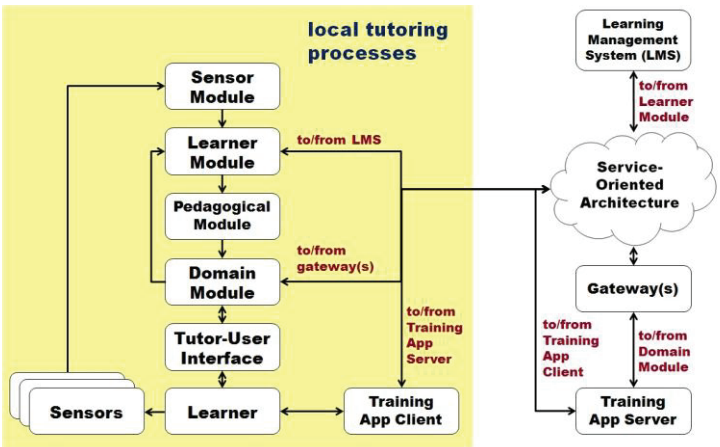

Fig. 3. Generalized Intelligent Framework for Tutoring (GIFT)

Before this conceptual approach of GIFT can be realized, a great deal work needs to be done to identify strategies found to consistently affect learning across multiple domains (codified in the pedagogical model) and the variables that influence the selection of these strategies (expressed in the learner model). In the remainder of this paper, we describe GIFT's functional application as an experimental testbed for conducting empirical research, followed by a descriptive use case of a recent instructional strategy-based experiment examining the effect varying modalities of feedback delivery have on learner performance and engagement within a game-based environment.

1.1 GIFT's Testbed Functionality

For GIFT to be effective across all facets of learning, there are a number of research questions that need to be addressed. These include, but are not limited to: (1) How can GIFT be used to manage the sequence, pace, and difficulty of instructional content before a learning session begins, as well as how to adapt instruction in real-time based on learner model metrics?; (2) What information is required in the learner model to make informed decisions on instructional strategy selection?; (3) How can GIFT best manage guidance and feedback during a learning session based on competency and individual differences?; and (4) What is the optimal approach for delivering GIFT communications to a learner during system interaction?

While GIFT provides the tools necessary to author and deliver adaptive learning applications, an additional function of the framework is to operate as a testbed for the purpose of running empirical evaluations on research questions that will influence future developmental efforts. Empirically evaluating developed models and techniques is essential to ensuring the efficacy of GIFT as a sound instructional tool. To accommodate this requirement, while maintaining domain-independency, GIFT's design is completely modular. This allows for the swapping of specific parts within the framework without affecting other components or models. Modularity enables easy development of comparative systems designed to inform research questions above. The framework is structured to support a variety of experimental design approaches, including ablative tutor studies, tutor vs. traditional classroom training comparisons, intervention vs. non-intervention comparisons, and affect modeling and diagnosis research [6]. The descriptive use case illustrated next is based on an intervention comparison approach.

2 GIFT Experimental Use Case

In this section, we describe in detail the process of using GIFT to design and run a study to evaluate varying methods for communicating information to a learner while they interact with a game-based environment. This experiment was designed to examine varying modality approaches for feedback information delivery during a game-based learning event that is not implicit within the virtual environment (i.e., feedback in the scenario as a result of a player/entity or environmental change). This is influenced by available features present in the GIFT architecture and the benefits associated with research surrounding learning and social cognitive theory [10-11]. The notion is to identify optimal approaches for providing social agent functions to deliver feedback content that is cost effective and not technically intensive to implement. As a result, research questions were generated around the various communication modalities GIFT provides for relaying information back to the learner.

A functional component unique to GIFT is the Tutor-User Interface (TUI). The TUI is a browser-based user-interface designed for collecting inputs (e.g. survey and assessment responses) and for relaying relevant information back to the user (e.g. performance feedback). In terms of providing real-time guided instruction, the TUI can be used as a tool for delivering explicit feedback content (i.e., guidance delivered outside the context of a task that relays information linking scenario performance to training objectives) based on requests generated from the pedagogical model. Because the TUI operates in an internet browser window, it supports multimedia applications and the presence of virtual entities acting as defined tutors. As a potential driver for interfacing with a learner, research is required to evaluate feedback delivery in the TUI and assess its effectiveness in relation to other source modality variations. The overarching purpose of the described research is to determine how Non-Player Characters (NPCs) can be utilized as guidance functions while learning in a virtual world environment and to identify tradeoffs among the varying techniques.

Research questions were generated around the limitation associated with using the TUI during game-based instruction. For a virtual human to be present in the TUI, it requires a windowed display of the interfacing game so the browser can be viewed in addition to the game environment, which may take away from the level of immersion users feel during interaction; thus removing a major benefit with utilizing a game-based approach in education. Specifically, this study will assess whether explicit feedback delivered by NPCs embedded in a scenario environment has a significant effect on identified dependent variables (e.g., knowledge and skill performance, and subjective ratings of flow, workload, and agent perception) when compared to external NPC feedback sources present in the TUI. In terms of serious games, the current research is designed to address how the TUI can be utilized during game-based interactions and determine its effectiveness versus more labor intensive approaches to embedding explicit feedback directly in the game world.

This experiment was the first implemented use of GIFT functioning as a testbed for empirical evaluation. During the process of its development, many components had to be hand authored to accommodate the investigation of the associated research questions. This involved integration with multiple external platforms (e.g., serious game TC3Sim, the Student Information Models for Intelligent Learning Environments (SIMILE) program, and Media Semantics); development of scenarios, training objectives, assessments, and feedback; exploration of available avenues to communicate information; and representing these relationships in the GIFT schema. In the following subsections, we will review the process associated with each phase listed above.

2.1 Testbed Development

GIFT provides the ability to interface with existing learning platforms that don't have intelligent tutoring functions built within. In these games, learners are dependent on implicit information channels to gauge progress towards objectives. Integrating the game with GIFT offers new real-time assessment capabilities that can be used to provide learner guidance based on actions taken within the environment that map to associated performance objectives.

For the instance of this described use case, the serious game TC3Sim was selected as the learning environment to assess the effect of differing feedback modality approaches. TC3Sim is designed to teach and reinforce the tactics, techniques, and procedures required to successfully perform as an Army Combat Medic and Combat Lifesaver [7]. The game incorporates story-driven scenarios designed within a gameengine based simulation and uses short, goal-oriented exercises to provide a means to train a closely grouped set of related tasks as they fit within the context of a mission [8]. Tasks simulated within TC3Sim include assessing casualties, performing triage, providing initial treatments, and preparing a casualty for evacuation under conditions of conflict (ECS, 2012). For the purpose of the experiment, GIFT had to be embedded within TC3Sim for the function of monitoring performance to trigger feedback that would ultimately influence data associated with the dependent variables of interest.

This required pairing of the two systems so that GIFT could consume game state messages from TC3Sim for assessment on defined objectives, and for TC3Sim to consume and act upon pedagogical requests coming out of GIFT. For this to happen, a Gateway Module had to be authored that serves as a translation layer between the two disparate systems. The Gateway Module was also modified to handle feedback requests that were to be delivered by programs external to the game. This included integration with MediaSemantics, desktop and server software that provides character-based applications and facilitated the presence of a virtual human in the TUI that would act as the tutor entity. Following, enhancements to the Communication Module/TUI had to be employed to support the variations in feedback modalities.

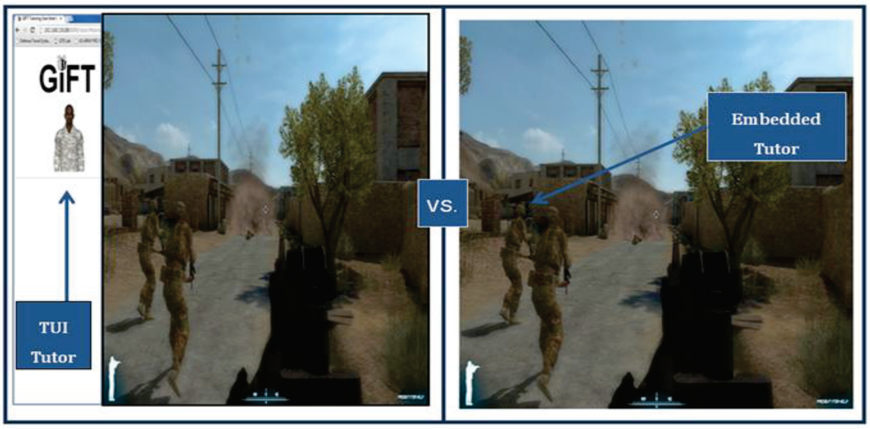

Fig. 4. Feedback Communication Modes

Communication Development. The functional component of GIFT primarily assessed in this research is the Communication Module/TUI, and focused on interfacing approaches for delivering feedback communications during a game-based learning event. For this purpose the major variations associated with the framework took place in GIFT's TUI, as well as identifying approaches for GIFT to manage agent actions within a virtual environment. This required two GIFT/TC3Sim versions with modifications to how the game was visually represented (see Figure 2). With a windowed version of the game, the MediaSemantics character was embedded into the TUI browser and was programmed to respond to feedback requests coming out of the domain module. Furthermore, two additional control conditions were authored to assess whether feedback delivered as audio alone made a difference and a condition with zero feedback to determine whether the guidance had any effect on performance. All participants interacted with the same scenarios, with two conditions including an EPA present in the virtual environment as an NPC. The remaining conditions will receive feedback from external sources to the game. With the functional modifications in place, the next step was designing scenarios, assessments, and feedback scripts.

Scenario Development. With the ability to apply varying techniques of feedback delivery during a game-based learning event, the next step was to design a scenario in TC3Sim to test the effects of all approaches. This requires multiple steps to ensure scenario elements are appropriate so that they lend to accurate inference based on the associated data captured during game interaction. This involved the definition of learning objectives the scenario would entail, associated assessments to gauge performance on objectives, and feedback to apply when performance was deemed poor.

Objectives were informed by competencies identified in ARL-STTC's Medical Training Evaluation Review System (MeTERS) program, which decomposed applied and technical skills for Combat Medics and Combat Lifesavers into their associated tasks, conditions, and standards for assessment purposes (Weible, n.d.). In development of the TC3Sim, the identified competencies were further decomposed into specific learning objectives in terms of enabling learning objectives and terminal learning objectives for each role and task simulated in the game environment. With guiding specifications, a scenario was developed that incorporated decision points for treating a hemorrhage in a combat environment. The scenario was designed to be difficult enough that participants would struggle, resulting in triggered feedback, while not being too difficult that successfully completing the task was impossible.

However, before explicit feedback linked to performance can be delivered in game-based environment, methods for accurately assessing game actions as they relate to objectives is required. The first step to achieve this is properly representing the domain's objectives within GIFT's Domain Knowledge File (DKF) schema by structuring them within the domain and learner model ontology. This creates a domain representation GIFT can make sense of, and results in a hierarchy of concepts that require assessments for determining competency. This association enables the system to track individual enabling objectives based on defined assessments, giving the diagnosis required to provide relevant explicit feedback based on specific actions taken. Following, methods for assessing the defined concepts must be applied that provide information for determining whether an objective has been satisfactorily met. For this purpose, ECS's Student Information Models for Intelligent Learning Environments (SIMILE) was integrated within GIFT.

Student Information Models for Intelligent Learning Environments (SIMILE). An innovative tool used in conjunction with TC3Sim for the purpose standardized assessment is SIMILE (ECS, 2012). In the context of this use case, SIMILE is a rule-engine based application used to monitor participant interaction in game environments and is used to trigger explicit feedback interventions as deemed by GIFT's learner and pedagogical models. In essence, SIMILE established rule-based assessment models built around TC3Sim game-state messages to generate real-time performance metric communication to GIFT. SIMILE monitors game message traffic (i.e., ActiveMQ messaging for this instance) and compares user interaction to pre-established domain expertise defined by procedural rules. As user data from gameplay is collected in SIMILE, specific message types pair with an associated rule authored and look for evidence determining if the rule has been satisfied; that information is then communicated to GIFT, which establishes if there was a transition in performance. Next, that performance state is passed to the learner model. GIFT interprets SIMILE performance metrics for the purpose of tracking progress as it relates to objectives. When errors in performance are detected, causal information is communicated by SIMILE in to GIFT, which then determines the feedback string to deliver.

Feedback Development. Following the completion of linking GIFT's domain representation with SIMILE-based assessments, specific feedback scripts had to be authored that would be presented when the pedagogical model made a 'feedback request'. In the design phase of these prompts, it was recognized that GIFT is dependent on a transition in performance before the pedagogical model can make any decision on what to do next. In the case of the TC3Sim scenario, this requires the player to perform certain actions that denote competency on a concept, but a question is, what information is available to determine they were ignoring steps linked to an objective?

From this perspective, it was recognized that time and entity locations are major performance variables in such dynamic operational environments. Outcomes in hostile environments are context specific, and time to act and location of entities are critical metrics that require monitoring. From there, if a participant had not performed an action in the game or violated a rule that maps to an associated concept, GIFT could provide reflective prompts to assist the individual on what action to perform next. An example applied in the experiment is 'Maintain Cover'. This requires staying out of the streets while walking through a hostile urban environment. For assessment, the player's distance from the street center was monitored, with a defined threshold designating if they maintained appropriate cover. For each concept, rules based on time and locations were identified, and reflective prompts were authored for each concept. Following, audio for each feedback prompt was recorded. This was the final step before the system could be fully developed.

3 Data Collection and Analysis Prep

Data collection was conducted over a five-day period at the United States Military Academy (USMA) at West Point, NY where a total of 131 subjects participated. This resulted in 22 participants for each experimental condition minus the control, which totaled at 21 subjects. The lab space was arranged for running six subjects at a time, with two experimental proctors administering informed consents and handling any technical issues that arose during each session. Once a subject logged in, GIFT managed all experimental procedures and sequencing, allowing the proctors to maintain an experimenter's log for all six machines. This feature shows the true benefit of GIFT in an experimental setting. Once properly configured, GIFT administers all surveys/tests and opens/closes all training applications linked to the procedure, thus reducing the workload on the experimental proctor and enabling multiple data sessions to be administered at a single time. GIFT offers the Course Authoring Tool (CAT) to create the transitions described above. A researcher can author the sequence of materials a participant will interact with, including transition screens presented in the TUI that assist a user in navigating through the materials.

Following the experimental sessions, data must be extracted from associated log files and prepped for analysis. A tool built into GIFT to assist with this process in the Event Reporting Tool (ERT). The ERT enables a researcher to pull out specific pieces of data that are of interest, along with options on how the data is represented (i.e., user can determine if they would like to observe data in relation to time within a learning event or to observe data between users for comparison). The result is a .CSV file containing the selected information, leaving minimal work to prepare for analysis. In this use case, the majority of analysis was conducted in IBM's SPSS statistical software, with the ERT playing a major role in the creation of the master file consumed by the program. This drastically reduced the time required to prep data for analysis, as it removed the need to input instrument responses for all subjects, it structured the data in a format necessary for SPSS consumption (i.e., each row of data represents an individual participant), and produced variable naming conventions listed on the top row.

4 The Way Ahead

GIFT provides a potent testbed in which studies of instructional techniques can be evaluated. Specifically, it allows researchers to investigate how best to implement tenants of intelligent tutoring, including optimal mechanisms for tracking performance, providing feedback and improving outcomes. At the current moment, GIFT provides limited feedback mechanisms that are generally used as formative prompts for correcting errors and reaffirming appropriate actions. New feedback approaches must be explored, such as natural language dialog, to expand the available options for relaying information in game environments. As well, research needs to identify approaches for using environmental cues in the game world to act as feedback functions informed by GIFT. In terms of GIFT as a testbed, advancements need to be applied to the ERT in terms of how data is exported to ease the required post-processing leading to analysis. This includes the ability to segment data in log files based around defined events in the environment that are of interest in analysis. Future research can build on the use case presented and/or conceptualize other investigations that benefit from GIFT.

5 References

1. Heylen, D., Nijholt, A., R., o. d. A., Vissers, M.: Socially Intelligent Tutor Agents. In: T. Rist, R. Aylett, D. Ballin & J. Rickel (Eds.), Proceedings of Intelligent Virtual Agents (IVA 2003), Lecture Notes in Computer Science, 2792: 341-347. Berlin: Springer (2008)
2. VanLehn, K.: The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. Educational Psychologist, 46(4): 197-221 (2011).
3. Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Diagnosing and acting on student affect: the tutor's perspective. User Modeling and User-Adapted Interaction, 18(1): 125-173 (2008).
4. Mayes, T., Freitas, S. d.: Review of e-learning frameworks, models and theories. JISC e-Learning Models Desk Study, from http://www.jisc.ac.uk/epedagogy/ (2004).

5. Beal, C., Lee, H.: Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction. In: Workshop on Motivation and Affect in Educational Software in conjuctions with the 12th International Conference on Artificial Intelligence in Education (2005).
6. Sottilare, R., Goldberg, B., Brawner, K., Holden, H.: Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems. In: Proceedings of the Inteservice/Industry Training, Simulation, and Education Conference, (2012).
7. S: vMedic. Retrieved from http://www.ecsorl.com/products/vmedic (2012).
8. Fowler, S., Smith, B., & Litteral, C.: *A TC3 game-based simulation for combat medic training* (2005).
9. Weible, J.: Preparing soldiers to respond knowledgeably to battlefield injuries. *MSTC U.S. Army*. (n.d.).
10. Bandura, A.: Social Cognitive Theory: An Agentic Perspective. Annual Review of Psychology, 52: 1-26. (2011).
11. Vygotsky, L.: Zone of Proximal Development. Mind in Society: The Development of Higher Psychological Processes: 52-91. (1987).

Authors

Benjamin Goldberg: is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL) Simulation and Training Technology Center (STTC) in Orlando, FL. He has been conducting research in the Modeling and Simulation community for the past five years with a focus on adaptive learning and how to leverage Artificial Intelligence tools and methods for adaptive computer-based instruction. Currently, he is the LITE Lab's lead scientist on instructional strategy research within adaptive training environments. Mr. Goldberg is a Ph.D. Candidate at the University of Central Florida and holds an M.S. in Modeling & Simulation. Prior to employment with ARL, he held a Graduate Research Assistant position for two years in the Applied Cognition and Training in Immersive Virtual Environments (ACTIVE) Lab at the Institute for Simulation and Training. Mr. Goldberg's work has been published across several well-known conferences, with recent contributions to both the Human Factors and Ergonomics Society (HFES) and Intelligent Tutoring Systems (ITS) proceedings.

Jan Cannon-Bowers: is Director of Research at the Center for Advanced Medical Learning and Simulation (CAMLS) at the University of South Florida (USF) in Tampa, FL. She holds MA and Ph.D. degrees in Industrial/ Organizational Psychology from USF. She served as Assistant Director of Simulation-Based Surgical Education for the Department of Education at the American College of Surgeons from 2009 - 2011. Previously, Dr. Cannon-Bowers served as the U.S. Navy's Senior Scientist for Training Systems where she was involved in a number of large-scale R&D projects directed toward improving performance in complex environments. In this capacity she earned a reputation as an international leader in the area of simulation and game-based training, team training, training evaluation, human systems integration and applying the science of learning to real-world problems. Since joining academia, Dr. Cannon-Bowers has continued her work in technology-enabled learning and synthetic learning environments by applying principles from the science of learning to the design of instructional systems in education, healthcare, the military, and other high performance environments. She has been an active researcher, with over 150 publications in scholarly journals, books and technical reports, and numerous professional presentations.