Motivation and research in architectural intelligent tutoring

Keith Brawner*, Anne M. Sinatra and Robert Sottilare

Army Research Laboratory,
Human Research and Engineering Directorate,
12423 Research Parkway, Orlando, FL, 32826-3274, USA
Email: Keith.W.Brawner.Civ@mail.mil
Email: Anne.M.Sinatra.Civ@mail.mil
Email: Robert.A.Sottilare.civ@mail.mil
*Corresponding author

Abstract: It is well-known that personalised and adaptive training, such as from a human tutor, is dramatically more effective than traditional classroom training (Bloom, 1984; VanLehn, 2011). Due to a variety of reasons, however, tutoring systems are not yet ubiquitous within the training market. The US Army Research Laboratory is working to address this problem and has recently published a series of research vector outlines, which guide research in the various areas. The research within the architectural vector naturally exists to support the other vectors and to investigate, standardise, componentise, and commodise the processes and functions of the various tutoring system aspects. This paper serves as an expansion and companion to the similarly named 2015 International Defense and Homeland Security Simulation Workshop paper and yet-to-be-published ARL architectural research plan, expanded in order to discuss the progress made to date, clarify the role of the architecture in the research, and discuss some of the advantages of a unified system as part of measuring training effectiveness and overall system improvement.

Keywords: adaptive and predictive computer-based training; intelligent tutoring systems; ITSs; architectural components; emerging standards.

Reference to this paper should be made as follows: Brawner, K., Sinatra, A.M. and Sottilare, R. (xxxx) ‘Motivation and research in architectural intelligent tutoring’, *Int. J. Simulation and Process Modelling*, Vol. X, No. Y, pp.xxx–xxx.

Biographical notes: Keith Brawner is a researcher within the Learning in Intelligent Tutoring Environments (LITE) Lab within the US Army Research Laboratory’s Human Research and Engineering Directorate (ARL-HRED) in Orlando, Florida. He has 11 years of experience within US Army and Navy acquisition, development, and research agencies. He holds a MS and PhD in Computer Engineering with a focus on Intelligent Systems and Machine Learning (ML) from the University of Central Florida (UCF). His current research is in ML, active and semi-supervised learning, ITS architectures and cognitive architectures. He manages research in adaptive training, semi/fully automated user tools for adaptive training content, and architectural programs towards next-generation training.

Anne Sinatra is a researcher within the LITE Lab at ARL-HRED. The focus of her research is in cognitive psychology, human factors psychology, and adaptive tutoring. She has interest in how information relating to the self and interest personalisation can aid in memory, recall, and tutoring. She is currently a Research Psychologist and Adaptive Tutoring Scientist at ARL after two years as an Oak Ridge Associated Universities/ARL Post Doctoral Fellow.

Robert Sottilare is the Adaptive Training Research Lead at ARL, a co-creator of GIFT, and Lead Editor for the Design Recommendations for Intelligent Tutoring book series. The focus of his research is automated authoring, instructional management, and data analysis ITSs. He is a member of the AI in Education Society and Editor for a special issue of their journal. His work is widely published and includes articles in the *Journal for Defense Modeling and Simulation*, *Cognitive Technology* and the *Educational Technology Journal and Society*. He is a Faculty Scholar and frequent Lecturer at the United States Military Academy and UCF. He is a recipient of two lifetime achievement awards in Modeling and Simulation: US Army Research Development and Engineering Command (2012 – inaugural recipient) and the National Training and Simulation Association (2015 – Governor’s Award).

This paper is a revised and expanded version of a paper entitled ‘Architectural research in intelligent tutoring technologies’ presented at DHSS 2015, Italy, 21 September 2015.

1 Introduction

The United States (US) Army Research Laboratory (ARL) has developed a program of research called adaptive training which includes six interdependent research areas or vectors: "individual learner and unit modelling, instructional management principles, domain modelling, authoring tools and methods, evaluation tools and methods, and architectural and ontological support for adaptive training" (Sottilare, 2013). Each of these research vectors has its own research goals and challenges, and is inseparably linked to desired capabilities within an architecture called the generalised intelligent framework for tutoring (GIFT; Sottilare et al., 2012).

The GIFT system draws its inspiration from a number of successful tutoring systems which address the problems of individual modelling, tools, and evaluation. As an example, the Wayang Outpost system draws similar distinctions among its core modules, including metacognitive instruction and open student modelling over the course of its 10+ reported years of development (Arroyo et al., 2014). Reviews of learning modelling work impress the need for modelling skill, affect, motivation, disengagement, metacognition, and long-term components (Desmarais and Baker, 2012).

Review of the evaluation of the effectiveness of various components of adaptive learning systems (content, instruction, modelling, etc.) additionally stress the need for the ability for separable components and evaluation (Brusilovsky et al., 2004). Each of these areas must be supported by an architecture looking to eventually develop standards, which serves as the motivating need for the GIFT system and the theoretical basis for each of the sections within the current work.

GIFT intelligent tutoring system (ITS) research additionally draws from the need for the system to interface with existing simulations. The difference between a simulator and a training system is the addition of teaching materials during student exercises, real-time feedback, progressively harder content, or other instructional system design. Examples of simulators ripe for the addition of training materials abound in areas such as surgery (Kassab et al., 2012), flight simulators (Richards, 2002), industrial operation (Bruzzone and Longo, 2013), and numerous others. While this area is not the core concept of the paper, it is worth noting that previous work has detailed the relatively large number of simulators in a variety of domains that have been integrated with GIFT for instructional purposes.

In addition to the research vectors mentioned earlier, ARL has been researching and developing GIFT to capture research and best practices for authoring ITSs, delivering instruction via ITSs, and evaluating ITSs. GIFT consists of a series of software modules which are able to interface through a set of standard messages. As is common to most ITSs seen generally (Brusilovsky et al., 2004) or specifically (Arroyo et al., 2014), the core components are: the learner module, pedagogical module, and domain module, along with a tutor-user interface to facilitate interaction with learners. The interactions between these modules form a significant focus for our research vectors.

The first three vectors tie directly to processing objectives in software. Research in individual learner modelling seeks to identify domain-independent learner attributes which will become standard elements of real-time learner models in GIFT-based tutors, such as acquisition of learner data and classification of learner state. Research in this vector also seeks to understand optimal methods for documenting achievements in long-term learner models and developing standards for team models to support training in collaborative activities. Research in the instructional management vector is focused on optimising the selection of domain-independent instructional strategies (plans for action) and domain-dependent instruction tactics (actions) by the ITS. The domain modelling vector is focused on how to represent the complexities of various training domains so the ITS can measure learner progress accurately and interact with the learner in real-time.

The last three research areas are support vectors in that they facilitate processes for authoring ITSs, evaluating ITSs, and enhancing the usability of GIFT as a tool. While ITSs are typically difficult and costly to develop, the authoring vector focuses on simplification through reducing the skill and time required for creation. This is accomplished through simplified interfaces and enhanced artificially-intelligent software agents. GIFT also provides evaluation functions which allow researchers to use GIFT as a testbed through sophisticated data acquisition techniques and data analysis. Finally, the GIFT architecture provides for implementation of adaptive training services, common components and libraries which make GIFT easy to use and easy to integrate with external training systems, such as simulations and serious games, such as the ones found elsewhere in this journal (Choi et al., 2015; Ekyalimpa et al., 2014).

In both the literal and philosophical sense, software architecture has pragmatic purpose and serves a supporting role. As such, the primary function of the 'architecture' component of the adaptive training group is to support and extend the abilities of the other active areas of research. This is performed through the capture of research performed in other vectors, functionality given to specific vectors, and through the practice of standardisation within communication. This paper will neglect the discussion of the origin of the GIFT project, as it is covered in the DHSS 2015 paper (Brawner et al., 2015), but will instead update its current directions, the key implementation challenges, the major architectural research and development challenges, and the opportunity for the international community to contribute.

2 The current gift architecture

At the time of writing, GIFT has nearly 800 users in 51 countries who have registered for accounts on the http://www.gifttutoring.org portal, and has achieved modest technology transition into the field of use with joint projects with both the US Navy and US Army. This adoption rate has been steady, with numbers increasing each month and year. GIFT has served as a basis for much of the US Army's research with adaptive tutoring. The expansion of the program to involve additional personnel, and the expansion of each of the research vectors has resulted in the development of a carefully constructed plan to avoid overlap, continue in a unified direction, and provide the functionally separate components that have been intended and designed towards at the outset of the project.

Significant progress has been made in pursuing the goals identified in ARL's adaptive training research outlines. The following provides a brief summary of progress to date in each vector in terms of its implementation in GIFT. Individual learner models have evolved to include standard (domain-independent) and non-standard (domain-dependent) fields. Several classifiers have been developed to identify cognitive and affective states along with detectors for off-task behaviours. Team models are just beginning to take shape with the completion of large meta-analysis of the team performance and tutoring literature (Sottilare et al., 2012).

Instructional management has taken a leap forward with the development of the engine for managing adaptive pedagogy (EMAP) which examines learner domain competency, motivation, goal-orientation, and grit to aid in recommending courses and course paths for the learner, based upon research evidence (Goldberg et al., 2012). Domain modelling remains a complicated and challenging area for standardisation, but progress is being made in branching tutors from simple desktop tools for cognitive domains to more complex and dynamic tutors for psychomotor tasks.

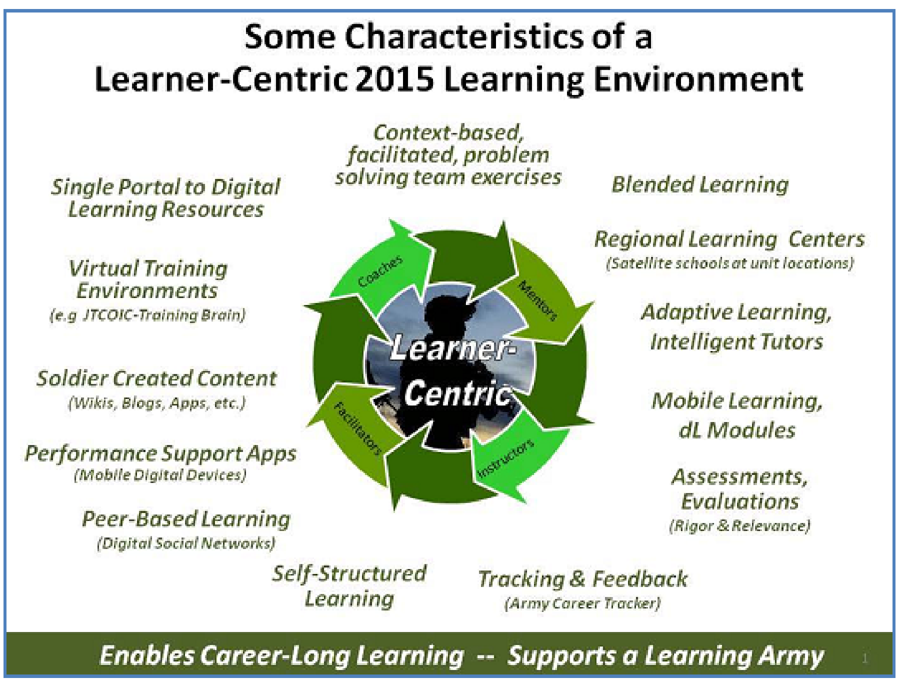
Authoring tools in GIFT have evolved significantly as has the architecture which is now available in a cloud-based version along with a locally-hosted version. The next steps are to provide a self-contained virtual machine version, which will be widely and publicly available by the end of 2016. The authoring tools have much simpler interfaces and the GIFT evaluation tools now include an experimental report tool to support tailored data collection and export, which will be discussed later in this work.

3 Relation to the us army learning model

The Army Learning Model/Concept of 2015, originally published in 2011, has served as a motivation for the development of GIFT in the research community as influences by the research 'pull' of the acquisition community. The authors would like to refresh some of the key concepts in Figure 1, with the knowledge that each of the research vectors is attempting to introduce adaptivity across all objectives:

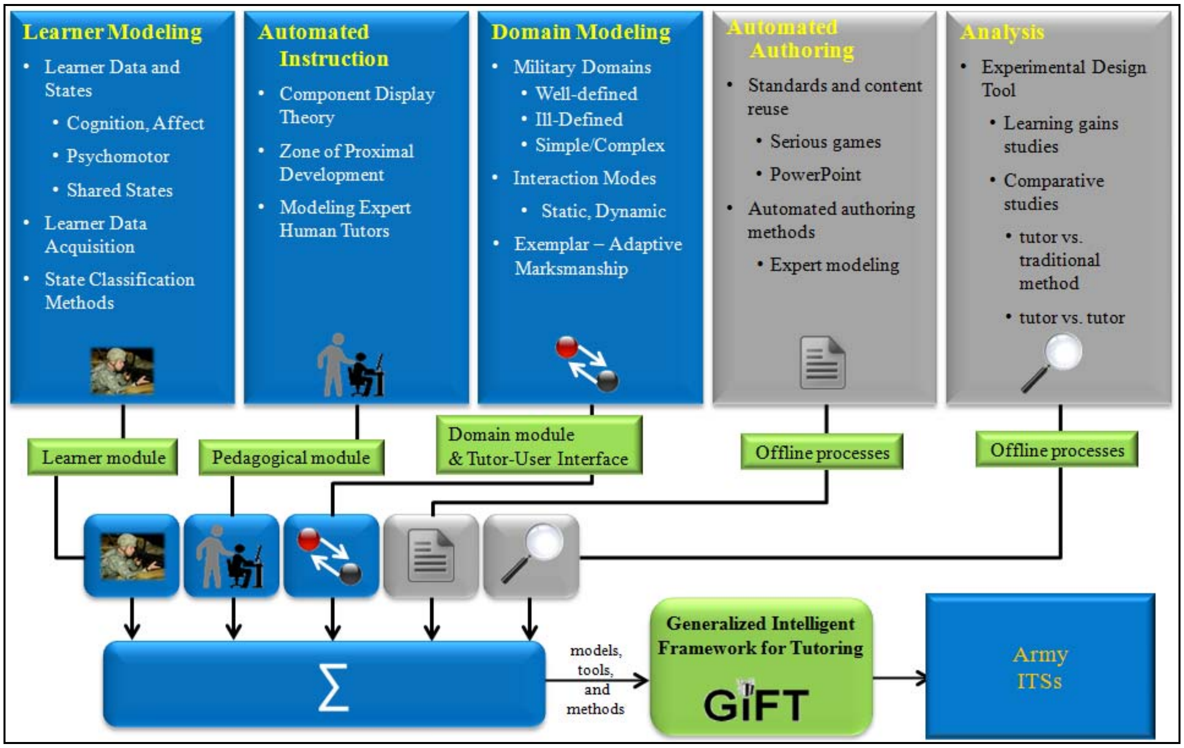
Some relevant portions of this combined learning picture are: tracking of a total career, digitising many learning resources, and the prevalence of 'continuous learning' environments. A continuous learning environment consists of a training environment which is linked to the tactical equipment (embedded training), a virtual environment/campus, and to refresher training on mobile devices, following the general idea that training will be available anywhere at any time.

Figure 1 Army learning model (see online version for colours)



Source: Army (2011)

Figure 2 Adaptive tutoring research vectors (see online version for colours)



Source: Sottilare (2013)

Regardless of the environment and delivery system, each of these training experiences should be adaptive and personalised to the individual in order to promote learning. Adaptive, in this sense, means responsive to the actions of the user: correcting misconceptions for a cognitive task (e.g., troop placement), or correcting performance errors for psychomotor tasks (e.g., marksmanship). Personalised, in this sense, means that the content has been customised for the user who is to receive it. As an example, a user with low motivation may receive material that is highly interactive, as managed by an instructional engine (Goldberg et al., 2012). These decisions are output as data from the modules which make them, and are reliant upon the input data which they receive from other models. The management of this data is shown abstractly as offline and online processes in Figure 2.

4 Guiding principles for a componentised approach

One of the authors, in 2010, worked with a successful military ITS known as the Tactical Action Officer (TAO) ITS which illustrates the state of ITS system design at the time (Stottler and Vinkavich, 2006). The system was designed to have computer 'virtual role players' take the place of live human instruction, such that a 6-man team could train with only one man present. It was designed with a scenario authoring tool to replicate military scenarios that were of interest, in order to stay relevant in modern military environments. Lastly, in unscientific study, it was shown to modestly increase learning, which was not particularly a project goal. The following story of this project provides an illustrative example of the state of the art at the time as well as portions of the guiding design principles behind the GIFT architecture currently.

While this ITS was useful for military training purposes, through elimination and reduction in the number of required instructors, the shortfalls of the field can be seen through the process of its design and support. Firstly, such a system was selected based on the partnership of an ITS company and a defence contractor; the resultant system required the expertise of instructional designers and subject matter experts in addition to the traditional development staff. Such partnerships, although well structured, should not be required to build a training system; *there should be a platform which encapsulates the current state of the science in an existing system for experimentation and use which can be implemented as a traditional engineering 'black box'*.

After system receipt, the schoolhouse desired new rules of operation. Although an authoring tool was developed for the effort, it created new scenarios for the existing assessment rules to be applied: no change could be made to the assessment logic or feedback. Changes in military policy necessitated changes in the system, which then required both instructional and programming knowledge in the type of partnership described earlier. *The system should be able to readjust its assessment logic without reengineering*.

Further, the Navy schoolhouse found the technology useful, as it made the task of instruction easier through partial automation. The training system program was expanded to include instructional content for the ship self-defence system (SSDS). It was found through practice, however, that it was impossible to take the existing instructional models and task assessments from one domain of instruction (e.g., TAO) and apply them to a new one (e.g., SSDS). This re-crafting of the resultant system was nearly as expensive as the creation of the initial system. *A modern ITS should be able to be repurposed for new tasks on an existing simulation without the reinvention of the system itself*.

Finally, the TAO ITS system required updates to some of its core functions, in two general categories: information assurance improvements, and new capability improvements. The information assurance improvements were relatively straightforward, as most modern software systems are designed for ease of maintenance. The modest capability improvements, however, proved difficult due to the closed and tightly coupled nature of the product requiring member of the initial construction team. *Open architectures are needed to facilitate long-term logistics cost of software*.

The lessons here are relatively clear, and have been learned both in other industries (e.g., car manufacturing) and within the computing industry (e.g., operating systems and drivers): common architecture and reusable components reduce time and cost. Specifically, the architecture for a common learning system should be able to encapsulate the knowledge of the supporting roles, such as instructional designers. Components should not be tightly coupled, but loosely integrated, as those that are found elsewhere within this journal (Wang et al., 2013), such that individual portions (i.e., assessment logic), can be changed without programming. The architecture should include a single model of the domain as a component, such that it can be replaced with another for a 'new' tutoring system. Finally, the interfaces and data to such a system should be clearly defined in order to create sustainable systems, or to be easily updated. In response to the needs detailed above, ARL has an ongoing program in adaptive training that is contributing to the state of the art in tailoring training along six research vectors (Figure 2) in support of the US Army Learning Model (Section 3), discussed next.

4.1 Individual and unit modelling component

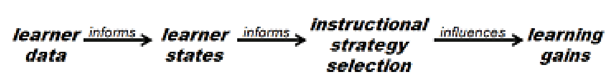
The first vector, individual learner and unit modelling, aligns with and supports both the 'individual learner' and 'social learning' subsections of 'innovation in learning'. In this area, we are researching the effect of transient (e.g., near-term learner states including performance), cumulative (e.g., achievements, competencies), and enduring learner characteristics (e.g., personality, gender) on instructional decisions and outcomes (e.g., learning, performance, retention, and transfer). This includes a recently completed literature review of the team performance and tutoring. We are developing team-level state models for team processes (e.g., coordination, communication, and leadership) and emergent team states (e.g., cohesion and conflict) based on their effect on performance and learning in the literature. These models will be validated in team training environments.

The goal of the learner module is to contain all of the information about the learner that is relevant for the ITS to monitor and adjust to. This can include many different types of information such as previous courses taken in the subject area, competency on a pre-test in the domain, and the learner's current state. Depending on the domain the specifically important learner characteristics may vary. For instance, spatial ability may be relevant for tutoring the subject of land navigation, but not for English literature. Therefore, it is important to design a framework where varying traits can be tracked. Tracking learner information over time (i.e., long-term modelling) is also important, as previous content performance, previous domain experience, or prior domain knowledge may have direct impacts on the types of intended instruction. The learner module also has the responsibility of monitoring the current state of the learner as he or she is engaged in tutoring, and the effect of current instructional actions.

4.2 Instructional management component

The instructional management principles for adaptive training are based on the learning effect model (LEM) and learning theory, shown in Figure 3 (Segedy et al., 2015). The EMAP, the default pedagogical module in GIFT, currently supports an instantiation of Merrill's component display theory (CDT) derived from Gagne's nine instructional events (Goldberg et al., 2012). The basic driver behind this theory is that there is the presentation of rules, example, recall, and practice, where each item builds on the previous items. A summary figure presenting this research is displayed in Figure 3. The work in this area is primarily focused on developing methods for optimal strategy selection based on learner states. The selected strategies drive selection of tactics or actions by the domain module.

Figure 3 Original learning effect chain



Source: Sottilare (2012)

The original real-time model represented in Figure 3 represents learner data (e.g., traits) and states (e.g., emotions), methods to optimally select instructional strategies or plan for action, and their impact on learning gains (e.g., knowledge and skill development, retention, performance, or comprehension). Learner behaviour (e.g., facial expressions) may be an indicator of the learner's emotional state (e.g., frustration) which can negatively affect his ability to learn. This model also includes methods to select optimal instructional strategies. The model now called the LEM has evolved over time to include domain-dependent tactics and non-real-time processes like authoring and long-term learner modelling. While the goal of GIFT was to specify primarily domain-independent components, some domain-dependent elements are needed to select specific tactics from data banks. The LEM now provides the ability to tie required knowledge and skills to course objectives which are in turn tied to course authoring processes to develop tailored learning events with embedded measures of success. The ability to define long-term attributes like domain competency provides an added benefit of allowing GIFT to tailor experiences based on previous training experiences and achievements.

While an instructional path may be relatively straightforward in a specific domain such as algebra, expanding to creating a domain-independent pedagogical module is challenging. As with learner modelling, the most important aspect is providing flexibility so that multiple domains and learner characteristics can be accounted for when strategies are recommended. GIFT's instructional engine, EMAP, is based on a meta-analysis of instructional strategies (Goldberg et al., 2012). Learner characteristics such as motivation or expertise level may result in presentation of different strategies or materials. However, the strategies that are implemented were based on general findings in the literature, and are being applied in a domain independent manner. In computer-based learning experiences, guidance and scaffolding may be critical, as metacognitive strategies and self-regulation are recognised to be key factors for success. Additional challenges to instructional management include the mode of presentation (i.e., does a virtual human provide it? Disembodied voice? Text only?), and context personalisation strategies that present information in ways that are consistent with knowledge the individual already possesses (Goldberg et al., 2015). Addressing these areas is a challenge, and flexibility is key in a domain-independent system.

4.3 Domain modelling component

Domain modelling for adaptive training focuses on the representation of knowledge for a particular task/concept and includes: relationships between goals, learning objectives, concepts, and learner experiences, domain content (a library of scenarios or problem sets), an expert or ideal student model with measures of success, and a library of tactics or actions (e.g., questions, assessments, prompts, and pumps) which can be taken by the tutor to engage or motivate the learner and optimise learning. Some of the prevailing challenges in domain modelling are the ability to represent them in a way that an artificially intelligent tutor can perceive, judge, and act upon. Until recently, ITSs domains were primarily restricted to tasks like problem solving (e.g., solving mathematics and physics problems) and decision making (e.g., exercising situational judgement). Research is needed to identify measures and methods of assessment to allow ITSs to represent more complex domains and then act to change them, such as training for agile thinking (Bruzzone et al., 2014). For example, training in psychomotor domains involves the development of physical and cognitive skills. Difficulty arises when there is significant variance in how each learner executes a physical task and the ITS is unfamiliar with how to measure or assess success.

4.4 Authoring tools component

Authoring tools and methods focuses on research to reduce the time, cost, and skill required to create tutors. This includes the development of standards to support reuse and interoperability among these systems, interface specifications to support easy integration of existing systems, and automation to reduce or eliminate the authoring burden (e.g., expert model development, and scenario evolution based on a single parent scenario).

There are several outstanding challenges in the authoring of tutoring systems. To begin, fundamentally, the creation of an ITS is a new task, somewhat unrelated to other content creation tasks (PowerPoint, PhotoShop, scenarios, etc.), making the inbound transferred knowledge somewhat low. Secondly, automated authoring and reuse processes are yet to be well-defined (e.g., what composes a single element of adaptive training? A single reusable object?). Thirdly, the research must address how each of these new authoring processes be supported inside of a collaborative environment and over the internet.

4.5 Ontological and architectural support component

Lastly, the ontological and architectural support for adaptive training is focused on standardising terms, functions, components and their relationships to support modularity, access at the point of need, and the vectors noted above. These challenges tend to be cross-cutting in the form of content, interoperability, and analytics to support a single point of training, single point for users, and single point of analysis. These challenges and items are discussed next.

5 Content and interoperability

By far, the most difficult design consideration for the GIFT architecture is how to be, and remain to be, domain independent while still contributing something valuable to an individual system. Providing such an architecture requires the removal of much of the context behind performance, and the generalising of instructional strategies. Information such as when and how to provide feedback is domain general, but information which involves specific mistakes or corrections must be handled by an interchangeable module. To support this end, the domain module has a few specific pieces of information made available to it:

- A concept/subconcept hierarchy of the tasks which should be instructed in an individual course.
- A link between each of these concepts/subconcepts and a manner in which to assess them, in the form of tasks, conditions, and standards.
- Tutoring information available for instructional actions, in the form of hints or adaptations.

The classification of information into this schema allows for a single configuration instance (domain knowledge file) to be mostly reused across simulators, for a single simulator to train different tasks according to its tutoring configuration, or to keep all of the other modules of GIFT stable while training a new task in a new domain.

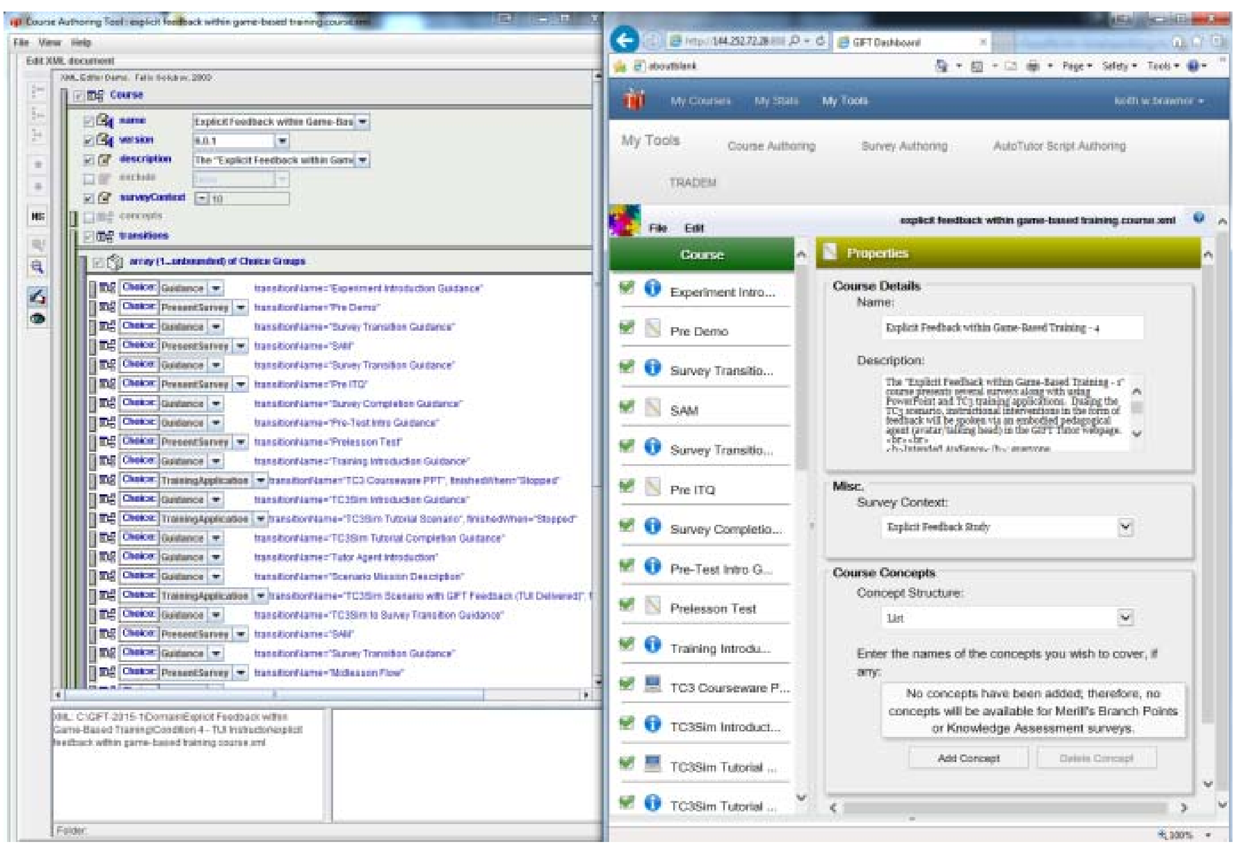
In addition to creating a required method of representing abstract domain structure, domain content is supplemented with information reflecting its content and usage, called metadata and paradata. This information, like the three types of information above, can be abstractly defined for a variety of domains. One of the key features of GIFT is that it allows these features to be built organically; if authored content is available in a compatible manner, it can be seemlessly integrated into the course of instruction, if information (content, assessments, metadata, etc.) is not available, the system defaults to its best guess at appropriate material. The construction of training material in this fashion allows for adaptive capabilities to be built after an initial training system, and to be incrementally constructed.

In reference to earlier content, it is somewhat anticipated that initially authored 'tutoring systems' will be of simplistic nature, perhaps as simple as computer-based training (CBT). The first intention is to allow for the easy creation of CBT-type content. This allows for the second step, of augmenting this existing string of content with manual branching structure, enhancing its general utility. Finally, enhanced through computer automation, recommendation, and/or automated experimental processes, the system can help to determine the best path through content for individual students. This content can be supplemented and delivered with meta- and paradata which can aid in its development and use.

5.1 *Metadata and paradata usage in practice*

EMAP, the default instructional engine behind GIFT, is able to select among the domain-general content to which it has access. It selects this content based upon domain-general content traits and learner-general traits. As an example, a learner who has been identified as having 'low motivation' can be served the content with the highest interactive multimedia instruction (IMI) level available. A 'high motivation' learner in the same situation may be given material where the IMI is lower, but the coverage is greater. The matching of these content traits and learner traits without domain information allows these actions to be performed in disparate instructional contexts. The default instructional engine is based upon a great deal of research, but can be easily reconfigured to support experiments, while tagging individual items with content has additionally been simplified, shown in direct comparison at a glance in Figure 4.

Figure 4 Example of authoring tool simplification (see online version for colours)



If there are two pieces of content, or instructional events, which have the same metadata descriptions, it raises the question of ‘which set should be given?’. GIFT uses paradata, or usage data, to adjudicate the case for the recommendation of matching or identically described content. Currently, this is a placeholder for larger and more appropriate social media-based rating systems to adjudicate individual content selections (Goodwin et al., 2015).

5.2 Interoperability concerns

As part of the creation of an ontological categorisation of domain-specific information, there is difficulty in maintaining the flexibility to the system to adjust to new domains of instruction while supporting existing research projects and transition. The research approach of cobbling together a system for the purpose of testing a theory is helpful in that it can quickly prove novel research ideas. The engineering approach of designing a widely applicable and standardised system allows for the use of proven research outside of its original laboratory.

A typical training model for current military instruction involves training in multiple environments. As a concrete example, a student may be assigned reading on the operation of a vehicle, trained in a simulated environment, trained in a practice environment, operate the vehicle, and receive embedded training during downtime. Sharing data across such disparate systems at sufficient granularity is a difficult problem which calls for interoperable standards. Examples of tasks are predictive modelling (will a student with X knowledge succeed at Y course), transferability (student with knowledge X can skip content Y), or effectiveness (student performing well on X performs well in the field). GIFT has chosen xAPI (Regan, 2013) as an emerging standard which can support the need for this type of actionable data and research question investigation. Other emerging standards such as the human performance markup language (HPML) (Gilbert et al., 2015) are additionally under consideration for representing fine-grained performance.

We seek to more easily identify domain competencies and transferrable skills between training experiences. To accomplish this GIFT will need to leverage the resource description framework (RDF) schema to support descriptions of classes and properties within taxonomies. This will allow GIFT to reduce the uncertainty involved with semantically identifying the effect of achievements documented in experience API (xAPI) statements. For example, having classes of learning experiences sorted by duration, recency, and refresher spacing, it is possible for GIFT to understand the effect of a one hour tutorial on quadratic equations, a four week algebra course, a semester long calculus course, and a degree in mathematics when examining transferrable skills for a new training experience: a three-week course in engineering economics.

6 Analytics and effectiveness

Many of the above concepts hinge on the ability to use materials while enhancing them over time. Instructional models improve selection, domain content can be further generated or supplemented, additional learner data and variables can be tracked, etc. These enhancements may either be machine-originated or human-originated, but should be based upon collected data. The challenge is the ability to analyse the data to allow for smart decisions. Integrated data analytics capability for multiple instructional domains and instructional tasks provides this solution.

The solution for automated analytics is based upon standards in the manner of information communication and storage. There are a few such competing standards available, such as xAPI (Regan, 2013) or DataShop/LearnSphere (Koedinger et al., 2010), which enables a shared format of learner data to be mined by potential algorithms. These algorithms, in turn, make models which can be used across multiple learning domains. The ability to accomplish this, in an automated fashion, and present insights to human users is needed in order for general course improvement. The GIFT project, and ARL’s architectural ITS research by extension, are beginning to integrate with LearnSphere in order to access the most cutting-edge modelling techniques from various available data sources.

In addition to providing tutoring, GIFT is also intended to serve as a means of conducting experiments through the wide-scale collection of data. Such experiments can involve displaying a series of surveys, training applications, and interactions in a linear order, or they can be designed to harness the adaptive features of GIFT and test elements of an ITS for effectiveness. For instance, since GIFT is domain independent, the effectiveness of different instructional strategies can be tested across domains. Additionally, adaptations based on different learner characteristics (varied by changing what is in the learner model) in the same domain could be examined. This flexible design allows for ITS research that would be difficult in other systems.

In order to perform an analysis on the data collected, the researcher can extract learner information and survey input from the desktop version of GIFT’s logs using the event reporting tool (ERT). The ERT allows for different logged data to be selected/merged based on analysis preference. Ultimately a .CSV file, compatible with Excel/SPSS is generated for further organisation and analysis. The ERT also serves as the primary way for instructors to examine student performance on specific concepts and questions that they were provided with during tutoring. The interface of the ERT in GIFT is being updated, with the cloud version having a simplistic interface, and the desktop version having additional power/options.

Experiments can be run with GIFT in both the desktop and cloud versions. Studies that can use an anonymous login and run online using GIFT cloud. A course can be ‘published’ as an experiment, and then the link can be sent to an individual who clicks the ‘start’ button to begin the course. Data can be extracted using the online version of the

ERT, and saved by the researcher. In the future it is intended that participants can login to the system and participate in online studies.

7 Supplementation of content with tutoring information

Initial presentation of content is merely the first part of the tutoring process. A full tutoring process involves content such as hints, prompts, pumps, assessing questions, or topic sequencing. The current manner of generating this type of supplemental content is manual; after the initial training content has been developed, the author is asked to create this type of material. In the creation of an item such as a hint, the domain expert may create an assessing question for each key concept in a supply of training material, a hint for each question, and a series of hints of escalating granularity for concepts which are known to give students issues.

The creation of this supplementary tutoring information generally takes comparable time to the creation of the initial training material. As a by product of the time required to create supplementary tutoring information, its creation by training instructors is performed with some trepidation. GIFT allows the creation of training material in the absence of its tutoring information, but these are the types of information are where learning gains over textbook reading are found.

There are projects involved with automating the tutoring supplemental content. As an example, it is possible, from a variety of texts to establish the order of instruction which is consistent among the domain (Robson et al., 2013). Assessing questions can be automatically generated through question generation techniques which generate multiple choice questions and distracters (Olney et al., 2012). Hints can be generated using a historical series of previous student actions, represented as a Markov chain, to provide a ‘hint factory’ (Stamper et al., 2008). Generally, there is some evidence that the types of supplemental material which authors are reluctant to author can be performed automatically.

8 Architectural research goals

8.1 A single point for training

GIFT does not aim to be a single point for all data to be stored and indexed. However, the goal is to be able to ease integration with a variety of training environments for the purpose of capturing training outcomes and standardising processes. A good architectural structure should allow for the easy import of existing training content, augmentation of its’ resources, sharing of ITS resources, delivery of tutoring instruction, provision of grading information back to instructors, and tracking of long-term learning data. In support of these goals, GIFT has a series of web-based authoring tools, a manner of integration with existing simulators, the ability to share a completed tutoring system. Each of these could have more functionality, but are provided as bare-bones to a diverse set of training systems. The goal is to provide the tools integrate with training systems, and to be able to capture training information where possible. To this end, GIFT may work as an enhanced version of the Gooru Learning platform, which indexes instructional content for use in classroom settings (GooruLearning, 2014).

While functionality is a vital part of any tutoring system, it is also important to consider usability. In this situation, usability refers to simplifying the process and interface that an individual interacts with such that the operation of a system is straightforward and consistent with the way the user expects it to work. Nielsen’s usability heuristics (Nielsen, 2005) are generally used to help guide design, with recommendations such as including iterative design and user expectation discussion (Gould and Lewis, 1985). A goal of many products is to be intuitive, such that a new user could pick it up and quickly figure out how to use it. The initial development of GIFT was focused on increasing functionality, and in its current state a focus has now been placed on improving usability. Additionally, the processes that users engage in for authoring (such as survey authoring) are being updated. Usability is not only important from the author side, but also from the learner side. Future improvements to GIFT are expected to improve the learner experience through providing system roles that only provide the learner with the information that is appropriate to them as opposed to additional authoring capabilities that may distract them from their tasks.

8.2 A single point for users

To the end that GIFT may function as a single point for training content, it is the intention for it to be a single point for users to access other systems, with tutoring optionally applied as an overlay or integrated into the system directly. User needs are simplistic: to access training content, to store a history of their training, and to provide curation and recommendation for future courses. Previous efforts in this area (Mangold et al., 2012) are being folded into the GIFT project in an effort to provide this single sign-on and tracking functionality for taking training, gaining access to new training, lodging social media objections, and other items. Future versions of GIFT will be distributed as virtual machines, for setup at individual schoolhouses, with interoperability with existing or external learner record stores (LRS) (Regan, 2013).

8.3 Single point for analysis

Using a single system to create and take training allows for research on the creation and use of training. This includes many interesting authoring research questions such as ‘which types of instructional domains are most difficult to create training for?’, ‘how can semi-automated tools improve to provide additional levels of automation?’ and additional learner modelling research questions such as ‘which courses are the most critical for future leaders to do well in?’ or ‘how long, on average, does it take before someone forgets critical aspects of their medical training?'. Standardising the data flow across disparate systems allows for the creation of analysis tools which can be applied to these systems. The introduction of powerful analysis tools to answer these research questions for disparate systems, at different types, at different granularity, for different users and groups of users is an architectural research goal. Cooperation with different teams in this area (Burke et al., 2015) will be a key point for reuse and success. GIFT currently allows for designing and conducting experiments See Figure 5 for a screenshot of the experiment tool. After data collection, the streamlined version of the ERT can be used to extract data (see Figure 6).

Figure 5 Screenshot of GIFT's experiments interface (see online version for colours)

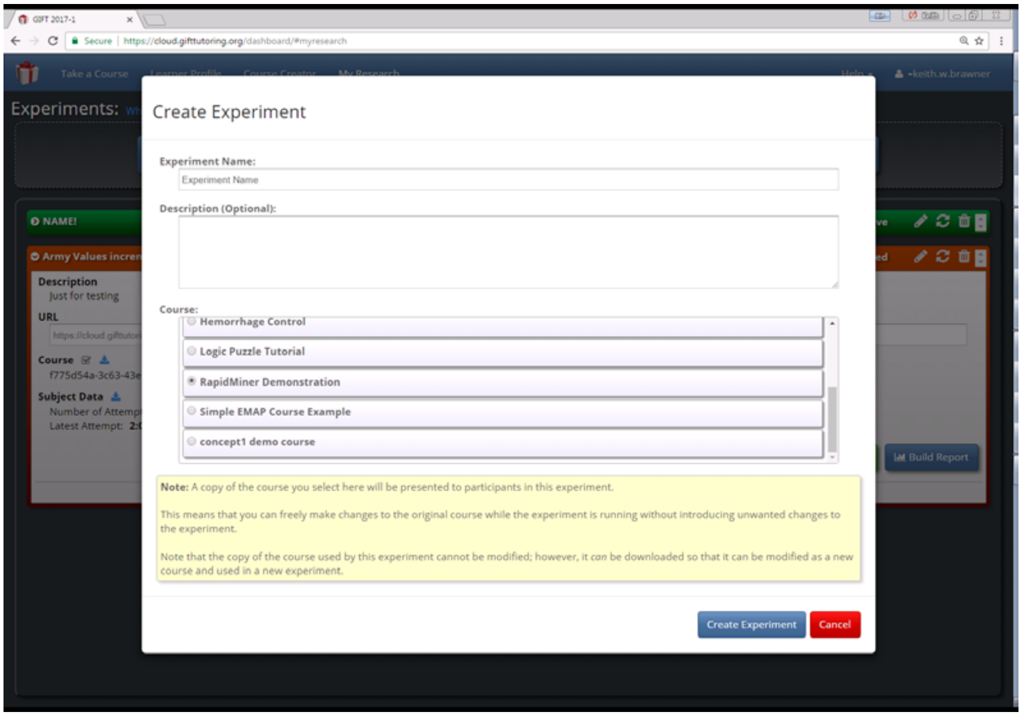
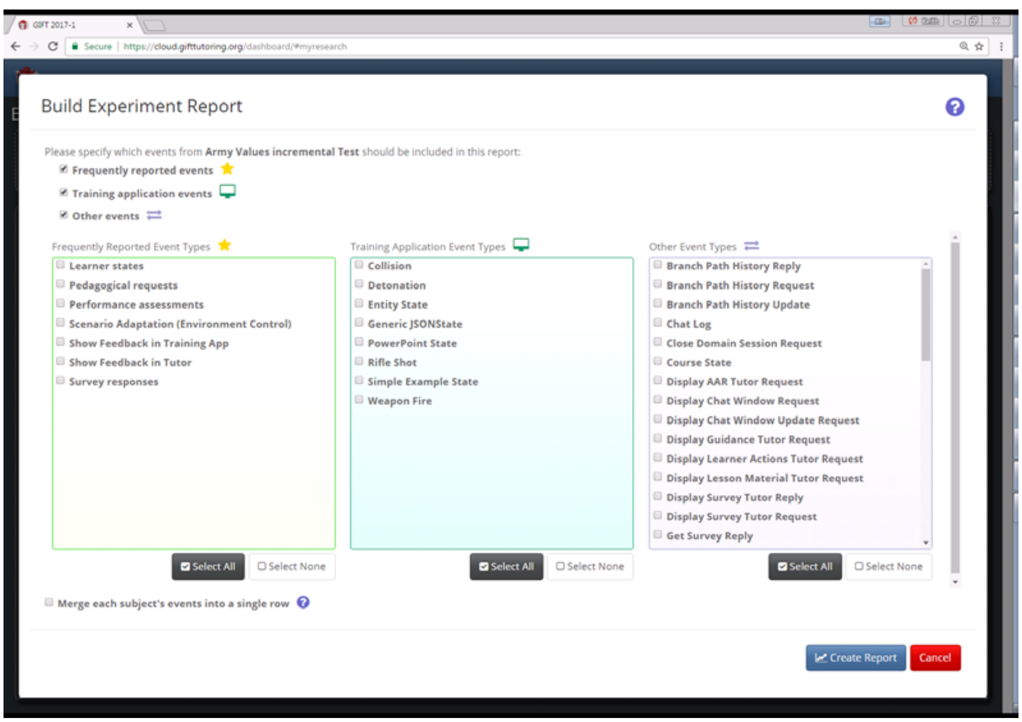


Figure 6 Screenshot of GIFT Cloud's online data extraction interface (see online version for colours)



8.4 Automated background processes

As mentioned in Section 7, automation can magnify individual impact. There are a number of opportunities in automation of learning systems. Some of these involve using AI processes to assist a course creator, such as the creation of course content and supplementary tutoring content. Some of these involve enhanced modelling of users for customised recommendations and assistance. Some of these involve the identification of poorly performing, or highly discussed, course content. Some of these involve items such as customised scenario generation to train automatically identified learner weaknesses. Others can involve running simulated students, or fully automatic simulations (Cuomo et al., 2013). Having data in a single point allows for the reuse of these processes across domains of instruction and gives the benefits to the final users of the software.

8.5 Single point of integration

Lastly, the lessons learned from the earlier TAO ITS system have not been forgotten. GIFT serves as a platform which encapsulates the current state of the science in an existing system for experimentation and use which can be implemented as a traditional engineering 'black box', and provides tools to do so. GIFT is able to readjust its assessment logic without reengineering, through relatively simple changes in configuration files by using existing tools. GIFT is frequently repurposed for new tasks on an existing simulation without the reinvention of the system itself. GIFT has an open architecture to facilitate long-term logistics cost of software, and is released publicly. All of these items allow for the ease of integration with other existing systems. These integration goals are intended to allow for the proliferation of systems, by making their creation easier. They allow for the change of modules, or introduction of new models within modules, without re-creation of the system. They additionally allow for easy data collection and analysis.

9 Future work with data-driven systems

The above text presents a number difficult challenges, which are currently addressed or being addressed by the GIFT project through architectural research. However, there are several major architectural challenges which will likely require a different type of approach towards development, authoring, and use. The first of these is the expansion of one-on-one tutoring to team tutoring, which naturally requires different models of instruction, learner modelling, domain modelling, and other items as appropriate for teams. The second of these is the ability to develop a system which, as a system of agent-driven policies, gradually learns from its mistakes over time to support various training needs (Bruzzone, 2013). The third of these challenges is making sense of a sea of applicable learner data in order to model competencies, skill decay, and other learner aspects which cannot be directly measured.

Creating a domain independent intelligent tutoring architecture is a difficult challenge with many different considerations. However, adding in team tutoring elements to the architecture leads to many more considerations and challenges including synchronisation of multiple learners in a training environment, assessment of both the learner and the team, as well as how feedback should be provided (at the individual or team level). One of GIFT's goals is to allow for the authoring of team tutors. Initial efforts for supporting team instruction and tutoring have begun (Bonner et al., 2015), with an ultimate goal of being able to simultaneously provide team tutoring to squad (nine individuals). Changes to the architecture are required in order to account for the new interactions that arise from team tutoring, as well as being able to author, and monitor both team and individual assessment.

Creating a system which is driven by a series of policies, instead of a series of modules, changes the type of engineering effort required. To use the analogy of a traditional engineering black box system, each component does not require knowledge of the component before or after it. When each component is allowed to change in response to its observed effect, the system becomes a dynamic system, which has different properties. Dynamic systems have seen much success, but the complexities of their creation grow, and traditional standards of operation are frequently discarded during their development.

Lastly, GIFT, as a tutoring system, will produce large swaths of data relating to every learner interaction, learning event, and other items. Fundamentally, this information is currently used by a human to assess basic questions such as 'is person X able to perform task Y?'. This question is complicated, as it is not only related to whether the person has the requisite knowledge and skills, but also how recently the skills have been learned and practiced. The analytics component of the architecture is still very much in the research phase, as these models of predicted performance and ability are currently unknown and untested.

10 Conclusions

Over 50 years of AI research has failed to produce generalised standards for authoring ITSs, automation of their instructional processes, or evaluating their effect. GIFT arose as an open-source, modular architecture to support more standardised processes in ITSs to allow interoperability of components and to reduce the skill/time required to author ITSs. This paper describes the research and development of GIFT capabilities (existing and future needs) and outlines challenge areas in adaptive training research in authoring, automated instruction, domain modelling, and supporting architecture. Many parts of the above currently exist, giving the ability for the user to have a single point for experimental data collection, analysis, authoring, content management, instructional engine, and adaptive training content delivery across affective, cognitive, and psychomotor domains, for individuals and teams, with AI policy-driven components.

GIFT serves as community-based project that needs a large group of practitioners to prosper, grow, and drive official standardisation. It is essential moving forward that GIFT is architected to support a wide-variety of domains (e.g., cognitive, affective, psychomotor, and social/collaborative) to validate its design principles and to demonstrate its authoring and evaluation tools and methods. To this end, we reach out to the global community to apply GIFT freely and provide feedback on its performance. The development of ITS standards will result in lower development time/cost, and higher levels of reuse across all of the participants.

Acknowledgements

The authors would like to acknowledge the significant input that the Army Research Laboratory's (ARL) Human Research and Engineering Directorate (HRED) Adaptive Training group has contributed to this paper, in the form of a multitude of meetings and discussions. In alphabetical order, the people on this team consist of Chuck Amburn, Michael Boyce, Benjamin Goldberg, Gregory Goodwin, Joan Johnston, Rodney Long, Chris Metevier, Jason Moss, and Scott Ososky.

References

Arroyo, I., Woolf, B.P., Burelson, W., Muldner, K., Rai, D. and Tai, M. (2014) 'A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect', *International Journal of Artificial Intelligence in Education*, Vol. 24, No. 4, pp.387–426.

Bloom, B.S. (1984) 'The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring', *Educational Researcher*, Vol. 13, No. 6, pp.4–16.

Bonner, D., Walton, J., Dorneich, M.C., Gilbert, S.B., Winer, E. and Sottilare, R.A. (2015) 'The development of a testbed to assess an intelligent tutoring system for teams', in *AIED Workshops*, Vol. 6, pp.31–37.

Brawner, K., Sinatra, A. and Sottilare, R. (2015) 'Architectural research in intelligent tutoring technologies', Paper presented at the *Defense and Homeland Security Simulation*.

Brusilovsky, P., Karagiannidis, C. and Sampson, D. (2004) 'Layered evaluation of adaptive learning systems', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 14, Nos. 4–5, pp.402–421.

Bruzzone, A.G. (2013) 'Intelligent agent-based simulation for supporting operational planning in country reconstruction', *International Journal of Simulation and Process Modelling*, Vol. 8, Nos. 2–3, pp.145–159.

Bruzzone, A.G. and Longo, F. (2013) '3D simulation as training tool in container terminals: the TRAINPORTS simulator', *Journal of Manufacturing Systems*, Vol. 32, No. 1, pp.85–98.

Bruzzone, A.G., Massei, M., Tremori, A., Poggi, S., Nicoletti, L. and Baisini, C. (2014) 'Simulation as enabling technologies for agile thinking: training and education aids for decision makers', *International Journal of Simulation and Process Modelling*, Vol. 9, Nos. 1–2, pp.113–127.

Burke, C.S., Feitosa, J. and Salas, E. (2015) 'The unpacking of team models in GIFT', Paper presented at the *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, Orlando, FL.

Choi, C., Seok, M-G., Choi, S.H., Kim, T.G. and Kim, S. (2015) 'Military serious game federation development and execution process based on interoperation between game application and constructive simulators', *International Journal of Simulation and Process Modelling*, Vol. 10, No. 2, pp.103–116.

Cuomo, A., Rak, M. and Villano, U. (2013) 'Concurrent simulation in the cloud with the mJADES framework', *International Journal of Simulation and Process Modelling*, Vol. 8, No. 4, pp.212–226.

Desmarais, M.C. and Baker, R.S. (2012) 'A review of recent advances in learner and skill modeling in intelligent learning environments', *User Modeling and User-Adapted Interaction*, Vol. 22, Nos. 1–2, pp.9–38.

Ekyalimpa, R., AbouRizk, S.M., Mohamed, Y. and Saba, F. (2014) 'A prototype for project management game development using high level architecture', *International Journal of Simulation and Process Modelling*, Vol. 9, No. 3, pp.131–145.

Gilbert, S., Winer, E., Holub, J., Richardson, T., Dorneich, M. and Hoffman, M. (2015) 'Characteristics of a multi-user tutoring architecture', Paper presented at the *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, Orlando, FL.

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., Billings, D.R. and Malone, N. (2012) 'Use of evidence-based strategies to enhance the extensibility of adaptive tutoring technologies', Paper presented at the *Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*.

Goodwin, G.A., Murphy, J.S. and Hruska, M. (2015) 'Developing persistent, interoperable learner models in GIFT', Paper presented at the *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, Orlando, FL.

GooruLearning (2014) [online] http://www.goorulearning.org (accessed 10 June 2014).

Gould, J.D. and Lewis, C. (1985) 'Designing for usability: key principles and what designers think', *Communications of the ACM*, Vol. 28, No. 3, pp.300–311.

Kassab, E., Tun, J.K. and Kneebone, R.L. (2012) 'A novel approach to contextualized surgical simulation training', *Simulation in Healthcare*, Vol. 7, No. 3, pp.155–161.

Koedinger, K.R., Baker, R., Cunningham, K., Skogsholm, A., Leber, B. and Stamper, J. (2010) 'A data repository for the EDM community: the PSLC DataShop', *Handbook of Educational Data Mining*, pp.43–55.

Mangold, L.V., Beauchat, T., Long, R. and Amburn, C. (2012) 'An architecture for a soldier-centered learning environment', Paper presented at the *Simulation Interoperability Workshop*.

Nielsen, J. (1994) 'Enhancing the explanatory power of usability heuristics', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, April, pp.152–158.

Olney, A.M., Graesser, A.C. and Person, N.K. (2012) 'Question generation from concept maps', *Dialogue & Discourse*, Vol. 3, No. 2, pp.75–99.

Regan, D.A. (2013) 'The training and learning architecture: infrastructure for the future of learning', Paper presented at the *Invited Keynote International Symposium on Information Technology and Communication in Education (SINTICE)*, Madrid, Spain.

Richards, R.A. (2002) 'Common cockpit helicopter training simulator', Paper presented at the *AVSIM 2002 Conference & Exhibition*.

Robson, R., Ray, F. and Cai, Z. (2013) 'Transforming content into dialogue-based intelligent tutors', Paper presented at *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*, Orlando, FL.

Segedy, J.R., Kinnebrew, J.S., Goldberg, B.S., Sottilare, R.A. and Biswas, G. (2015) 'Using GIFT to model and support students' metacognition in the UrbanSim open-ended learning environment', Paper presented at the *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, Orlando, FL.

Sottilare, R.A. (2012) 'Considerations in the development of an ontology for a generalized intelligent framework for tutoring', Paper presented at the *International Defense & Homeland Security Simulation Workshop*, Vienna, Austria, September.

Sottilare, R.A. (2013) *Special Report: Adaptive Intelligent Tutoring System (ITS) Research in Support of the Army Learning Model – Research Outline*, Army Research Laboratory, Orlando, FL.

Sottilare, R., Burke, S., Johnston, J., Sinatra, A., Salas, E. and Holden, H. (2015) 'Antecedents of adaptive collaborative learning environments', in *Proceedings of the Interservice/Industry Training Simulation & Education Conference*, Orlando, FL, December.

Stamper, J., Barnes, T., Lehmann, L. and Croy, M. (2008) 'The hint factory: automatic generation of contextualized help for existing computer aided instruction', Paper presented at the *Proceedings of the 9th International Conference on Intelligent Tutoring Systems Young Researchers Track*.

Stottler, R.H. and Vinkavich, M. (2006) *Tactical Action Officer Intelligent Tutoring System (TAO ITS)*, DTIC Document.

United States Army Training and Doctrine Command (2011) *The United States Army Learning Concept for 2015*, Fort Monroe, VA.

VanLehn, K. (2011) 'The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems', *Educational Psychologist*, Vol. 46, No. 4, pp.197–221.

Wang, H., Hu, B. and Li, H. (2013) 'An agent-based simulation study for exploring organisation design under environmental uncertainty', *International Journal of Simulation and Process Modelling*, Vol. 8, Nos. 2–3, pp.118–132.