Learner Models in the Generalized Intelligent Framework for Tutoring: Current Work and Future Directions

Gregory A. Goodwin
U.S. Army Research Laboratory – Human Research and Engineering Directorate

INTRODUCTION

The function of an intelligent tutoring system (ITS) is to adapt or tailor training to an individual learner. As with a human tutor, this requires the ITS to have some "knowledge" of the learner (i.e., a learner model). The ITS uses and updates the learner model as the learner progresses through the material. For example, if the learner masters some concept, the learner model must be updated to reflect this. On the other hand if the learner has difficulty with a concept, the ITS needs to be able to understand where deficiencies lie in order to prescribe the appropriate remediation.

Understanding why the learner might have had difficulty with a particular concept is no simple task as the list of reasons could be quite extensive. Perhaps the learner lost focus during the presentation of a key piece of information, lacks some key prerequisite knowledge, or has a low aptitude for the domain. The list could go on and on.

All of these possible explanations require assessment of the learner. As can be seen from the above example, assessments can include information about the learner's background, experiences, traits, and aptitudes, as well as measures of the learner's affect, behavior, and performance during the training session. The more completely the learner model represents the learner, the better the ITS will be able to effectively adapt training.

Dimensions of Learner Modeling

In September of 2015, we published a report outlining research challenges in the area of individual learner modeling (Goodwin, Johnston, Sottilare, Brawner, Sinatra, Graesser, 2015). This report described a framework for assessment of the learner to support learner modeling. This framework provides a way of classifying different types of measures and relates those measures to adaptive methods.

The framework categorizes measures into four groups in a 2 x 2 matrix. One axis in the matrix divides measures into state-like or trait-like categories. Trait like measures are what the learner brings to the training event. Examples would include physical strength and aptitude. State-like measures on the other hand are things resulting from the training. Examples include fatigue or confusion. State-like measures are fairly stable and either don't change, or change very slowly. Trait-like measures change fairly quickly and and are often transient.

The other axis in the matrix divides measures into content-dependent or content-independent categories. Content dependent categories are learner measures that are directly relevant to the content being trained. Examples include prior knowledge or comprehension. Content independent measures are traits and states that are relevant to training generally rather than to specific content. Examples include aptitude and personality traits. Each of these four cells apply to three domains of learning (cognitive, affective, and psychomotor, *vis*. Bloom, 1956).

State-like and trait-like measures have some interdependencies (Goodwin, Murphy, & Hruska, 2015). For example, a student with high aptitude or prior expeince would be expected to perform better in training (Schafer & Dyer, 2013). Additionally, some state-like measures could update trait-like measures. For example, as the learner completes a block of training, his or her performance (state-like measures) would then update the trait-like measures, (e.g., indicating the learner had mastered a particular skill or completed a certification course).

ITSs need both state like and trait like measures to adapt training effectively (VanLehn, 2006). For example, before an ITS can initiate training, it needs to know something about the learner. What does the learner already know? What is the learner's aptitude? How motivated is the learner to complete the training? The ITS might use this information to determine the difficulty level of the training or what topics to skip. These are often described as outer-loop adaptation. As the ITS delivers training, it will measure student comprehension, attention, as well as the types of errors made, and level of frustration and/or boredom. The ITS can use these measures to choose remedial content or to change the pace or difficulty of the training – so called inner loop adaptation (VanLehn, 2006). Table 1 summarizes the kinds of measures that can be used for adaptation of training in GIFT.

Table 1. Components of the Learner Model.

	Learner Measure Category	Trait-Like (Outer Loop Adaptation)	State-Like (Inner Loop Adaptation)
Content Dependent	**Cognitive**	Relevant prior cognitive experience/knowledge/training	Comprehension of concepts presented in the training
	Psychomotor	Relevant prior psyhomotor experience or training,	Measures of Skill improvement
	Affective	Fears, likes, goals, attitudes relevant to the training.	Arousal and emotions in response to the training
Content Independent	**Cognitive**	Intellect/Aptitude, Memory, Meta-cognitive skills	Attention, Cognitive Workload
	Psychomotor	Physical strength, stamina, sensory acuity	Endurance and fatigue
	Affective	Personality Traits, general test anxiety	Arousal, emotions resulting from factors independent of training

Using this assessment framework for developing learner models has a couple of benefits. First of all, by understanding that there are different uses for each type of assessment, it is possible to think about ways that those uses might be standardized in GIFT modules. This might be especially true for content-independent measures. Second, it is useful in identifying research and technical challenges that affect certain types of assessments.

For example, in-training assessments of learner state are challenging because they must be frequently and rapidly assessed in a nonobtrusive way by the training system. Such assessments rely on measurement technologies like eye-trackers and physiological measures that can be expensive and may only be available in certain training facilities. This highlghts the need for research and development to bring the cost of these capabilities down and to increase their validity.

Assessment of trait like factors is time consuming and so we want to avoid doing this every time a learner starts a training session. Ideally GIFT would access pre-existing databases containing that information (e.g., personnel records, learner records). Research is needed to develop ways to access that information in a secure way using open standards. Services also need to be developed to facilitate interoperability among databases. The next section outlines ongoing research in the area of learner modeling.

AREAS OF RESEARCH ON INDIVIDUAL LEARNER MODELS FOR GIFT

The following are areas of research on individual learner models for GIFT that are currently being investigated:

Modeling Learner Competencies

We know that ITSs can be expected to operate within a larger ecosystem of training events and systems. For example, for a given skill or course, a learner may receive training in a live or distributed classroom led by a live instructor, participate in hands-on training, virtual simulation training, multimedia training, and/or game-based training. Often these separate events are developed and sequenced so that the learner's skill or expertise progresses throughout the course. The ITS may only deliver a single block of instruction within the larger course or may be used to provide remedial training. Both of these circumstances indicate that there is a need for a learner model that tracks learner competencies as they develop across multiple training venues and that can be shared among multiple training systems.

Competencies are domain specific knowledge and skills possessed by the learner. Competencies can encompass a large set of skills acquired over a long time (e.g., being a researcher or a physician) or they can be very specific (e.g., launching a Raven unmanned aerial vehicle). The challenge is that there are no standard, broadly accepted, validated ways to assess most competencies. Competencies are reflected not only in the training the learner has received, but also by their experience and performance of that competency in battlefield conditions. Competencies change over time though gradually. They may increase if the learner practices the competency regularly but they can decline in the absence of practice.

Because there is no standard set of assessments for most competencies, and because competencies are not static, there is a need to be able to determine competencies at the time of training. An effort (Engine for Quantifying User Intelligence and Performance – EQUIP, see Goodwin, Murphy, & Hruska, 2015) is investigating an approach to provide this capability to GIFT. The components of the EQUIP architecture include a Learner Record Store (LRS) that contains performance data relevant to the competency being assessed in an experience application programming interface (xAPI) format; an interoperable learner competency model (ICM), and of course the GIFT application.

Figures 1 and 2 illustrate this system. Let's suppose that a course delived in GIFT were to tailor training to a learner based on the learner's current competency in some domain. As shown in figure 1 below, the course concepts are read from the domain knowledge file (DKF) and are then passed through the gateway module to a web service that hosts a set of ICMs. The web service queries all ICMs to determine which ones may be relevant to the concepts of the course. Each ICM contains an index of performance measures and methods for interpreting those measures which are returned to GIFT.

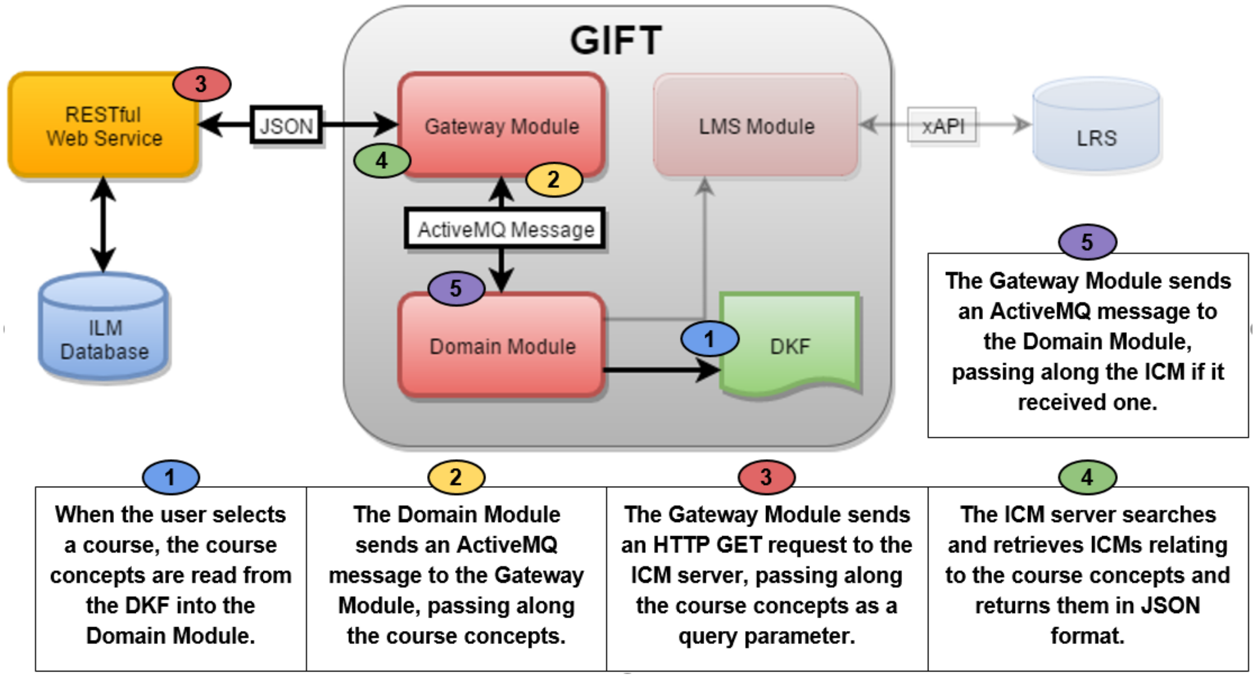

Figure 1. GIFT Integrated Architecture Flow: Steps 1 - 5

Once GIFT has the ICM data, it can then query the LRS for the appropriate performance data for that learner and subsequently interpret that data to estimate the learner's current competency level as illustrated in figure 2. GIFT can also add assessments to the LRS. To estimate the competency level, it is necessary to have validated models to predict them.

For example, suppose we were to develop an ICM for marksmanship. the Army presently scores marksmanship competency/proficiency in four categories based on the number of hits in a standard course of fire:

1. Expert (38-40 hits; max = 40)

2. Sharpshooter (33-37 hits)

3. Marksman (26-32 hits)

4. Unqualified (25 or fewer hits)

In an initial entry training environment, students complete the marksmanship qualification test at the end of training. In that training environment, the ICM could use learner measures to make predictions about competency using the Army standard. However, it would probably also be useful to be able to use learner measures to make predictions about performance in intermediate training events. For instance, an ICM might map performance in the simulator to predictions about performance during the subsequent period of live instruction.

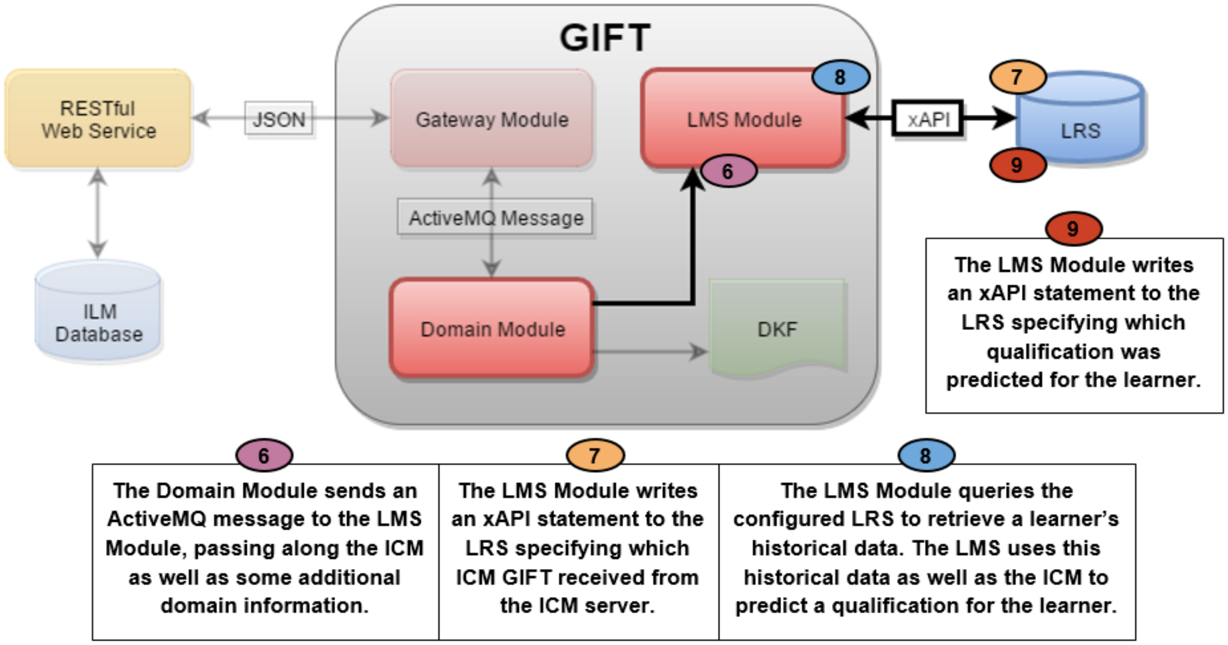

Figure 2. GIFT Integrated Architecture Flow: Steps 6 – 9

As the learner completes training in GIFT, those learner behaviors and assessments would be fed back into the LRS. That new performance data would then impact subsequent assessments of learner competencies.

The primary advantage of ICMs is that they allow for standardization of competency modeling across different training systems. Furthermore, they allow for GIFT to know more about its learners than what they have done in GIFT applications. By opening up this window to GIFT, it can much more efficiently target training to learners. In this way, GIFT can act much more like a human tutor would, as an adjunct to a course for example. Clearly, this allows GIFT to operate within a larger ecosystem of training systems including live, virtual, constructive, and gaming in a seamless way.

Increasingly, these assessments are being written using an industry standard known as the xAPI specification. This standard was developed by the Advanced Distributed (ADL) Co-Lab as a means of logging learner activities across a wide variety of platforms, systems, and media. Each xAPI statement includes a subject, verb, and object and contextual information (ADL, 2013). The specification also includes data transfer methods for the storage and retrieval of these statements from a learner record store (LRS) and security methods for the exchange of these statements between trusted sources.

Currently, data pertaining to learner actions, states, and accomplishments stored using the xAPI specification provide the best means of creating and updating a persistent interoperable learner model. In order to do this, GIFT and other adaptive training systems will need to both consume and generate xAPI statements of learner assessments that can be used to update competencies in a learner model.

Assessing Differences in Motivation: Long Term Learner Modeling

Another effort underway in the learner modeling domain involves an examination of the ways in which motivation affects the rate of learning and forgetting of a given learning task. The approach taken is to develop and validate a motivator taxonomy that matches motivators to personality traits of learners. For example, it may be that individuals who score high on measures of extraversion are most strongly motivated by acknowledgement from peers or higher ups. On the other hand, an introvert may be more motivated by free time and relaxation.

This is a three-phased project and work is currently in the first phase. The first phase focuses on the development of a Motivator Assessment based on individual differences. The motivator assessment identifies motivation in the learner. It builds upon efforts to incorporate additional classification variables that include student personality, learning performance history, and motivational responses. Motivational responses refer to a measured increase of sustained effort, because of the end goal resulting in a reward based on personality. Sustained effort would be indicated by physiological measures, such as a higher amount of oxygen produced for a longer sustained time or an increase in heart rate due to stress/arousal to meet the goal.

The second phase of this project will involve an experimental verification of the Motivator Taxonomy and/or the Motivator Assessment. Specifically, this will test how personality and the Motivator Taxonomy/Assessment affects the learning rate and retention of training. The learning objective could be presented in a simulation-enabled mission command, intelligence, surveillance, and reconnaissance mission, UMedic, or some other application to be determined. The goal for this phase, is to identify the relation between the classifications of motivational tools and individual factors with the learning rate and retention, specifically the Long Term Learner Model

In the final phase of this project, data collected from the previous scenario will be tested across a different domain, population, and/or scenario. All results will then be used as the basis for a framework that will provide pedagogical recommendations based on the evaluation of the Student's real-time data on motivation and personality factors into a specific learning intervention for GIFT training.

Modeling the Determinants of Training Time in GIFT

Adaptive training promises more effective training by tailoring content to each individual insuring that it is neither too difficult nor too easy. Another, less discussed benefit of adaptive training, is improved training efficiency. This efficiency comes from minimizing the presentation of unnecessary material to learners. Typically, non-adaptive training is developed for the lowest tier of learners. While this insures that no learner will be unable to complete the training, it also means that many students are given material that is not well suited to their current level of understanding.

The focus of this effort (Goodwin, Kim, Niehause, 2017) is to determine how the fit between learner characteristics (e.g., aptitude, reading ability, prior knowledge), learning methods employed by the adaptive training system, course content (e.g., difficulty and length, adaptability), and test characteristics (e.g., difficulty, number of items) all determine the time to train for a population of learners.

We use a probabilistic model to represent the different factors and instructional strategies that impact the completion time of a MAST module, as well as probabilistic inference techniques to determine a distribution of a course completion time.

For example, if a trainee normally reads at 100 words per minute, there are 100 words in the text, and the trainee is tired, the reading time of the trainee could be distribution uniformly from 1 to 2 minutes. The reading speed of the trainee is also a non-deterministic variable that depends on how much prior knowledge the trainee possesses about statistics about how fast the general population of trainees read.

One of the benefits of building a probabilistic model to represent the completion time is that not all of the information in the model is needed to estimate the completion time. For example, if we know how much prior knowledge the user has about the subject (for example, from a pre-instruction questionnaire), we can post that knowledge as *evidence* to the model that would be taken into account when estimating the completion time. If we do not possess that information, we can treat the variable as *latent* and use a prior distribution to represent the state of the variable. For example, we can estimate that only 20% of trainees taking the course have prior knowledge of the subject. These prior distributions can be estimated from the literature review or expert knowledge, and then *learned* over time based on the outcomes of actual testing.

RESEARCH CHALLENGES

To date, the research into how best to adapt training content based on student performance in intelligent tutoring systems is inconclusive (Durlach & Ray, 2011). As can be seen, GIFT baesd research on learner modeling is still relatively nascent. Some key areas of research that need to be investigated are described below.

Cross platform training. The major benefit of interoperable student models is the ability to adapt training across technology platforms. Using the xAPI specification, performance data can be recorded and interpreted from a wide variety of platforms, including desktop and mobile devices. While some Army-sponsored efforts have focused on assessing student performance across a range of training platforms (e.g., Spain, et al., 2013), maintaining a complex student model across these platforms – and adapting training accordingly – has yet to be successfully accomplished in a military context. Integrating GIFT with xAPI data would enable investigations into the best practices for adapting training across platforms.

Macro- versus micro-adaptive interventions. Multi-faceted student models based on cognitive, psycho-motor, and affective components are inherently complex, and may be representative of both "state," or situationally dependent components such as level of workload and "trait," or more persistent student characteristics such as personality traits. Whether to adapt training on a macro level (e.g. course selection) or a micro level (e.g. real time adaptation of content) based on these complex models has yet to be fully investigated. While some research suggests macro-adaptative strategies are more appropriate for more persistent characteristics (Park & Lee, 2004), this question has not been addressed across domains.

Adaptation based on a combination of learner states. Assessing a learner's affective state during the course of training has been a focus of ITS research over the past decade (e.g., D'Mello & Graesser, 2007). However, research into how to adapt training based on this state is in its infancy (e.g., Strain & D'Mello, 2015). Arguably the state of the art in intelligent tutors, Affective AutoTutor (D'Mello & Graesser, 2007), senses student cognitive and emotional states such as boredom and frustration and acts to alleviate states. If a negative emotion is detected, the avatar within the tutor responds with an encouraging phrase and facial expression. In Affective AutoTutor, student affect and learning are managed through separate models; that is, interventions that are geared toward managing frustration are distinct from interventions aimed at manipulating content difficulty. The extent to which different interventions could be used to address combinations of these states has yet to be determined, but is a research question GIFT could support.

Scenario-based training. GIFT is unique in that it supports intelligent tutoring in scenario-based platforms such as the Army's *Virtual Battlespace 3* (VBS3). How to assess competencies across complex student models using key events within one of these scenarios has yet to be investigated. If scenario data were recorded in xAPI specification scenario events could be diagnostic of both performance and affect. Key to this development is the careful mapping of competencies to decision events in a scenario. Best practices for accomplishing this have yet to be established.

Predictive analysis of performance. Persistent learner models provide the opportunity to prescribe interventions based not only on performance during training but also prior to training on both the macro- and micro-adaptive level. Based on performance in one training setting, a student model could reflect a number of cognitive, psychomotor, and affective attributes which could then predict performance in another setting, given the domains were sufficiently interrelated. These data could be used to prescribe courses of instruction, training platforms, and even micro-adaptive strategies. To date, this potential has not been investigated.

Return on investment of different types of interventions. To date, research into addressing interventions based on complex student models is feasible. However, whether or not a learning intervention is effective is not that same issue as whether or not it is effective *enough.* With defense budgets becoming increasingly limited, the question is whether adapting training based on complex representations of student competency is worth the investment. Implementing intelligent tutoring systems to date has been limited due to their domain specificity and cost to develop. While the GIFT initiative aims to address these issues specifically, the relative cost of some interventions has yet to be determined. For example, emerging physiological technology enables the unobtrusive measurement of student cognitive and affective state (Murphy et al, 2014), but does adapting training based on these types of measures produce sufficient learning gains to warrant their cost? These questions have yet to be fully investigated.

CONCLUSIONS

This discussion highlights a number of research questions that can be addressed as the result of integration of complex, interoperable learner models into the GIFT architecture. Through the use of xAPI data, representations of student performance can incorporate data from a multitude of sources. The GIFT team envisions a multi-faceted learned model consisting of psychomotor, cognitive and affective aspects of competencies. This model can be used to drive training adaptations across technological platforms, across domains, and across the course of a learner’s career. While the potential to fully model the lifelong learning of a student is promising, research is needed to fully evaluate the utility of these learner models. Some of this work is currently underway at the Advanced Distributed Laboratory under a program known as the Total Learning Architecture (TLA, Johnson, 2013).

As an initial attempt at addressing these issues, several projects are using a marksmanship use case for an initial investigations of this capability. Marksmanship is an ideal domain for implementing multi-faceted learner models. While marksmanship skills may appear to be straightforward, effective performance is much more than simply hitting a target with a bullet. The marksman must master a range of psychomotor, cognitive, and affective skills in order to be successful, and must have an understanding of how myriad environmental factors play into his or her accuracy. Furthermore, marksmanship is a skill that every Soldier must master, so it has a broad applicability to the Army and its sister services.

It is important to note research in learner modeling is still in its infancy. Consequently, our efforts are a first step toward developing definitive guidelines and best practices for how to best leverage interoperable performance data. Further research will be needed to expand an understanding of how these learner models play into the development and use of intelligent tutors across domains, training audiences, and platforms.

REFERENCES

Advanced Distributed Co-Laboratories (2013). xAPI-1.0.2. Retrieved from https://github.com/adlnet/xAPI-Spec/releases/tag/xAPI-1.0.2.

Bloom, B.S. (Ed.). Engelhart, M.D., Furst, E.J., Hill, W.H., Krathwohl, D.R. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay Co Inc.

D'Mello, S. & Graesser, A. (2007). Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments. In R. Luckin, K. Koedinger & J. Greer (Eds.), Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007) (pp 161-168). Amsterdam, The Netherlands: IOS Press.

Durlach, P. J. & Ray, J. M. (2011). Designing adaptive instructional environments: Insights from empirical evidence (ARI Technical Report 1297). Arlington, VA: U.S. Army Research Institute

Goodwin, G.A., Kim, J.W., Niehaus, J. (2017). Modeling Training Efficiency and Return on Investment for Adaptive Training, Presented at the 5th *Annual GIFT Users Symposium* 10-11 May, 2017, Orlando, FL

Goodwin, G.A., Murphy, J.S., & Hruska, M. (2015). Developing Persistent Interoperable Learner Models in GIFT. Presented at the *3rd Annual GIFT Users Symposium* 17-18 June, 2015, Orlando, FL

Goodwin, G., Johnston, J., Sottilare, R, Brawner, K, Sinatra, A., & Graesser, A. (2015). Individual Learner and Team Modeling for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline. ARL Special Report 0336. U.S. Army Research Laboratory.

Johnson, A. (2013). The Training and Learning Architecture: Meeting the Needs of the Next Generation of SCORM. Slides from a Webinar. Downloaded from: http://adlnet.gov/wp-content/uploads/2013/02/TLAWebinarFeb2013-1.pdf on 4 Jan, 2017.

Murphy, J.S., Carroll, M.B., Champney, R.K., & Padron, C.K. (June, 2014). Investigating the Role of Physiological Measurement in Intelligent Tutoring. Paper presented at the GIFT Users's Symposium, Pittsburgh, PA

Park, O., & Lee, J. (2004). Adaptive instructional systems. In D.H. Jonassen (ed.), *Handbook of Research on Educational Communications and Technology 2nd edition* (pp. 651-684). Mahwah, NJ: Lawrence Erlbaum.

Schafer, P., & Dyer, J (2013). Defining Tailored Training Approaches for Army Institutional Training. (ARI Research Report 1965). U.S. Army Research Institute for the Social and Behavioral Sciences. Fort Belvoir, VA.

Strain, A., & D'Mello, S. K. (2015). Affect regulation during learning: The enhancing effect of cognitive reappraisal, *Applied Cognitive Psychology, 29*: 1–19.

Spain, R., Mulvaney, R., Cummings, P., Barnieu J., Hyland, J., Lodato, M., & Zoileck, C. (2013). Enhancing Soldier-centered learning with emerging training technologies and integrated assesments. *Interservice and Industry Training and Simulation and Education Conference*, Orlando., FL.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, *46* (4), 197–221.

ABOUT THE AUTHOR

***Gregory Goodwin** is a senior research scientist and acting Branch Chief at the Army Research Laboratory-Human Research and Engineering Directorate, Simulation and Training Technology Center (STTC) in Orlando, Florida. His research focuses on methods and tools to maximize the effectiveness of training technologies. After completing his Ph.D. at the State University of New York at Binghamton in 1994, Dr. Goodwin spent three years in a post-doctoral fellowship at the Columbia University College of Physicians and Surgeons followed by a year as a research associate at Duke University Medical Center before joining the faculty at Skidmore College. In 2005, Dr. Goodwin left academia and began working at the Army Research Institute (ARI) field unit at Fort Benning Georgia and six years later, he came to the ARI field unit in Orlando, FL where he has been examining ways to leverage technologies to reduce the cost and improve the effectiveness of training.*