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:

Personality: A key to motivating our learners

This report by Biddle, Lameier, Reinerman-Jones, Matthews, and Boyce (2018) describes an effort to utilize the personality of the learner (a trait-like factor) to identify key motivators that will improve learning performance. This association between personality and motivators can then be used by GIFT to use those motivators to tailor training to each individual. For example if people who are outgoing find social affirmation to be a powerful motivator, GIFT might utilize something like leaderboards or feedback from other learners to incentivize those learners.

In fact, prior research has shown that personality and motivational factors are related. For example, learners with intrinsic motivation, which refers to an internal desire to succeed, are more likely to have a high level of the personality trait Conscientiousness (Duckworth et al, 2007).

Last year, the authors presented work which identified items for a Motivator Assessment Tool (MAT). This tool identified individual motivational traits and specific associated reinforcers. This year, the authors have added items to these scales and check the reliability and factor structure and provide final refinement to the MAT items and then examine the relationship between the MAT items and the Big Five personality traits finding some interesting associations between personality and motivators. For example, they report that individuals who are open, conscientious, and/or agreeable tend to associate with self-directed learning. On the other hand, individuals who score high on neuroticism tend to find the learning environment threatening and would be difficult to motiviate.

Currently GIFT only tailors training based on a classification of learners as novice, journeyman, or expert. The next phase of this work will focus on integrating this survey into GIFT to provide a classification by personality. Using the associations that were discovered between personality traits and motivators, the course could then be tailored by the pedgagocial module accordingly.

This work also highlights the need for GIFT to implement a long-term learner model to avoid having to re-assess learners each time they take a GIFT course. As noted, traits tend not to change over time and so there is little need to readminister a survey that should essentially yield the same score each time. In fact, as noted by the authors, subjecting learners to the same survey over and over would probably be a demotivator.

Perceptual-cognitive Training Improves Cross-cultural Communication in a Cadet Population

In this paper by Folsom-Kovarik, Boyce, and Thomson (2018), the authors explore ways to more efficiently develop remediaton for training in GIFT using a cross-cultural communications lesson plan. More specifically, the investigators explored ways to adapt training using patterns of learner behaviors, common misconceptions, and a specific type of adaptation known as mid-lesson reports.

The concept patterns refers to the ways in which learners tend to progress through the lesson. Some learners may persist until they achieve success. These learners are willing to try different strategies to solve the problems until they get it right. Other learners may not shift a response strategy, trying the same strategy over again, possibly several times, before quitting. Still others fall somewhere in between these two extremes.

The concept of common misconceptions is fairly self-explanatory. For any given question, incorrect responses are often associated with a specific misconception. In the case of this project, the learning objective had to do with cross-cultural communication. The questions required the learner to balance different values or outcomes and then choose the best, though imperfect, course of action. Misconceptions identified by the authors included an authoritarian response in which the learner was mostly focused on being respected or obeyed, or a rules focus in which the learner inflexibly adheres to rules. These, and other, misconceptions could be applied across a wide range of question responses.

In this experiment, the identification of the misconception, allowed the appropriate remediation to be selected by the pedagogical module in GIFT. The remediation was provided in the form of mid-lesson feedback pointing out the error by challenging the misconception and encouraging further reflection before responding. The interventions worked on most of the responses, improving learning outcomes.

One of the outcomes of this report is a recommendation to enhance the learner model to understand the patterns of responding by particular learners. Does a learner easily adapt his or her response strategy or doe the learner seem to persist in using an unsuccessful strategy? By understanding the learner's response pattern, GIFT may be able to tailor prompts to these different types of learners.

Another suggestion made by the authors of this report was to identify common or general misconceptions that learners make when responding to topical questions. The reusability of those misconceptions could make it easier to author remediation. If the content author simply identifies the misconception associated with a response, the pedagogical module can apply the appropriate remediation (e.g., encouraging the learner to apply a different response strategy) avoiding the need to author a unique remedation for each response of each question.

Predicting Students' Unproductive Failure on Intelligent Tutors

In this report Park and Matsuda (2018) examine a method for detecting a type of unproductive failure known as wheel-spinning. Wheel spinning occurs when a student seems to be unable to figure out how to solve a particular problem or problem type. The result is that students spend an extended period of time on a problem without making progress. Students can become frustrated and will eventually give up. Needless to say, this is not effective or efficient learning and being able to detect students that are heading into this hole before they get too discouraged is critical to improving learning outcomes.

The investigators in this report used archival data in DataShop to explore modeling methods for predicting this pattern of learner behavior. They used four student factors: performance, hint usage, sum of response time, and difficulty of problem type. Employing a data mining method using a gradient boosted decision tree model yielded a model that could predict wheel spinning patterns of behavior about 62% of the time after the third opportunity to solve a problem and 83% of the time by their sixth opportunity. Future work will need to focus on how to adaptively and constructively respond to this pattern of learning so that students do not get frustrated and improve their learning outcomes.

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, Niehause, 2018) 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.

In this second year of this effort, the focus has been on further elaboration of the MAST model, identification of GIFT training content for use in the validation experiment for the final year of this effort and developing interoperability between the predictive model and GIFT.

RESEARCH CHALLENGES

As can be seen, GIFT baesd research on learner modeling is still relatively nascent. However, the projects described above are pursuing a number of interesting approaches to both developing learner models and using them to adapt training to improve both training effectiveness and training efficiency. All of the key research challenges identified last year continue to need more work. These 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, psychomotor, 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.

Biddle, E., Lameier, E., Reinerman-Jones, L., Matthews, G., & Boyce, M. (2018). Personality: A key to Motivating our Learners. Presented at the 6th *Annual GIFT Users Symposium* 9-11 May, 2018, Orlando, FL

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.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly , D. R. (2007). Grit: perseverance and passion for long-term goals. Journal of personality and social psychology, 92(6), 1087.

Folsom-Kovarik, J.T., Boyce, M. W., & Thomson, R.H. (2018). Perceptual-cognitive training improves cross-cultural communication in a cadet population. Presented at the 6th *Annual GIFT Users Symposium* 9-11 May, 2018, Orlando, FL

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.

Goodwin, G., & Niehaus, J.(2018). Modeling training efficiency and return on investment for adaptive training: GIFT integration. Presented at the 6th *Annual GIFT Users Symposium* 9-11 May, 2018, Orlando, FL

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 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.*