Proceedings of the Sixth Annual GIFT Users Symposium

May 2018
Orlando, Florida

Edited by:
Robert Sottilare

Part of the Adaptive Tutoring Series

Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6)

Edited by:

Robert Sottilare

The views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Army Research Laboratory.

Use of trade names or names of commercial sources is for information only and does not imply endorsement by the U.S. Army Research Laboratory.

This publication is intended to provide accurate information regarding the subject matter addressed herein. The information in this publication is subject to change at any time without notice. The U.S. Army Research Laboratory, nor the authors of the publication, makes any guarantees or warranties concerning the information contained herein.

Printed in the United States of America
First Printing, May 2018

*U.S. Army Research Laboratory
Human Research & Engineering Directorate
Orlando, Florida*

International Standard Book Number: 978-0-9977257-5-9

Dedicated to current and future scientists and developers of adaptive learning technologies

CONTENTS

FROM THE EDITOR

GIFT is a free, modular, open-source tutoring architecture that is being developed to capture best tutoring practices and support rapid authoring, reuse and interoperability of Intelligent Tutoring Systems (ITSs). The authoring tools have been designed to lower costs and entry skills needed to author ITSs and our research continues to seek and discover ways to enhance the adaptiveness of ITSs to support self-regulated learning (SRL).

This year marks the sixth year of GIFT Symposia and we accepted 30 papers for publication in this year's proceedings. None of this could happen without the efforts of a fantastic team. Our program committee this year did an outstanding job organizing and reviewing, and we want to recognize them for their efforts.

- Kat Amaya
- Michael Boyce
- Keith Brawner
- Elyse Burmester
- Noel Cal
- Paula Durlach
- Benjamin Goldberg
- Gregory Goodwin
- Jong Kim
- Rodney Long
- Paul Shorter
- Anne Sinatra
- Bob Sottilare

We are proud of what we have been able to accomplish with the help of our user community. This is the sixth year we have been able to capture related research and development efforts for the Generalized Intelligent Framework for Tutoring (GIFT) community which at the writing of these proceedings has well over 1500 users in over 76 countries.

In addition to providing a record of the symposium content, these proceedings are also intended to document the evolutions of GIFT as a tool for the authoring of intelligent tutoring systems (ITSs) and the evaluation of adaptive instructional tools and methods. Papers in this volume were selected with the following goals in mind:

- The candidate papers describe tools and methods that raise the level of knowledge and/or capability in the ITS research and development community

- The candidate papers describe research, features, or practical applications of GIFT

- The candidate papers expand ITSs into previously untapped domains with recommendations for future GIFT capabilities

- The candidate papers build/expand models of automated instruction for individuals and/or teams

The editor wishes to thank each of the authors for their efforts in the development of the ideas detailed in their papers and their thoughtful presentations. As a community we continue to move forward in solving some significant challenges in the ITS world and in this light GIFTSym has evolved. GIFT and the GIFT Symposium continue to take on a broader perspective as the new Center for Adaptive Instructional Sciences (CAIS) was formed this year begins formal operations under ARL's Open Campus Initiative. The purpose of CAIS is to encourage the community development of adaptive instructional systems (AISs), capabilities & standards. Since our last GIFTSym, the IEEE Learning Technologies Standards Committee (LTSC) has approved the development of an AIS standards study group and GIFTSym has allocated a full-half day session to inform AIS stakeholders. You can learn more about CAIS and our IEEE AIS standards activities by signing up at https://www.arl.army.mil/opencampus/centers/cais.

We would also like to encourage readers to follow GIFT news and publications at www.GIFTtutoring.org. In addition to our annual GIFTSym proceedings, GIFTtutoring.org also offers volumes of the Design Recommendations of Intelligent Tutoring Systems, technical reports, journal articles, and conference papers. GIFTtutoring.org also includes a users' forum to allow our community to provide feedback on GIFT and influence its future development. We encourage you to subscribe and to send us your stories and experiences using GIFT as part of our GIFT_around_the_Globe series.

Finally, ARL has developed about GIFT instructional videos which are now available on YouTube at: https://www.youtube.com/results?search_query=generalized+intelligent+framework+for+tutoring. We encourage you to subscribe.

Many thanks to all of our GIFT users…

Bob

Robert A. Sottilare, Ph.D.

GIFTSym6 Chair and Proceedings Editor

Theme I: GIFT Architecture

Architecture and Ontology in the Generalized Intelligent Framework for Tutoring: 2018 Update

Keith Brawner, U.S. Army Research Laboratory
Mike Hoffman, Dignitas Technologies

INTRODUCTION

The first version of the Generalized Intelligent Framework for Tutoring (GIFT) was released to the public in May of 2012. One year later, the first symposium of the GIFT user community was held at the Artificial Intelligence and Education conference in Memphis, Tennessee. Since then, the GIFT development team has continued to gather feedback from the community regarding recommendations on how the GIFT project can continue to meet the needs of the user community and beyond. This current paper continues the conversation with the GIFT user community in regards to the architectural "behind the scenes" work and how the GIFT project is addressing the user requirements suggested in the previous GIFTSym5 proceedings. The development team takes comments within the symposium seriously, and this paper serves to address requirements from prior years.

As a follow up to the "GIFT 2015 Report Card and State of the Project" (Brawner & Ososky, 2015), the GIFT 2016 Community Report (Ososky & Brawner, 2016), and the GIFT 2017 Architecture Report (Brawner, Heylmun, & Hoffman, 2017), the feature requests and responses have been broken out among a number of papers, and into logical sections of this work. This paper discusses the ongoing architectural workings and changes in support of the various sets of projects. The number of projects which the GIFT overall projects is now well over 50, which represents a) the inability for significant direct support of any individual project and b) the relatively little support that individual projects need to be successful. GIFT generally works well enough to support research studies without direct developer guidance or specifically developed features.

The research and technology innovation efforts presented in the current document include those that are informed by the GIFT user community, and only represent a fraction of the overall research, development, and implementation work associated with GIFT. We invite the reader to review the other chapters in this volume, publications on GIFTTutoring.org, and other references described below, to get a sense of the total body of work on the GIFT project. Major themes in this current, 2018 GIFT report include tighter integration with wide-scale systems such as EdX and LearnSphere, further work in enhancing authoring, significant load tests for supporting many simultaneous users, the first and second GIFT Summer Camps, an upcoming shift to better conversational agents, and the move to individualized training for teams and during psychomotor tasks.

WELCOME

First, to the new members of the GIFT community and new GIFT users – Welcome! There are a number of recommended resources that will help to orient you to this project and ecosystem. GIFT has come a long way since its original goals were defined in its description paper (Sottilare, Brawner, Goldberg, & Holden, 2012). First, we would encourage you to simply get started, as the tools and example courses have been designed to try to be as easy as possible for the creation of intelligent tutoring systems.

If you struggle with any individual aspect of the system, however, the team has produced short "how to" videos to try to help around the sticking points. There are now around 20 such videos, available at the following youtube channel URL: https://www.youtube.com/channel/UCWtI_V8f2mN5XD6h2lCjsAA/videos, which is the first result if you search "Generalized Intelligent Framework for Tutoring Youtube" on Google. If you would like additional help getting started, please consider the GIFT Quick Start Guide (Ososky, 2016) as another place to start.

In addition to a Quick Start Guide, usable tools, and videos, there is support for developers in the help forums and documentation. The GIFT user community is also invited to ask questions and share your experiences and feedback on our forums (https://gifttutoring.org/projects/gift/boards). The forums are actively monitored by a small team of developers, in addition to a series of Government project managers. The forums are a reliable way to interact with the development team and other members of the GIFT community. The forums, at the time of this writing, have over 1200 postings and responses. Documentation has been made freely available online at https://gifttutoring.org/projects/gift/wiki/Documentation, with interface control documentation https://gifttutoring.org/projects/gift/wiki/Interface_Control_Document_2018-1, and a developer guide https://gifttutoring.org/projects/gift/wiki/Developer_Guide_2018-1. These documents are updated each software release. The concept document was also updated in 2017 (Sottilare, Brawner, Sinatra and Johnston, 2017).

CLOUD GIFT GENERAL REPORTING

Cloud GIFT has now been up and running for the last two years. Increasingly, users start on the Cloud GIFT instance to make and take their first courses. With minimal outages, the system has now been up for a number of years. While initially envisioned as a "try before you buy" program (Brawner & Ososky, 2015), user expectations and general usability have demanded more mature software functionality from this research project. We have responded to the community demand for reliability in the Cloud GIFT instance by increasing its accessibility significantly. We, the development team, did not anticipate that users would author surveys with multiple hundreds of questions, open the system up to 100+ users on Amazon Mechanical Turk simultaneously, or other relatively high-demand tasks. This is a good problem to have, and we have taken several actions to harden the system to the level of robustness demanded from the community.

First, updates to GIFT Cloud now significantly precede the updates to the downloadable GIFT. Downloadable GIFT still operates on the 12-month developmental cycle, while Cloud GIFT is now operating on a 7-day release cycle. This effort has required significant re-tooling to move to a dev-desk, dev-cloud, and production model. As a byproduct, the team responds much quicker to bug requests. These changes are transparent to the end users but involve significant effort from the team. The latest stable regression-tested GIFT release is still available for download at gifttutoring.org, but a clone of what is available at cloud.gifttutoring.org is always available upon request.

Second, as part of the move to Cloud GIFT developmental cycles, we have been coordinating stress tests in order to identify system weaknesses and harden against them. Early weaknesses were identified in survey editing, survey requests, course validation, content upload, and other database-intensive requests. One initial stress tests of the system showed as few as 6 simultaneous users could successfully perform database-intensive operations. Modern tests after performance improvements have been made, at the time of writing, are reporting on the order of magnitude of 100 simultaneous users. These changes are transparent to the end users but involves significant effort from the team.

Third, as a part of hardening the system for research, the end use capability has been the ability to run educational experiments with cloud-deployed software instantaneously across the country. This capability is relatively mature, and the author is aware of several such experiments which have been run with 100+ users, by the teams for Adaptive MOOCs (Aleven et al., 2017), Long Term Learner Modeling (Biddle, Lameier, Reinerman, Matthews, & Boyce, 2018), Structural Equation Modeling (Robson, Ray, & Sinatra, 2017), the After-Action Review (Brawner, Carlin, Oster, Nucci, & Kramer, 2018; Carlin, Nucci, Kramer, Oster, & Brawner, 2018), and Tutorial Planning (Rowe, Pokorny, Goldberg, Mott, & Lester, 2017). This capability is available for use by the general public.

Virtual Machines Available Upon Request

As part of the move to Cloud GIFT, we have a number of specialized processes which run in the background. Figure 1 shows the current structure of the Virtual Machine (VM) instances which operate Cloud GIFT. At its basic level, GIFT runs on two VMs; a Windows VM for all of the core GIFT features, and a Linux VM hooked up to an Amazon Relational Database Service (RDS) for the content. These items are what are contained in the downloadable GIFT instance. In addition to the basic instances, however, are services for monitoring GIFT; PiWik monitors user behaviors within the system, while the GIFT monitoring service monitors usage for future performance improvements. GIFT now includes an instance to a Social Media Framework (SMF) and Learner Record Store (LRS), which are based around Elgg and Learning Locker, respectively. GIFT's copies of these configurable items are available upon request, and posted to github, but the authors would urge users to select their own instances of commercial sharing and data warehousing items dependent upon their own individual needs; there is nothing tying GIFT to a specific SMF, LRS, PiWik, or monitoring framework. We do not think of these items as core to GIFT, only that they are reported outwards.

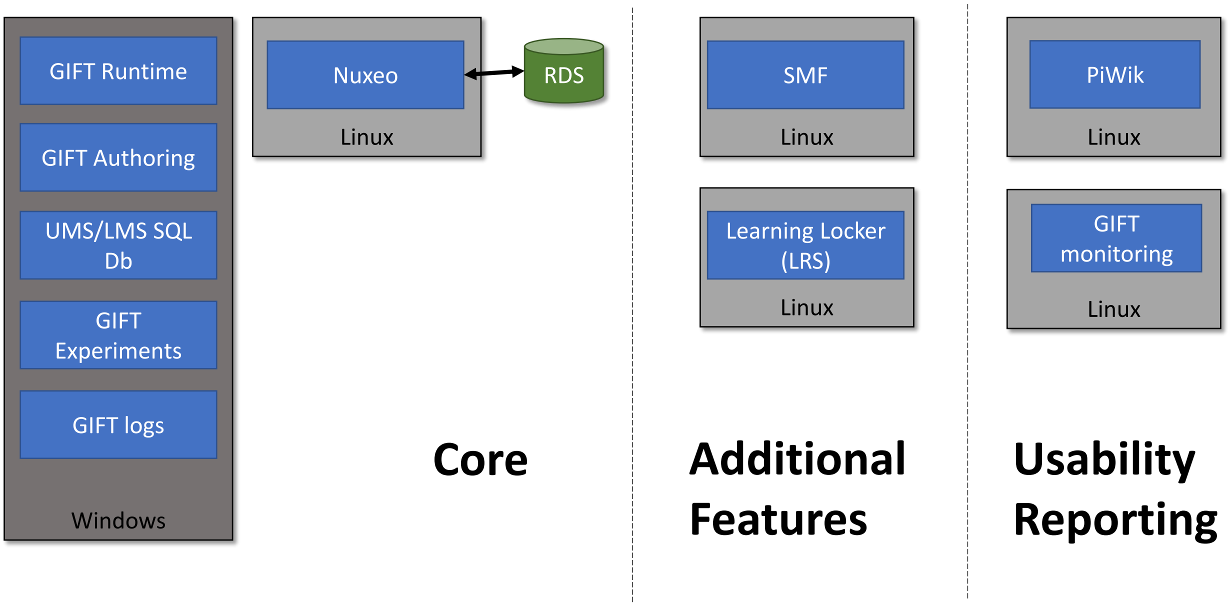


Figure 1: Simplistic Diagram of Cloud Gift Items

NEW INSTRUCTIONAL MODELS

GIFT has historically been based on the Engine for Management of Adaptive Pedagogy (EMAP) processes. These processes were based upon an extensive literature review which diagnosed the best types of content to give learners, based on the traits of the content and learner. This framework has expanded to accommodate Chi's interactive, Constructive, Active, and Passive (iCAP) framework (Chi, 2009; Rowe et al., 2017). Within the authoring tools, this expansion involves the addition of the "remediation" area to the existing the Rules, Example, Recall, and Practice areas of the Adaptive Courseflow object. Content from this remediation area, if require, is then given preferentially to content in the other areas. If no remediation content is available, or all of the remediation content has already been given, the system will then give a single piece of content from the content within the other bins. This is a change to the behavior of the GIFT adaptive Courseflow object in two manners:

1. Remediation content will be considered before other content when presenting remedial content.

2. Regardless of student performance, only one piece of content will be given prior to retesting

Existing courses are being automatically upgraded in order to use this instructional model, and two instructional events have been added with the ability to be authored within the remediation content block as active/constructive activities – highlight passage and summarize passage. The seamless migration from one instructional model to another instructional model is one of the features of the GIFT system, was designed from the beginning, and is now put to the test.

The move to this instructional model is based upon evidence of effectiveness and is being done in order to support machine learning processes for the optimal selection of remedial content based upon the evidence of effectiveness within an individual course (observed effectiveness) as opposed to effectiveness based upon research projects (theorized effectiveness). More about this project, its results, and the machine learning processes which are being used can be found in (B. S. Goldberg, 2018).

VIRTUAL HUMAN TOOLKIT (VHTK)

There are many problems with the "talking head" process which GIFT has used since the beginning of the project. Firstly, we used this talking head for relatively simplistic reasons – it was already being used as part of the AutoTutor integration and using it by default limited integration cost. Secondly, the character usage was not an open source item, as opposed to the rest of GIFT. Changing the avatar, voice, or character responses usually involved paying Media Semantics a fee. Thirdly, the Media Semantics Character Server and Builder were required to run on Windows, which is also not open source or free. Fourth, Media Semantics has discontinued support of the avatar for modern browser compatibility standards. In summary, it costs money, costs OS maintenance, limits new user adoption, and isn't supported by the company which created it.

For the reasons above, we have wished to switch to another virtual human technology. Previous efforts in allowing GIFT to be more ontology-driven (Nye, Auerbach, Mehta, Hartholt, & Fast, 2017) have allowed for us to use interchangeable agents, and were demonstrated at the last GIFT Symposium. The lack of developmental support for the MSC forced our hand to switch to the new ontology-driven agent processes.

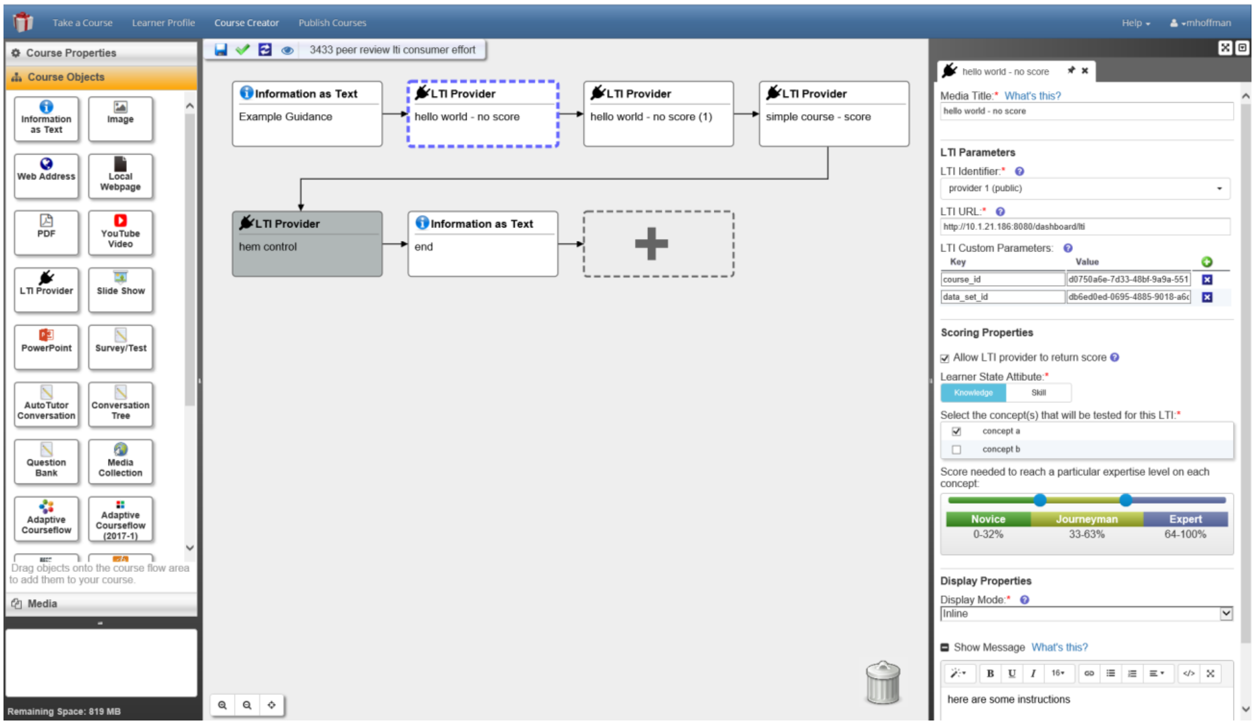
The Virtual Human Toolkit (VHTk) is a collection of modules, tools, and libraries designed to aid and support researchers and developers with the creation of virtual human conversational characters (Hartholt et al., 2013). It provides a way for users to generate virtual humans and integrate them across many projects.

Experiments have been performed to assess the ease of the creation of agents, with outputs driving tool design. VHTk is now available open source, and the characters that GIFT will use in the future are VHTk-based, with a VHTk-based agent planned to be available upon the cloud before the publication of this work.

LEARNING TOOLS INTEROPERABILITY

In previously publication, GIFT supported one part of a full LTI connection (Aleven et al., 2017; Brawner et al., 2017). This functional enabled GIFT to be part of an EdX course, or any other LTI Consumer. A GIFT course was run as part of an EdX course, through the LTI interface. EdX passed control of the module to GIFT, students took the GIFT course, and control was passed back to EdX. This flow of connection makes GIFT an "LTI Provider."

GIFT is now also an LTI Consumer, meaning that it can serve the same role as EdX did for GIFT – control during a GIFT course can be relinquished to an external training application, such a Cognitive Tutor exercise, and then returned back to GIFT with score reporting, which can be used elsewhere in the GIFT course per configurable assessment shown within Figure 2. This information can then be used later in the course.



Media Title:* What's this?
hello world - no score

LTI Parameters

LTI Identifier:* provider 1 (public)

LTI URL:* http://10.1.21.186:8080/dashboard/lti

LTI Custom Parameters:

Key	Value
course_id	d0750a6e-7d33-48bf-9a9a-551
data_set_id	db6ed0ed-0695-4885-9018-a6c

Scoring Properties

☑ Allow LTI provider to return score

Learner State Attibute:* [✓ Knowledge] Skill

Select the concept(s) that will be tested for this LTI:*
☑ concept a
☐ concept b

Score needed to reach a particular expertise level on each concept:

Novice	Journeyman	Expert
0-32%	33-63%	64-100%

Display Properties

Display Mode:* Inline

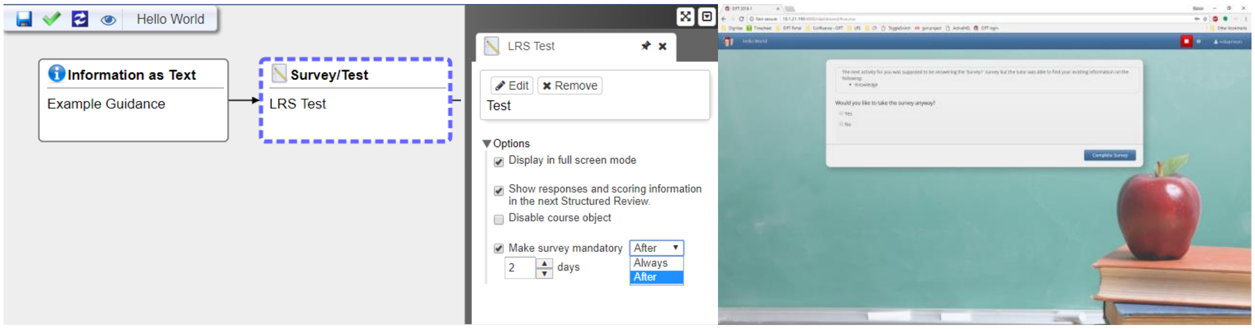
Show Message What's this?

here are some instructions

Figure 2: LTI handoff interface

LEARNER RECORD STORE

For a number of years GIFT has supported the functionality of reporting data to a Learner Record Store (LRS) in a configurable xml file. By default, this redirected to a publicly accessible LRS. At the time of writing, Figure 1 shows the connection of GIFT to a hosted LRS, which is now in active use in Cloud GIFT. A sample of the authoring and user interface settings for GIFT is shown in Figure 3. The coming developmental cycle will see the use of LRS data for filtering courses and for pulling learner information for future courses; creating an overarching learner profile used in many places. LRS data is planned to be able to be used across a wide variety of other systems from other Governmental agencies, such as within the Competency and Skills System (CASS).



Test

Options

☑ Display in full screen mode

☑ Show responses and scoring information in the next Structured Review.

□ Disable course object

☑ Make survey mandatory After ▾ Always After

2 days

The next activity for you was supposed to be answering the "Survey1" survey but the tutor was able to find your existing information on the following:

- Knowledge

Would you like to take the survey anyway?

○ Yes

○ No

Complete Survey

Figure 3: LRS Survey configuration and user experience

AUTHORING

The previous GIFT Symposium put forth the idea of creating a GIFT Course Wizard, which walks a novice author through the process of creating a course, eventually leading them to a created course on the existing course creator page (Murray, Pico, Redmon, & Rowan, 2017). This process has not been implemented, but efforts have been made to streamline the authoring tools, and to help novice authors with the creation of the Quick Start Guide (Ososky, 2016) and the GIFT YouTube video series mentioned earlier.

The most challenging area of authoring remains to be the authoring of the assessment logic which occurs within simulations. In the public example GIFT courses, the reader can see assessment logic configurations for the following, with the following:

- PowerPoint courses
  - over/under-dwell assessments
- Unity simulations
  - Assessment based on button-click events
- Medical training scenarios
  - many domain-specific assessments, such as time to apply tourniquet
  - Assessment is handled with external assessment engine called SIMILE (Mall & Goldberg, 2014)
- Excavator training scenarios
  - Assessment based on movement of the machine
- VBS training scenarios

- Assessment based on learner movements and actions

The overwhelming challenge is how to support authoring of this diverse set of assessments, without requiring coding knowledge, in a manner independent of the simulation, preferably while authoring simulation scenarios. Further, this functionality should be available for domain experts who are not experts in instruction, simulations, or GIFT.

Technically, what this authoring tool authors is a Domain Knowledge File, which contains a hierarchal task breakdown of the domain in the form of tasks, conditions, and standards. In the authoring tool, at the time of writing, this is called a "Real Time Assessment" and is authored as a series of Tasks, Concepts and Conditions. A project for building this capability has been ongoing and the functionality that is has developed is reported elsewhere within these proceedings (F. Davis, Riley, & Goldberg, 2018) , as requested and needed in the previous GIFTSym proceedings (F. C. Davis, Riley, & Goldberg, 2017; Ososky, 2017). The functionality of the new tool is anticipated to be deployed in the upcoming GIFT release.

UPCOMING RESEARCH DIRECTIONS: TEAM AND PSYCHOMOTOR TRAINING

Part of the goal of the GIFT project is to expand tutoring systems from relatively well-defined domains to ill-defined domains, from desktop training to "in the wild" training, and from individual training to team training. This is part of the military interest in intelligent tutoring technologies – Warfighters train as a group, and within the training environment.

Team Training

In the realm of team training, the GIFT project has recently finished a project reviewing the literature for what works with team instruction (RA Sottilare et al., 2017). Further, a number of small studies of teams were completed by the team at Iowa State University (Gilbert et al., 2017). These research studies were useful for the initial assessment of the team models, although are lacking in a number of manners. As part of these research discoveries, the system is being re-architected in a manner so as to support team "roles", with tutoring being role specific, but not team-member specific. The reasoning behind these decisions can be read within other research papers (Brawner, Sinatra, & Gilbert, 2018). Specific research implementations can be read elsewhere within this proceedings (Sinatra, 2018).

Psychomotor Training

Psychomotor, or "in the wild" training is a significant part of the reason for military investments in the intelligent tutoring technologies. As part of this effort, work within the domain of marksmanship has been well-published (B. Goldberg, Brawner, Amburn, & Westphal, 2014). Since the previous GIFTSym, the GIFT project has put measures in place to support training of tactical breathing (Kim, Sottilare, & Brawner, 2018) and land navigation. It does so through the use of a mobile application which reads and reports sensor data for physical actions or positioning, respectively, reported to the GIFT server. In prototype fashion, this has worked for one experiment, and a second experiment has been scheduled. The GIFT Mobile App is available upon request, but, at the time of writing, has not been fully tested for functionality.

OTHER NEW FUNCTIONALITY

There are a number of other features which have completed their experimental and developmental cycle are a now either scheduled for integration and deployment, as urged in prior GIFTSym publications, or completed. For the sake of completeness, these are included in the below list:

- Copy Course, downloadable in the latest release, deployed to Cloud GIFT
- VBS3 support, downloadable in the latest release
- Unity support, downloadable in the latest release
- Importing surveys from Qualtrics, downloadable in the latest release, deployed to Cloud GIFT
- Microsoft Band support, downloadable in the latest release
- Adaptive After Action Review (Brawner, Carlin, et al., 2018; Carlin, Brawner, Nucci, Kramer, & Oster, 2017; Carlin et al., 2018), scheduled at time of writing

GIFT AND IEEE STANDARDS

As part of this year's GIFT Symposium, there is an associated standards meeting. This standards meeting will be among those which occurred over the course of the year, including telephone calls, in-person meetings, proceedings presentations, and other activities. The IEEE Learning Technologies Standards Committee, with support from the GIFT community and the Government, is now seeking involvement in standardization activities. The GIFT community invites the reader to join the conversation on what data exchange standards for learning technologies might look like in the future – there is now active IEEE community on the subject, to which the GIFT project is contributing meaningfully. Interested readers are encouraged to go to www.instructionalsciences.org or the IEEE LTSC meetings to become involved.

REFERENCES

Aleven, V., Baker, R., Long, R., Sewall, J., Andres, J. M., Wang, Y., … Blomberg, N. (2017). *Integrating MOOCs and Intelligent Tutoring Systems: edX, GIFT, and CTAT.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5), Orlando, FL.

Biddle, E., Lameier, E., Reinerman, L., Matthews, G., & Boyce, M. (2018). *Personality: A Key to Motivating our Learners.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), Orlando, FL.

Brawner, K., Carlin, A., Oster, E., Nucci, C., & Kramer, D. (2018). *Adaptive, Policy-Driven, After Action Review in the General-ized Intelligent Framework for Tutoring.* Paper presented at the human Computer and Intelligent Interaction, Las Vegas, Nevada.

Brawner, K., Heylmun, Z., & Hoffman, M. (2017). *The GIFT 2017 Architecture Report.* Paper presented at the Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5).

Brawner, K., & Ososky, S. (2015). *The GIFT 2015 Report Card and the State of the Project.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3), Orlando, FL.

Brawner, K., Sinatra, A., & Gilbert, S. (2018). Lessons Learned Creating a Team Tutoring Architecture: Design Reconsiderations. In R. Sottilare, A. Graesser, X. Hu, & A. Sinatra (Eds.), *Design Recommendations for Intelligent Tutoring Systems: Teams (Volume 6).* Army Research Laboratory.

Carlin, A., Brawner, K., Nucci, C., Kramer, D., & Oster, E. (2017). *Educational Data Mining using GIFT Cloud.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5), Orlando, FL.

Carlin, A., Nucci, C., Kramer, D., Oster, E., & Brawner, K. (2018). *Data Mining for Adaptive Instruction.* Paper presented at the Florida Articificial Intelligence Research Society (FLAIRS), Melbourne, FL.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science, 1*(1), 73-105.

Davis, F., Riley, J., & Goldberg, B. (2018). *Iterative Development of the GIFT Wrap Authoring Tool.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), Orlando, FL.

Davis, F. C., Riley, J. M., & Goldberg, B. S. (2017). *Development of an Integrated, User-Friendly Authoring Tool for Intelligent Tutoring Systems.* Paper presented at the Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5).

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., . . . Winer, E. (2017). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Education*, 1-28.

Goldberg, B., Brawner, K., Amburn, C., & Westphal, M. (2014). *Developing Models of Expert Performance for Support in an Adaptive Marksmanship Trainer*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Goldberg, B. S. (2018). *Instructional Models in the Generalized Intelligent Framework for Tutoring: 2018 Update.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), Orlando, FL.

Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., . . . Gratch, J. (2013). *All together now: Introducing the Virtual Human Toolkit.* Paper presented at the International Workshop on Intelligent Virtual Agents.

Kim, J., Sottilare, R., & Brawner, K. (2018). *Integrating Sensors with GIFT to Maximize Data Exploitation for Improved Learning Analytics.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), Orlando, FL.

Mall, H., & Goldberg, B. (2014). *SIMILE: An Authoring and Reasoning System for GIFT*. Paper presented at the GIFTSym2, Pittsburgh, PA.

Murray, S., Pico, C., Redmon, K., & Rowan, C. (2017). *GIFT Course Creator Wizard.* Paper presented at the Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5).

Nye, B., Auerbach, D., Mehta, T., Hartholt, A., & Fast, E. (2017). *Building a Backbone for Multi-Agent Tutoring in GIFT.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5), Orlando, FL.

Ososky, S. (2016). Generalized Intelligent Framework for Tutoring (GIFT) Cloud / Virtual Open Campus quick start guide. (ARL-CR-0796). Orlando, FL: US Army Research Laboratory.

Ososky, S. (2017). *The 2017 Overview of the GIFT Authoring Experience.* Paper presented at the Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5).

Ososky, S., & Brawner, K. (2016). *The GIFT 2016 community report.* Paper presented at the Proceedings of the 4th Annual GIFT Users Symposium.

Robson, E., Ray, F., Sinatra, A. M., & Sinatra, A. M. (2017). *Integrating the outer loop: Validated tutors for portable courses and competencies.* Paper presented at the Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5).

Rowe, J., Pokorny, B., Goldberg, B., Mott, B., & Lester, J. (2017). *Toward Simulated Students for Reinforcement Learning-Driven Tutorial Planning in GIFT*. Paper presented at the Proceedings of R. Sottilare (Ed.) 5th Annual GIFT Users Symposium. Orlando, FL.

Sinatra, A. (2018). *Team Models in the Generalized Intelligent Framework for Tutoring: 2018 Update*. Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6), Orlando, FL.

Sottilare, R., Brawner, K. W., Goldberg, B. S., & Holden, H. A. (2012). The Generalized Intelligent Framework for Tutoring (GIFT).

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: US Army Research Laboratory. May 2017. DOI: 10.13140/RG.2.2.12941.54244.

Sottilare, R., Burke, C., Salas, E., Sinatra, A., Johnston, J., & Gilbert, S. (2017). Towards a design process for adaptive instruction of Teams: A Meta-Analysis. *International Journal of Artificial Intelligence in Education*.

ABOUT THE AUTHORS

***Keith Brawner, PhD** is a researcher for the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED), and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He has 12 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current research is in 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.*

***Michael Hoffman** is a senior software engineer at Dignitas Technologies and the technical lead for the GIFT project. He has been responsible for ensuring that the development of GIFT, meeting community requirements, and supporting production ITS systems, ITS research, and the growing user community. Michael manages and contributes support for the GIFT community through various mediums including the GIFT portal (www.GIFTTutoring.org), annual GIFT Symposium conferences and technical exchanges with ARL and their contractors. In addition he utilizes his expertise in integrating third party capabilities such as software and hardware systems to enable other organizations to integrate GIFT into their training solutions.*

Potential to Migrate ElectronixTutor to GIFT

Andrew J. Hampton[1], Xiangen Hu[1], Arthur C. Graesser[1], Zhiqiang Cai[1], & Andrew C. Tackett[1]
The University of Memphis, Institute for Intelligent Systems[1]

INTRODUCTION

Integrating disparate learning resources into a common framework presents several standard challenges. The learning resources are potentially diverse: texts, videos, diagrams, VR, open-ended and multiple-choice questions, natural language tutoring, simulations, and so on. How do we equate progress from one system to another? How do we assess a learner's progress within a learning resource and across resources? How do we recommend the best way forward for the learner? How do we handle different roles of users? All these critical questions have answers arising from the structure of the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare et al., 2012a; 2013). However, if a system has not been created in GIFT from the ground up, the potential for migrating into that structure requires careful consideration.

ElectronixTutor represents the culmination of several years of development in electrical engineering intelligent tutoring systems (ITS) (Graesser et al., 2018). In the Office of Naval Research STEM Grand Challenge, the University of Memphis is leading an effort to integrate several separately developed computer-based learning aids on the topic of electronic circuits. The resulting system constitutes an expansive, adaptive pedagogical tool with the potential to substantially elevate conventional instruction. This paper discusses the commonalities and differences of ElectronixTutor and GIFT, with an eye toward migrating the innovative functionality and breadth of the former into the standardized structure established by the latter. There are two primary reasons for this migration: (1) To improve the quality of existing content by following GIFT standards, and (2) To allow easier expansion of content and learning resources.

COMMON FEATURES IN ELECTRONIXTUTOR AND GIFT

Current implementation of ElectronixTutor is in the form of Moodle (version 3.4.1). Moodle (Dougiamas & Taylor, 2003) is a learning platform or course management system. It is a free open source software package designed to help educators create effective online courses based on sound pedagogical principles (http://www.moodle.org) and it is now the most popular adapted open source learning management system worldwide, notably used by US government agencies such as Advanced Distributed Learning and the Office of Personnel Management.

GIFT (www.gifttutoring.org) is an empirically-based, service-oriented framework of tools, methods and standards to make it easier to author computer-based tutoring systems (CBTS), manage instruction and assess the effect of CBTS, components and methodologies (Sottilare et al., 2012b). GIFT is being developed under the Adaptive Tutoring Research Science & Technology project at the Learning in Intelligent Tutoring Environments Laboratory, part of the U.S. Army Research Laboratory's Human Research and Engineering Directorate.

There are high-level similarities between GIFT and ElectronixTutor. The similarities include, but not limited to:

1. ElectronixTutor (as a Moodle implementation) and GIFT are open source and highly used by learning organizations. While Moodle is used widely by various learning organizations, ElectronixTutor and GIFT are primarily for the government/military of the United States.

2. ElectronixTutor (not necessarily Moodle) and GIFT are especially designed to integrate theory-driven, research-based learning resources.

3. ElectronixTutor and GIFT use the same standards-based learning behavior data repository. ElectronixTutor utilizes a module to connect Experience Application Programming Interface (xAPI, Advanced Distributed Learning, 2016) and a Learning Record Store (LRS, such as LearningLocker, https://learninglocker.net/). GIFT has a utility that sends xAPI statements to the LRS.

4. ElectronixTutor and GIFT have a Learning Content Management System with built-in authoring tools for native learning resources.

DIFFERENCES BETWEEN ELECTRONIXTUTOR AND GIFT

There are a few distinctive features that differentiate ElectronixTutor (Moodle) and GIFT that need special attention when we migrate ElectronixTutor to GIFT. Some of the distinctions are technological in nature, whereas others are based on application details.

1. Moodle and GIFT are implemented using different underlying technologies. The Moodle interface is HTML5 generated by PHP pages with a backend mySQL relational database. The open source nature is applicable to almost every aspect of the application, including module integration and a look & feel theme integration (and responsive design) that fits a variety of client platforms. GIFT was originally designed for military use and has much more restricted underlying technology. It is less flexible, but more stable with specially designed modules.

2. Moodle is optimized as an Internet application and best used as a browser-based application[1], such that there are no limitations on the source of the learning material as long as it is accessible. GIFT has two versions: a cloud-based version and standalone version. While the cloud-based version is similar to Moodle, where there are no special limitations on the source of the learning content, the stand-alone version limits the source of the learning content. This limitation requires that all learning resources are from authenticated sources (in the current implementation of GIFT, they need to be from *.gifttutoring.org). This limitation will have some impact when we migrate ElectronixTutor to GIFT, if we want to have a GIFT version of ElectronixTutor as a standalone learning platform.

3. The GIFT domain knowledge is an XML file that contains the information needed to execute a single lesson. The information in this file is essential for other GIFT modules, such as the learner module and the pedagogy module. ElectronixTutor does not have (and is not intended to have) detailed information within each of the integrated learning resources. For example, when ElectronixTutor selects one of its component resources, such as AutoTutor or Dragoon, ElectronixTutor only uses limited information from the instantiated lesson because it is run on a different server and may use a different pedagogy. ElectronixTutor only requires that the learning resource returns a value of 0 to 1 on an associated knowledge component and/or topic.

[1] There are mobile applications made for Moodle, so there is a non-browser version of Moodle. However, the viewing of the learning content still uses a browser on mobile devices.

SYSTEM MAPPING FROM ELECTRONIXTUTOR TO GIFT

Given the similarities and differences between Moodle-based ElectronixTutor and GIFT detailed above, we consider the following mappings between ElectronixTutor and GIFT.

From ElectronixTutor Knowledge Components to GIFT Concepts

The substantial challenge in creating a single, sensibly integrated system like ElectronixTutor includes determining a way to have the component systems communicate with one another in a mutually comprehensible way. To do this, we use Knowledge Components (KCs) as basic units at the conceptual level (Koedinger, Corbett, & Perfetti, 2012). These KCs map onto skills or information in electrical engineering and periodically appear in a given learning resource. Each learning resource can contribute a score (varying from 0 to 1) on a given KC or combination of KCs for a problem presented to the learner. Scores contribute to the learner's level of mastery on that KC.

KCs are analogous to the GIFT *concept* whose assessments are conveyed via game state messages. This structure allows them to be integrated as game state messages with two variables: name and value. In a migrated GIFT/ElectronixTutor system, the *MessageTypeEnum* would be updated to include "SaveKCScore" as a message type, which would be the message type sent each time a learner completes an item and generates a KC (concept) score. Conditions that assess the game state messages could simply return the name/value pair provided.

From ElectronixTutor Learning Resources to GIFT Modules/Lessons

ElectronixTutor includes several distinct learning resources that range from ITSs to conventional learning aids. They include simple multiple-choice questions that provide feedback and adaptivity (BEETLE-II, LearnForm), questions on skill building (ASSISTments), component manipulation and simulated circuit problems (Dragoon), and conversational deep reasoning and knowledge checks (AutoTutor). These intelligent and adaptive systems complement more traditional static resources such as topic summaries and Navy manual readings (NEETS). These learning resources are analogous to the GIFT Learner Module. Each learning resource could be integrated as a course object.

Of all these modules, some (such as Dragoon and ASSISTments) are integrated as external applications. Others (hypertext such as videos, slides, etc.) can be re-authored and improved using existing GIFT course objects for topic introductions, and conventional surveys or tests for assessments. The most complicated resource, AutoTutor, is already an object as part of GIFT. As we have pointed out earlier, if ElectronixTutor resources are external (such as Dragoon and ASSISTments), they will not be available for the standalone GIFT unless they are implemented with the authenticated servers (such as *.gifttutoring.org).

From ElectronixTutor Resource Organization to GIFT Domain Course File

In ElectronixTutor, learning resources are organized by our Recommender System, combining typical course progression and user characteristics to identify optimal next steps. These take the form of Topic of the Day and Recommended Items. Users can also self-direct learning. Within GIFT modules, learning resources within a course are organized by a domain course file. A domain course file is an XML file that contains the information needed to follow a single course, which may contain one or more lessons. The domain course file allows substantial control and flexibility in determining the flow between course objects. Externally integrated learning resources and GIFT-native resources are organized in the domain course file (in the form of XML). ElectronixTutor's resource organization and GIFT domain course file can be made structurally equivalent, so that ElectronixTutor's Recommender System can be mimicked by GIFT.

Topic of the Day

Determining what content to present to the learner at a given time is handled by our Recommender System. This pedagogical component considers a typical progression through electrical engineering education (roughly equivalent to a syllabus), where the learner has exhibited proficiency (from the Learning Record Store), and on which types of learning resources the learner has performed well or poorly. Topics always begin with a topic summary to orient (or reorient) the learner, then progress to a conversational reasoning question. These fall roughly in the upper-middle of the difficulty spectrum among the ITSs and hold the most potential for discriminating the level of proficiency among aspects of a single question. Based on performance, the Recommender System can send users "up" to the most difficult Dragoon problems, or "down" to multiple choice, decomposed circuit problems, skill builder items on Ohm's or Kirchhoff's laws, and possibly to summary static readings. This process involves differential determinations based on KCs constituent to the topic, so excellence in one area does not supersede the learning trajectory of another topic. The selection of learning resources for the topic of the day could be handled in GIFT's *pedagogical module.*

Recommended Items

Recommended items are generated from a combination of learner KC scores and pre-defined rules. Among these rules, topics are repeated if a learner's topic performance score falls below a threshold. Next there is a focus on underperforming knowledge components. Topics with medium performance scores and individual knowledge component scores below a threshold are recommended. In addition, we include an option to "push the envelope", where learners who often perform above a threshold receive resources that have a higher intrinsic difficulty. Finally, we have motivated and unmotivated "bottom dwellers", where bottom dweller is defined by topic performance scores often occurring below a threshold whereas motivation is determined by falling outside of processing time thresholds. The Recommender System is more complex than expressed here, but it is beyond the scope of this document to give a full specification.

These rules that the Recommender System uses are analogous to some but not all of the GIFT *strategies.* More generally, the Recommender System is handled by GIFT's pedagogical module and could be encoded as such in our migration. Further, the Recommender System's consideration for learner aptitude on various learning resources could dovetail with the ICAP framework (Chi & Wylie, 2014) available within GIFT. This framework delineates stages of interactivity with a learning system—interactive, constructive, active, and passive—that correspond neatly with current learning resources.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

ElectronixTutor is currently implemented within the open source Moodle infrastructure. Moodle provides different user roles, a common housing for learning content (in the "activity window"), and a broad community of users from which to draw inspiration or consult on obstacles. But Moodle does not afford analysis and alteration at a fine-grained level, a positive feature of GIFT (Sottilare et al., 2013). This paper identifies some of the structural similarities of ElectronixTutor and GIFT at a high-level specification with the ultimate goal of migration to GIFT. Our focus now turns to how that migration can proceed.

The challenges listed above, and their respective solutions in the current manifestation of ElectronixTutor, have many similarities to the GIFT architecture. First, GIFT's User Module is directly analogous to our

Learning Record Store. Both use xAPI and serve the purpose of informing the system of user characteristics relative to the system. The Domain Module in GIFT corresponds closely to the knowledge components and topic mastery adopted in ElectronixTutor. These domain-specific aspects of learning serve as the currency to evaluate learner progress. That progress is managed in GIFT by the Pedagogical Module, structurally similar to the Recommender System described above.

We are exploring the process of migrating ElectronixTutor from the Moodle infrastructure to GIFT. The primary challenges lie in the details. For example, the custom-made Recommender System serves a similar function to the Pedagogical Module, but the pedagogical rules are not exactly the same. Likewise, Moodle presents the Learning Resources in an idiosyncratic way, with unclear mappings to GIFT interface structures. This paper describes a preliminary evaluation of the challenges and opportunities for integration of the ElectronixTutor system within GIFT.

REFERENCES

Advanced Distributed Learning. (2016). The xAPI overview. Retrieved from https://www.adlnet.gov/xapi/.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243.

Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. Retrieved from https://research.moodle.net/33/1/

Graesser, A. C., Hu, X., Nye, B. D., VanLehn, K., Kumar, R., Heffernan, C., … & Andrasik, F. (2018). ElectronixTutor: An intelligent tutoring system with multiple learning resources for electronics. *International Journal of STEM Education: Innovations and Research*, 5(1), 15.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge‐Learning‐Instruction framework: Bridging the science‐practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757-798.

Sottilare, R., Brawner, K., Goldberg, B. & Holden, H. (2012a). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.

Sottilare, R., Goldberg, B., Brawner, K., & Holden, H. (2012b). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.

Sottilare, R., Holden, H., Goldberg, B., & Brawner, K. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottilare, R. (Eds.) *Fundamental Issues in Defence Simulation & Training*. Ashgate Publishing.

ABOUT THE AUTHORS

Dr. Andrew J. Hampton *manages the ElectronixTutor project at the Institute for Intelligent Systems, also contributing expertise in psycholinguistics and human factors.*

Dr. Xiangen Hu *is a professor of psychology and of computer science at the University of Memphis, while also serving as Dean of Psychology at Central China Normal University. He serves as co-PI of the ElectronixTutor effort.*

Dr. Arthur C. Graesser *has pioneered approaches to conversational intelligent tutoring systems and co-founded the Institute for Intelligent Systems. He serves as principal investigator of the ElectronixTutor effort.*

Mr. Zhiqiang Cai *serves as an assistant research professor in the Institute for Intelligent Systems, responsible for managing and implementing the knowledge component structure in AutoTutor.*

Mr. Andrew C. Tackett *is the Senior Research Software Developer at the Institute for Intelligent Systems, responsible for much of the implementation of ElectronixTutor in the Moodle framework.*

Ontology-driven Methods and a Framework for Enabling Hybrid Adaptive Team Training using Task and Sensor-based Performance Evaluation

Perakath Benjamin, Andrew Stephenson, and Kumar Akella
Knowledge Based System, Inc., College Station, TX

Rodney Long
Army Research Laboratory Human Research and Engineering Directorate Advanced Training and Simulation Division, Orlando, FL

INTRODUCTION

Adaptive Training is intelligently tailored, computer-guided experiences for individuals and units focused on optimizing training performance, training efficiency, deep learning, and transfer of skills to the operational environment. Training adaptation is multi-faceted. For example, training must adapt to the needs of the individual trainee as well as organizational groupings of trainees (e.g., an Army unit). Training must be tailored based on trainee and team state (cognitive, affective, social, etc.) and to trainee and team task performance. Adaptations might be determined and delivered in real time during training events or determined through assessment of learner data over extended time and delivered periodically (non-real time). Adaptations may seek to inform and optimize instructional strategies both during training and off-line (between training sessions). From a "training systems" life cycle perspective, the adaptation approaches must seek to optimize training over learner and team lifecycles through optimal blending of training types and modalities (e.g., computer-based, game-based, simulation-based, Live, Virtual, Constructive, and Game (LVC&G), etc.). A central barrier that impedes increased use of adaptive team training is the time and cost required to build and maintain these complex training applications. This paper describes an ontology-driven framework method that targets this challenge. The paper describes: (i) an ontology-driven method for hybrid (multi-domain, multi-task, multi-objective) adaptive team training; (ii) an enhanced Generalized Intelligent Framework for Tutoring (GIFT) architecture to support the hybrid adaptive team training method; (iii) sensor and task based individual and team performance evaluation approach; and (iv) hybrid adaptive training application examples that show the practical benefits of the method.

Current simulation-based training systems are incapable of dynamically generating and maintaining scenarios in an instructionally sound manner. Instead, scenarios are hand-crafted, static representations of training and mission contexts (Benjamin et al, 2012). The research described in this paper targets the multi-domain team adaptive training challenge – the ability to affordably build training applications in different domains that dynamically adapt training to rapidly changing learner needs. The long term goal of our research is to establish a multi-domain team, adaptive training capability suitable for application to a variety of warfighter contexts.

MOTIVATIONS

Current simulation-based training systems are incapable of dynamically generating and maintaining scenarios in an instructionally sound manner. Instead, scenarios are hand-crafted, static representations of training and mission contexts (Benjamin, Akella, Malek, & Fernandes, 2005). The research described in this paper targets the multi-domain hybrid team adaptive training challenge – the ability to affordably execute team trainings in different domains that dynamically adapt to rapidly changing learner needs. The long term goal of our research is to establish a multi-domain adaptive team training capability suitable for application to a variety of warfighter contexts.

Federated military simulation-based training exercises typically require the exchange of information between multiple warfighter functional areas and echelons. The complexity of mediating these information exchanges is intensified because of the multiplicity of simulation-based training tools and systems that are required in such training exercises.

Simulation-based training models require the representation of complex information structures. The information contained in these models depends on a systematic connection between the components of the representation and the real world. It is this connection that determines the semantic content of the data being represented. Generally, the semantic rules of a representation system for a given application of a simulation-based training tool and the semantic intentions of the tool designers are not advertised or in any way accessible to other agents in the warfighter organization. This makes it difficult for such agents to determine the semantic content of the simulation-based training models. We refer to this as the problem of *semantic inaccessibility* (Benjamin et al, 2005). This problem often manifests itself in different ways, including *unresolved ambiguity* (as when the same term is used in different contexts with different meanings) and *unidentified redundancy* (as when different terms are used in different contexts with the same meanings).

An important practical problem is – how to *determine* the presence of ambiguity and redundancy in the first place? In other words, how can we assess the semantics of simulation-based training data across different contexts? How can we define the semantics objectively in a way that permits accurate interpretation by agents outside the immediate context of this data? Our focus in this paper is to provide a solution approach to address this problem for simulation-based adaptive training applications that use GIFT.

GIFT

The Army Research Laboratory (ARL) is developing GIFT as part of its adaptive training research program. *Adaptive Training* is “intelligently tailored, computer-guided experiences for individuals and units focused on optimizing training performance, training efficiency, deep learning, and transfer of skills to the operational environment” (Sottilare, 2014). Training ‘adaptation’ can be multi-faceted. For the trainee, the delivery of training must adapt to individual trainee needs, as well as to the organizational groupings of trainees (e.g., an Army unit). Training must be tailored to trainee state (cognitive, affective, psychomotor, social, etc.) and to trainee task performance (Sottilare, 2013). Adaptations might be determined and delivered in real time during training events or determined through assessment of learner data over extended time and delivered periodically (non-real time). Adaptations may seek to inform and optimize instructional strategies both during training and off-line (between training sessions). Training content adaptations might be automated, semi-automated, or human (instructor)-driven. From a ‘training systems’ lifecycle perspective, the adaptation approaches must seek to optimize training through optimal blending of training types and modalities (e.g., computer-based, tutor-based, game-based, simulation-based, live training-based, etc.). In support of ARL’s adaptive training research, GIFT is being developed as open-source software, with a modular architecture whose goals are to reduce the cost and skill required for authoring adaptive training and educational systems, to automate instructional delivery and management, and to develop and standardize tools for the evaluation of adaptive training and educational technologies.

ONTOLOGY-DRIVEN METHOD FOR HYBRID TEAM ADAPTIVE TRAINING

The ontology-driven method for hybrid team adaptive training is summarized using the IDEF0 function modeling method (Figure 4).

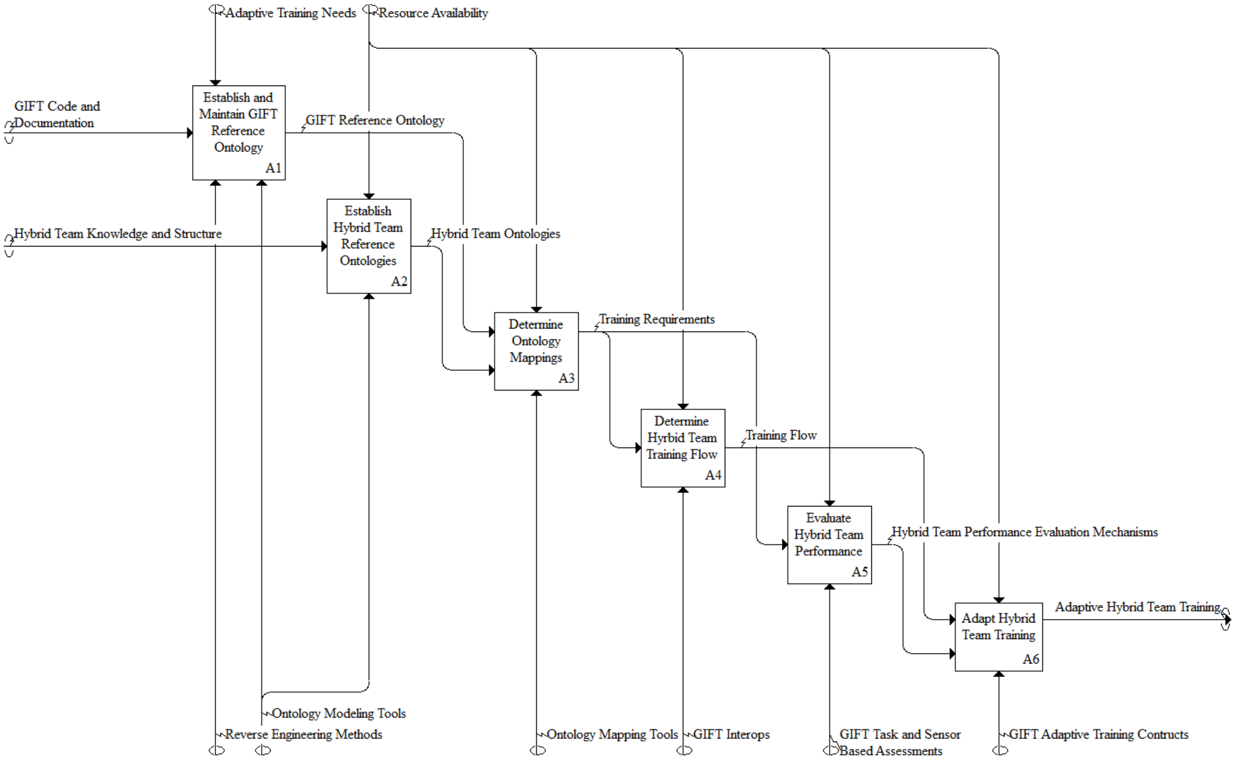

Adaptive Training Needs
Resource Availability
GIFT Code and Documentation
Establish and Maintain GIFT Reference Ontology
A1
GIFT Reference Ontology
Hybrid Team Knowledge and Structure
Establish Hybrid Team Reference Ontologies
A2
Hybrid Team Ontologies
Determine Ontology Mappings
A3
Training Requirements
Determine Hyrbid Team Training Flow
A4
Training Flow
Evaluate Hybrid Team Performance
A5
Hybrid Team Performance Evaluation Mechanisms
Adapt Hybrid Team Training
A6
Adaptive Hybrid Team Training
Ontology Modeling Tools
Reverse Engineering Methods
Ontology Mapping Tools
GIFT Interops
GIFT Task and Sensor Based Assessments
GIFT Adaptive Training Contructs

Figure 4. Ontology-Driven Training Application Integration Method

The steps of the ontology-driven training application integration approach are: 1) Establish and Maintain GIFT Ontology; 2) Establish Hybrid Team Reference Ontologies; 3) Determine Ontology Mappings; 4) Determine Hybrid Team Training Flow; 5) Evaluate Hybrid Team Performance; and 6) Adapt Hybrid Team Training. These activities are described in greater detail in the following paragraphs.

Establish and Maintain GIFT Ontology

The creation and maintenance of a GIFT ontology is an important first step towards building GIFT-enabled integrated simulation-based training applications. The multifaceted GIFT ontology includes concepts such as course, scenario, task, assessment, and conditions, in addition to classes, vocabulary, and attributes (Figure 5). An important aspect of the ontology are the relationships between the concepts and the cardinality restrictions of GIFT attributes. An initial GIFT ontology has been developed (Benjamin et al 2016). Once the GIFT ontology has been developed, it needs to be maintained over extended time.

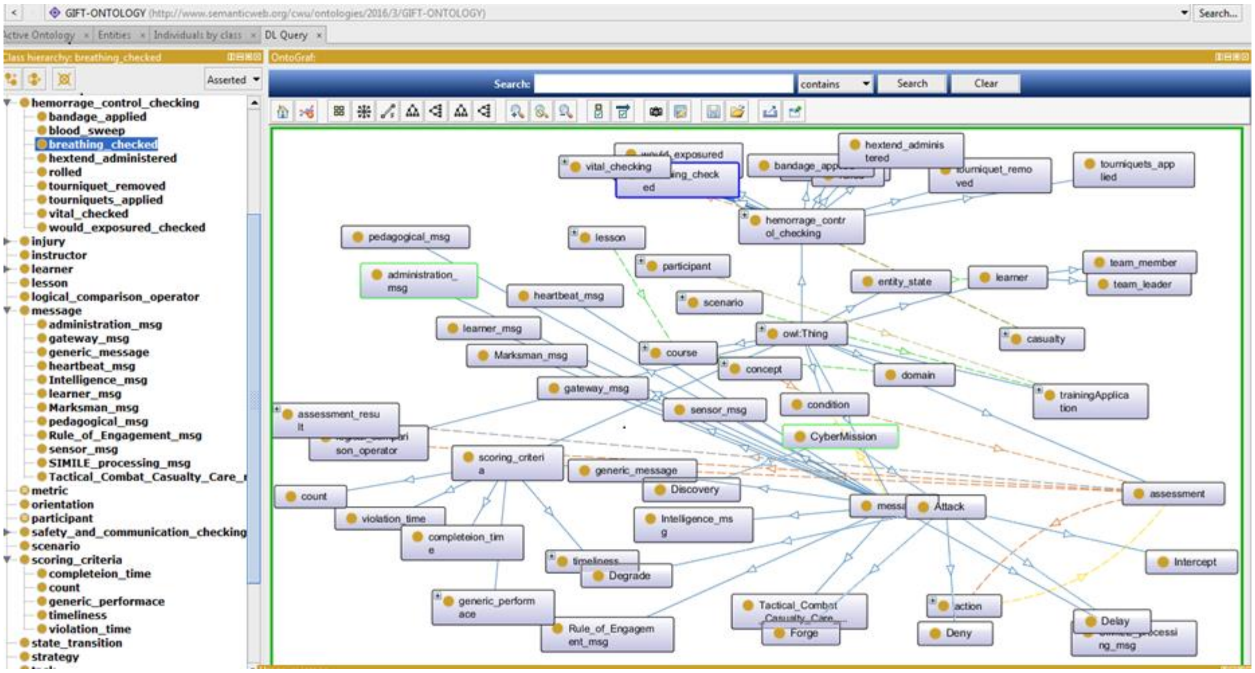


Figure 5. GIFT Ontology Fragment

Hybrid Team Ontologies

This important activity formulates mappings between the GIFT Ontology and the ontologies of the team or teams involved in the training (Figure 6). Note that for a given military training event, a federation of several simulation tools and models often need to be integrated and made to work together in an effective manner that addresses the warfighter training objectives.

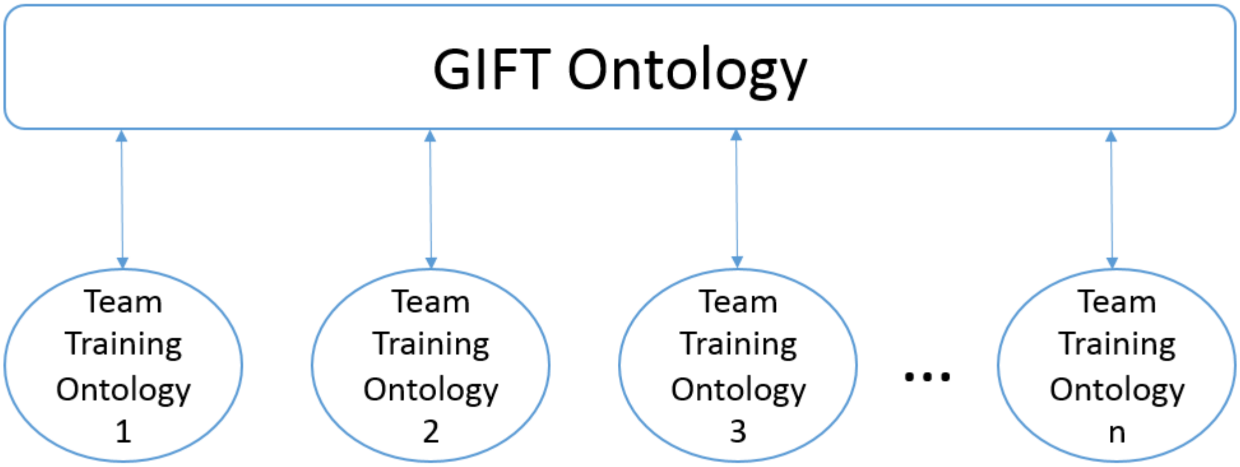


Figure 6. Mapping Team Ontologies to the GIFT Ontology

Such a method results in a framework that is configurable to different target domains through the insertion of an ontology of that target domain. We refer to this reconfiguration strategy using the term '***ontology-based***'. The basic idea is to create and maintain an adaptive-training reference ontology that is utilized to semantically determine and mediate the needed data and information exchanges between the adaptation framework elements and the core training system elements. Additionally, the ontology provides the adaptation framework with the domain knowledge required to evaluate team performance (using both task and sensor based methods).

Example Hybrid Team Ontology

An example application ontology for hybrid adaptive team training is shown in Figure 7. This ontology includes various components of an intelligence domain team interacting with a medical team to execute attending to wounded during attacks. The ontology is not meant to be comprehensive but is intended to provide basic constructs of an ontology that is useful for hybrid multi-domain adaptive team training. The ontology includes team member roles, training applications, tasks, dependencies (e.g. member 1 task 1 triggers member 2 task 2), and evaluation methods. If this type of ontology model was to be utilized by GIFT, it would need to be mapped to various components of the GIFT domain knowledge files (DKFs) for each member of team.

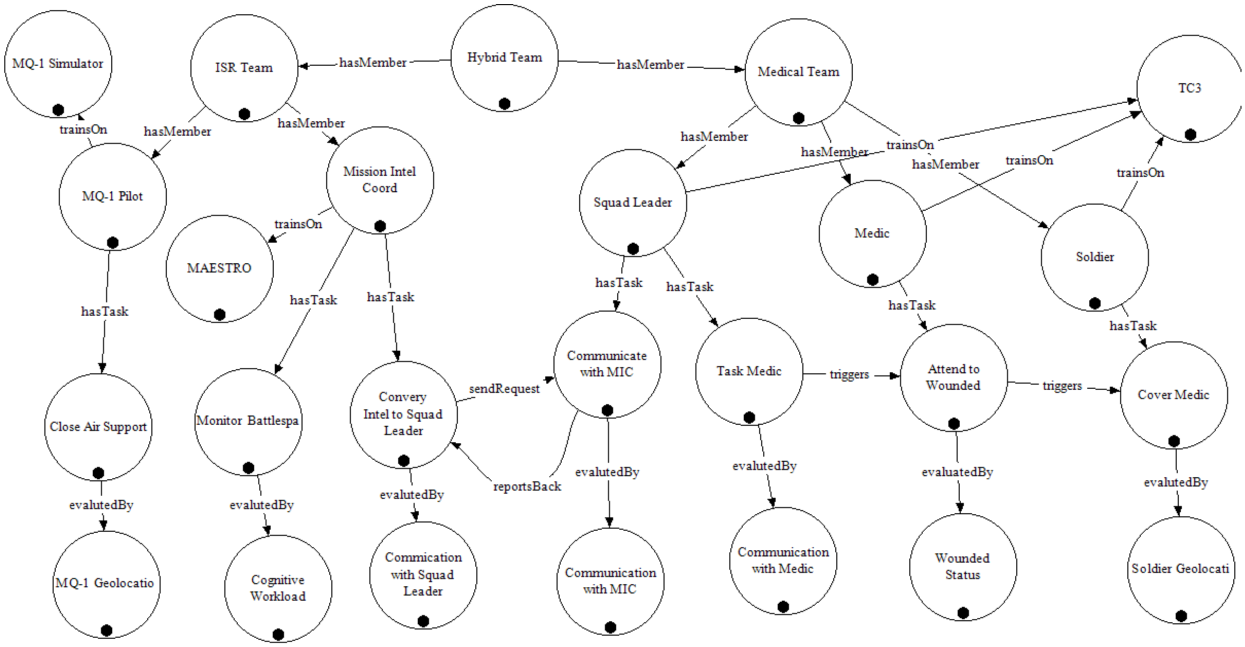


Figure 7. Notional Application Ontology Example for Hybrid Adaptive Team Training

Team Performance Evaluation Approach

The method incorporates the fusion of both sensor and task based team performance evaluation. Using an ontology (or set of ontologies), a system such as GIFT would help with the extraction of the necessary domain knowledge for robust hybrid team performance assessment. The method utilizes Bayesian data fusion techniques to integrate team sensor and task data in order to determine overall team performance scores.

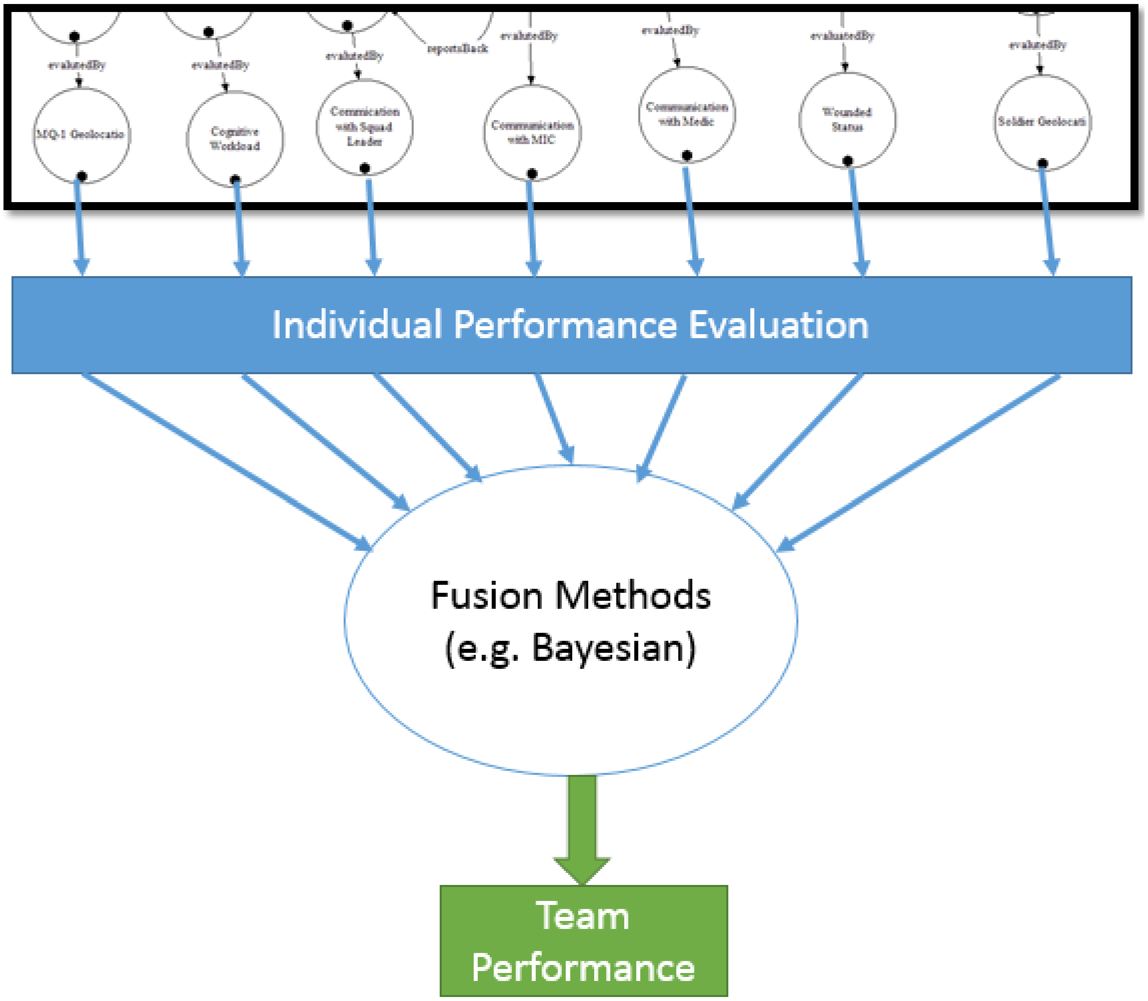


Figure 8. Evaluation Criteria Extracted from Ontology and Fused for Team Performance Evaluation

Sensor-Based Team Training Performance Evaluation

The sensor-based performance evaluation activity involves: (i) measuring cognitive indices from multiple sensors; and (ii) inferring cognitive states and trainee learning conditions using multi-senor data fusion. The reference ontology is used to: (i) select a set of cognitive indices and cognitive states relative to the training application objectives; (ii) adapt a multi-sensor data analyses suite to determine values of the selected indices and states; and (iii) map the cognitive states to trainee and team learning conditions. Figure 9 shows a high level concept of sensor data information fusion. It is assumed that artificial neural networks (or other analytical methods) are being utilized within the sensor module to evaluate learner state based on data coming from sensors.

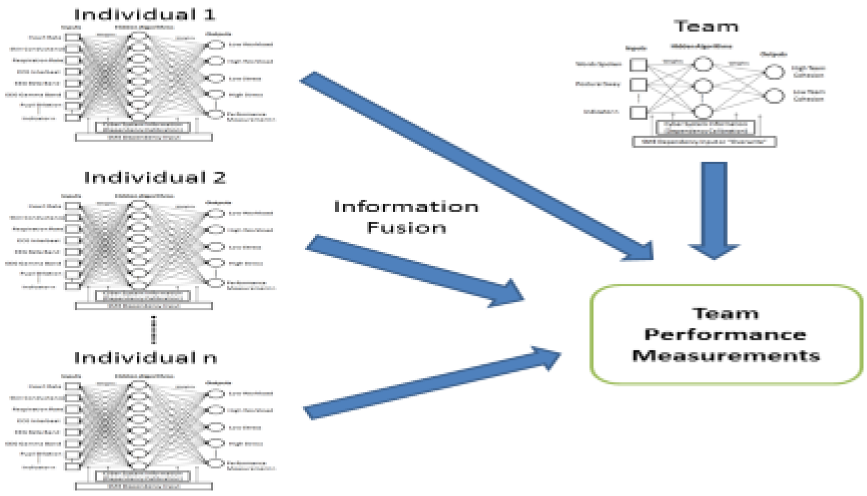


Figure 9. Sensor Data Fusion for Performance Evaluation

Task-Based Team Training Performance

The task-based measurements are tied to: (i) the overall mission outcomes; (ii) individual trainee skills; and (iii) team skills. The reference ontology is then utilized to determine how specific task data applies to the outcomes and skills (both individual and team). A rule-based approach is then utilized to encode the logic to compute the values of objective metrics from training system output data.

Team Training Performance Evaluation Example

for a proof of concept demonstration implementation of multi-domain adaptive team training using Tactical Combat Casualty Care Simulation (TC3 Sim) tool, the Generalized Intelligent Framework for Tutoring (GIFT) software, and KBSI's MAESTRO™ (ISR training tool) is described here. TC3 Sim is used to run a battlefield medical evacuation training scenario where the trainee is a medic who is embedded with a unit patrolling hostile streets. The squad leader is tasked with locating the village elder to discuss opportunities for local support and humanitarian aid. Intelligence reports indicate possible insurgent activity in the surrounding buildings. The unit is to secure the area while discussions are held to improve safety. When the unit is engaging insurgents, the medic should apply proper techniques of care under fire and tactical field care where appropriate. In parallel, MAESTRO™ is training intelligence personnel to gather real-time hostile data, process them, and feed situational awareness information to the squad leader in the TC3 Sim. The functions supported for training in the MAESTRO™ scenario are ISR supported by the Mission Intelligence Commander (MIC), CAS supported by MQ-1 and A-10 platforms, JTAC who also interacts with the squad leader in TC3 Sim scenario, and the Ground Force Commander (GFC).

In this notional multi-domain team training, two sets of tasked-based performance evaluation metrics have been designed, one set of metrics for the ISR Team training in MAESTRO™ and the second set of metrics for the Patrol Team training in TC3 Sim. The metrics for the ISR Team include: (1) did the MIC review the COP and send out follow up information on time? (2) did the MIC send the message to the right person? (3) did the MIC follow up with the person to whom he send the information? and (4) did the MIC use communication standards (with brevity and using the right terminology) while relaying information? Example metrics for the Patrol Team include: (1) was the criteria "stay close" violated? (2) by what margin (distance and time) did the team violate safe distance from building? (3) did the medic stop bleeding and stabilize victim? (4) did MEDEVAC process get initiated at the right time? and (5) did the Patrol Team leader send acknowledge message to the MIC after receiving recommendations? Each team is evaluated in their own environment and corrective (adaptation) strategies are used in each environment to rectify deficiencies. The advantage of using a multi-domain team in this example application situation is that there is real synergy with different team members complementing each other's effort while cooperatively working to achieve overarching and shared mission goals.

In addition to the task-based performance evaluation criteria, sensor-based measurements are also captured to determine the cognitive state of the MIC and the Medic. Example metrics for the ISR Team include (1) maintaining acceptable stress levels, (2) fatigue management, and (3) attentiveness. Metrics for the Patrol Team include (1) limiting nervousness (for example, manifested 'shaking') by the Medic, (2) alertness, and (3) maintaining acceptable stress levels.

Once the team performance evaluations are completed using the metrics described earlier, the individual trainees and the teams are 'graded'. The results of the performance evaluation were used to recommend adaptation strategies for (1) the two teams in MAESTRO and TC3Sim; (2) the ISR Team in MAESTRO; and (3) the Patrol Team in TC3 Sim. To illustrate, suppose that the MIC does not use communication standards to relay information and that the unit leader does not acknowledge the message after receiving information from the MIC, then recommending that both the teams (ISR and patrol) must review "communication standards" learning module is an example of an appropriate adaptation strategy. An example adaptation strategy for the ISR Team is as follows: when the MIC is overwhelmed because of 'information overload', he/she may not relay timely information to his/her squad leader. To rectify this deficiency, the ISR Team is introduced to several 'drills' (simple scenarios) to help them achieve higher levels of the "situational awareness" skill. If it is observed that a Patrol Team, in TC3 Sim, is violating the "stay close" criteria then the instructor would relax the 'distance margin' in order to help the team get more familiar with the team coordination effort and to better recognize uncertainties.

A GIFT-BASED ARCHITECTURE FOR MULTI-DOMAIN TEAM ADAPTIVE TRAINING

This section describes the two conceptual design options of a GIFT-based architecture for multi-domain hybrid team adaptive training using task and sensor based performance evaluation.

Overview

Currently, GIFT supports training in various domains with performance evaluations and adaptations specific to the training applications in those domains. Our goal is to enhance and extend GIFT so that it will be able to support multi-domain hybrid team adaptive training without the need for team-specific extensions to GIFT. In order to reach this goal, we have designed a method (outlined in the previous section) and two different architecture options for extending GIFT to support multi-domain hybrid team training. As noted in the previous section, the ontologies provide the basis for allowing GIFT to 'understand' team structures and appropriately adapt the training content. In GIFT terms, this would mostly involve an extension/plug-in utilized by the Domain Module. At the point of writing this paper, it is assumed that GIFT is being / has been extended to support team training (e.g. Team DKF, Team Model, and Team Pedagogy).

GIFT Architecture Extension Option 1

The first potential architecture extension (see Figure 10) is the less complex of the two options identified in this paper. It would include only extensions to the existing GIFT code base, with very little modification of the current code. This architecture would contain three new components/plugins/services: 1) an Ontology Mapper; 2) a DKF Builder; and 3) a Bayesian Fusion Engine. The Ontology Mapper would be utilized to map a team ontology to the GIFT ontology and the DFK Builder would build appropriate DKF files (both team and individual) based on the mappings. There would then be picked up and utilized by GIFT's current team and individual training execution and evaluation components. The third new component, the Bayesian Fusion Engine, would be utilized by the Learner/Team Module to fuse individual and team states into overall team performance states. In order for the Bayesian Fusion Engine to "know" how to fuse the states, the team DKF file would need to include state weighting information.

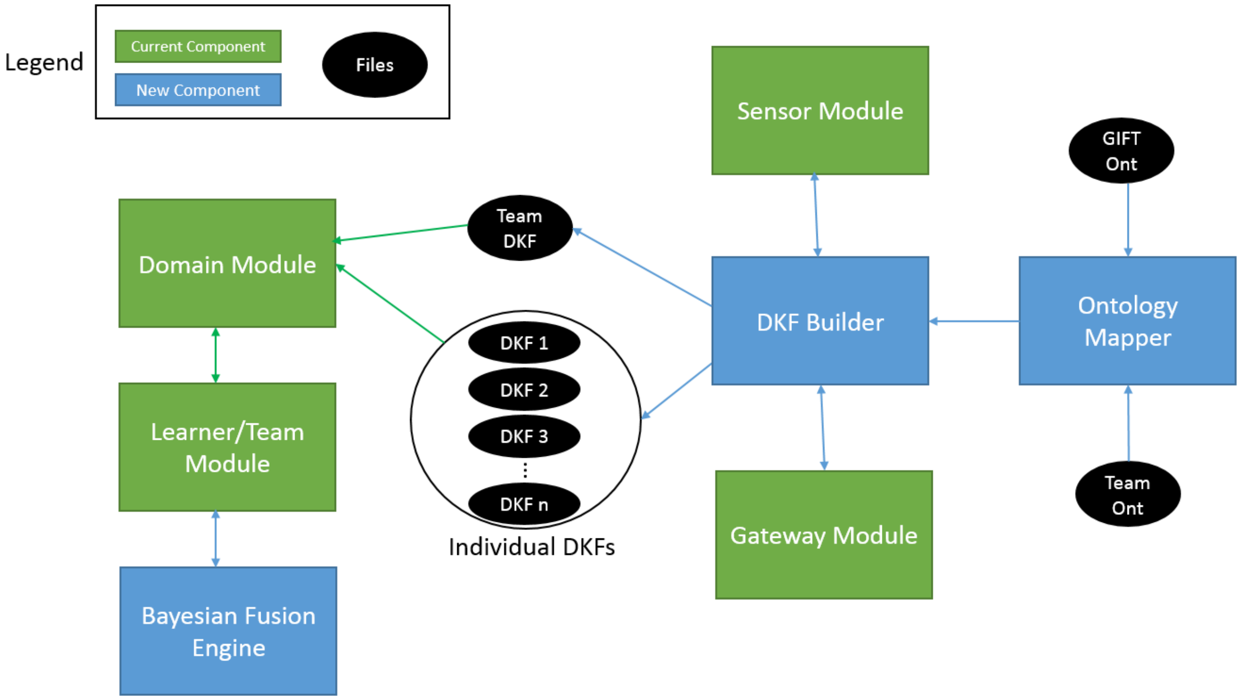


Figure 10. Option 1 GIFT Architecture Extensions

GIFT Architecture Extension Option 2

The second potential architecture enhancement (see Figure 11) would require more extensive modification to the existing GIFT code base. This architecture would contain two new components/plugins/services: 1) an Ontology Mapper; and 2) a Bayesian Fusion Engine. Additionally, the Domain Module would have to be modified so that it could not only interpret/read DKF files, but also various ontology files and format. The Ontology Mapper would be utilized by the Domain Module to map a team ontology to the GIFT ontology. The mapped ontology would then be utilized directly by the Domain Module to configure domain specifics of GIFT training sessions. Furthermore, similar to Option 1, the Bayesian Fusion Engine would be utilized by the Learner/Team Module to fuse individual and team states into overall team performance states.

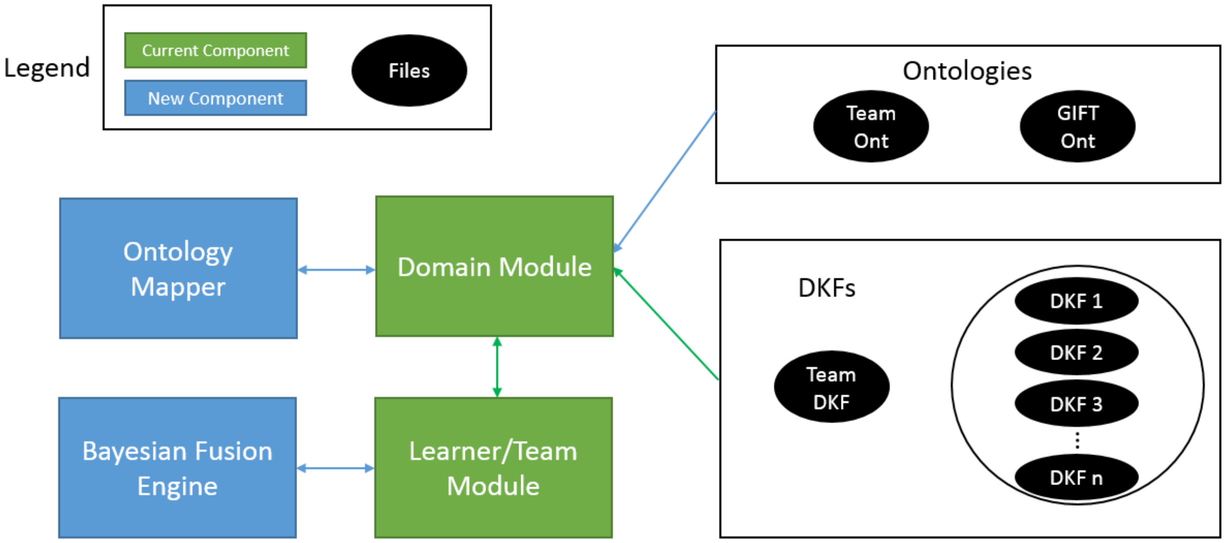
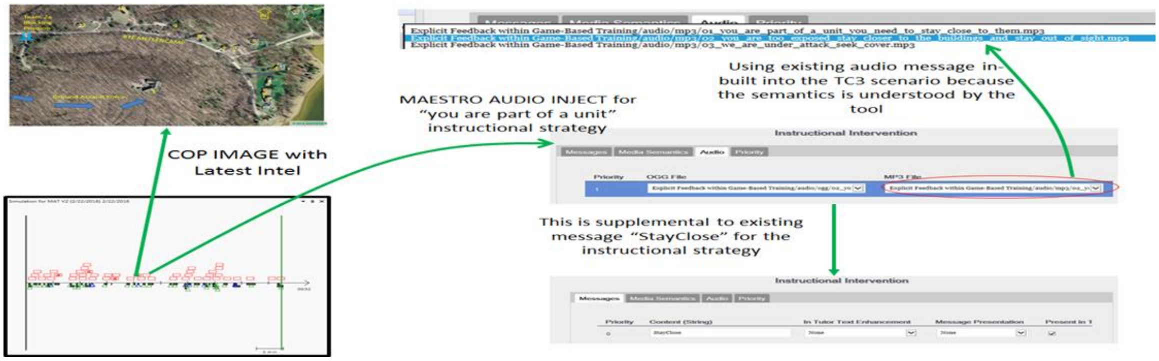


Figure 11. Option 2 GIFT Architecture Extensions

Example: Maestro^{TM} with GIFT and TC3 Sim for Multi-domain Team Training

The notional scenario outlined in the "Task-Based Team Training Performance" section is described in detail here. The previously descried architecture options would support this training example in GIFT. For the example, we will refer to the overall team, which includes the ISR Team and the Patrol Team, as the Hybrid Team. In Figure 12, the image of ground assault teams engaged in mission is presented as (Common Operational Picture) COP inject to the MIC in MAESTRO^{TM}. Looking at the image, the MIC should relay this information and follow up action (recommendation) to the unit on the ground within reasonable amount of time so that the relevance of the information remains current and timely. The recommendation can be relayed either via chat messages or audio messages and a notional message for this situation can take the form of "You are too exposed, stay closer to the buildings, and stay out of sight." This message is sent to GIFT software which is routed to the TC3 Sim scenario and evaluated by the learn module as a correct response (at expectation). As the unit is patrolling, suppose that an IED goes off at a distance and shrapnel hits a member of the unit. The medic embedded with the unit now initiates a 'victim stabilization' process and then GIFT evaluates the medic's performance as being 'at expectation'.


Explicit Feedback within Game-Based Training/audio/mp3/01_you_are_part_of_a_unit_you_need_to_stay_close_to_them.mp3
Explicit Feedback within Game-Based Training/audio/mp3/02_you_are_too_exposed_stay_closer_to_the_buildings_and_stay_out_of_sight.mp3
Explicit Feedback within Game-Based Training/audio/mp3/03_we_are_under_attack_seek_cover.mp3

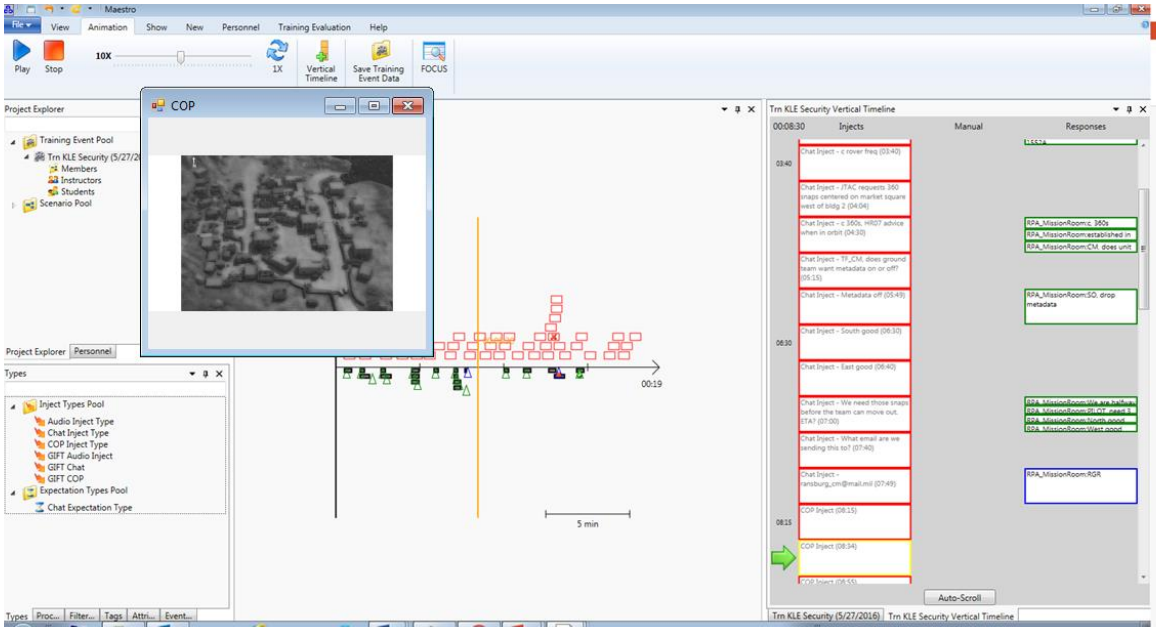
Using existing audio message in-built into the TC3 scenario because the semantics is understood by the tool

MAESTRO AUDIO INJECT for "you are part of a unit" instructional strategy

COP IMAGE with Latest Intel

This is supplemental to existing message "StayClose" for the instructional strategy

Figure 12: Information Flow between MAESTRO and TC3 Sim



Chat Inject - c rover freq (03:40)

Chat Inject - JTAC requests 360 snaps centered on market square west of bldg 2 (04:04)

Chat Inject - c 360s, HR07 advice when in orbit (04:30)

Chat Inject - TF_CM, does ground team want metadata on or off? (05:15)

Chat Inject - Metadata off (05:49)

Chat Inject - South good (06:30)

Chat Inject - East good (06:40)

Chat Inject - We need those snaps before the team can move out. ETA? (07:00)

Chat Inject - What email are we sending this to? (07:40)

Chat Inject - ransburg_cm@mail.mil (07:49)

COP Inject (08:15)

COP Inject (08:34)

COP Inject (08:55)

RPA_MissionRoom:c, 360s

RPA_MissionRoom:established in

RPA_MissionRoom:CM, does unit

RPA_MissionRoom:SO, drop metadata

RPA_MissionRoom:We are halfwa

RPA_MissionRoom:PILOT, need 3

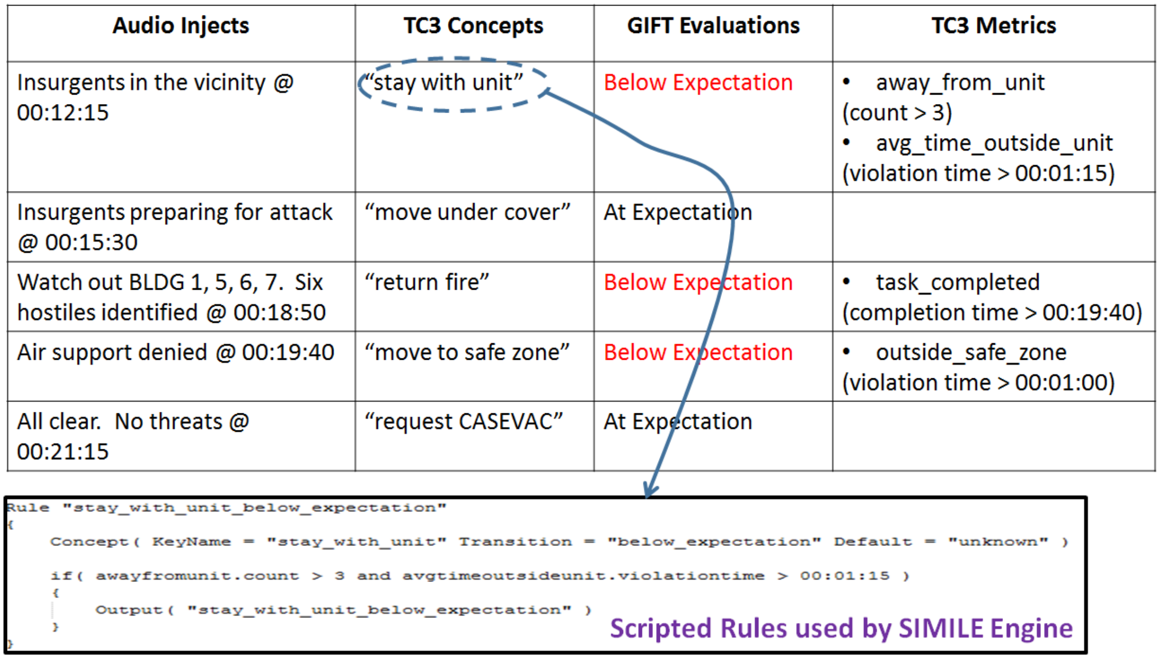
RPA_MissionRoom:North good

RPA_MissionRoom:West good

RPA_MissionRoom:RGR

Figure 13: ISR Team Simulation Data Captured in MAESTRO™

When a scenario is executed in MAESTRO^{TM}, incorrect responses, correct responses, and instructor recorded comments are logged and persistently stored in the database. The categories of performance metrics are late response, echo in wrong chat room, incorrect response, response in wrong chat room, positive tag (best practices of trainees), and negative tag (egregious mistakes observed by instructor). Responses to injects, in the form of chat messages, are evaluated in MAESTRO^{TM} as shown in the timeline view of Figure 13. MAESTRO^{TM} has the ability to persistently store trainees responses and this give the ability to collect vital statistics on their performance like how many times a trainee responded incorrectly, how many times trainee missed to response, average late response time of a trainee, etc. The performance metrics derived from MAESTRO^{TM} evaluation are sent to SIMILE workbench, performance evaluation engine in GIFT, to determine trainee's grade. For example, trainee's grade is set as below expectation if all the following conditions are met: (a) Echo in Wrong Chat Room > 3; (b) Incorrect Response >= 2; (c) Late Response > 3; (d) Negative Tag > 2; and (e) Positive Tag = 0. Likewise, other rules are scripted for at expectation and above expectation grades. These rules can be edited and tailored made for scenarios being trained. Adaptation rules can be developed to address observed deficiencies and recommend training scenarios for ISR Team in MAESTRO^{TM} tool.



Audio Injects	TC3 Concepts	GIFT Evaluations	TC3 Metrics
Insurgents in the vicinity @ 00:12:15	"stay with unit"	Below Expectation	• away_from_unit (count > 3) • avg_time_outside_unit (violation time > 00:01:15)
Insurgents preparing for attack @ 00:15:30	"move under cover"	At Expectation	
Watch out BLDG 1, 5, 6, 7. Six hostiles identified @ 00:18:50	"return fire"	Below Expectation	• task_completed (completion time > 00:19:40)
Air support denied @ 00:19:40	"move to safe zone"	Below Expectation	• outside_safe_zone (violation time > 00:01:00)
All clear. No threats @ 00:21:15	"request CASEVAC"	At Expectation	

```
Rule "stay_with_unit_below_expectation"
{
    Concept( KeyName = "stay_with_unit" Transition = "below_expectation" Default = "unknown" )

    if( awayfromunit.count > 3 and avgtimeoutsideunit.violationtime > 00:01:15 )
    {
        Output( "stay_with_unit_below_expectation" )
    }
}
```

Scripted Rules used by SIMILE Engine

Figure 14: Notional Evaluation Rules Scripted in the SIMILE Engine Using TC3 Sim Data

There are a total of 60 injects defined in MAESTRO^{TM} and five of those injects (audio format) are routed to TC3 Sim through GIFT software. The routed injects, initiated at certain times, are mapped to specific TC3 concepts. Three of the concepts – "stay with unit," "return fire," and "move to safe zone" – are found to be at 'below expectation' grade. The trainee responses in TC3 environment is logged, which are evaluated to derive several metrics such as the one listed under the TC3 Metrics column in Figure 14. Derived metrics are used to evaluate the TC3 concepts by SIMILE workbench engine, which uses scripted rules as shown in the figure. Adaptation rules can be developed to address observed deficiencies and recommend training scenarios for Patrol Team in TC3 Sim environment.

Now that performance states have been capture for both the TC3 trainees and MAESTRO^{TM} trainees, results are sent to the Bayesian Fusion Engine. The Bayesian Fusion Engine combines the performance states into a final team state, resulting in an at expectation grade for the team as a whole.

Adaptation rules are scripted for various performance grades to help trainees learn skills better by gradually presenting complex training concepts in a methodical way. Examples of adaptation rules for the ISR Team in MAESTRO^{TM} environment include: (1) for 'below expectation' performance grade -- remove any TWO role types, remove injects that have more than ONE expectation, and remove injects that have expectation duration of less than 40 seconds; (2) for 'at expectation' performance grade -- remove any ONE role type and remove injects that have expectation duration of less than 20 seconds; and (3) for 'above expectation' performance grade -- reduce ALL expectations duration by 30% and eliminate FEW (three to five) COP images to impact situational awareness. The first two adaptation rules are meant to reduce the complexity of the scenario so that the trainees can assimilate concepts better. Third adaptation rule will increase the complexity of the scenario and help trainees enhance skills.

Adaptation rules are scripted for various performance grades to help trainees learn skills better by gradually presenting complex training concepts in a methodical way. Examples of adaptation rules for the Patrol Team in TC3 Sim environment include: (1) if less than 15% of the concepts are graded to be 'below expectation' then have the trainees review PowerPoint presentation on TTPs of key concept areas and repeat the scenario;

(2) if 20% to 30% of the concepts are graded to be 'below expectation' then relax grading criteria on concepts, for example, the distance range for being away from the unit can be increased from 10 meters to 20 meters and with these adjustments the scenario can be repeated; and (3) if more than 30% of the concepts are graded to be 'below expectation' then the interface with the external team (ISR Team in MAESTRO^{TM}) can be removed and have the Patrol Team train exclusively within TC3 Sim environment. These adaptation rules are gradually reducing complexity based on logical reasoning with the purpose of helping trainees learn better.

Finally, adaption rules are built for various performance grades of the Hybrid Team as a whole. These can become very complex if individual performance grades were taken into consideration. For simplicity, we will consider only the overall team grade for these adaptation rules. Examples of Hybrid Team adaption rules include: (1) if 'below expectation', remove MAESTRO^{TM} injects containing more than one response expectation and send a PowerPoint presentation to the TC3 team to review TTPS; (2) if 'at expectation', increate simulation speed for MAESTRO^{TM} and TC3 by 10%; (3) if 'above expectation', reduce response time criteria for the ISR Team and the Patrol Team by 30%. In a more complex adaption configuration, team member weighting values could be utilized to determine more robust team adaptations.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The paper describes: (i) an ontology-driven method for hybrid adaptive team training; (ii) an enhanced Generalized Intelligent Framework for Tutoring (GIFT) architecture to support the hybrid adaptive team training method; and (iii) a hybrid adaptive team training application example that shows the practical benefits of the method. Innovative aspects of the research described in this paper include: (i) a new ontology-based approach for hybrid adaptive team training; (ii) a standards-compliant and component-based architecting strategy that allows for rapid and affordable deployment of the adaptive training framework; and (iii) the ability to automate the generation of adaptive training scenarios. Benefits include: (i) reduced training costs; (ii) improved team training effectiveness; (iii) reduced cognitive workload for instructors; (iv) significantly reduced time and effort for *semantic* knowledge sharing, communication, and *semantic integration* for distributed training applications; and (v) improvements in learner and team performance.

Areas that would benefit from R&D include: (i) methods for extending and generalizing the GIFT adaptive team training reference ontologies; (ii) design of automated support for ontology analysis and harmonization to support training application integration; (iii) design and implementation of inter-application information exchanges with GIFT for a broader range of training application areas; and (iv) design of mechanisms to mediate and exchange adaptive training content across multiple training modalities and types.

REFERENCES

Benjamin, P, Graul, M, Akella, K, Gohlke, J, Schreiber, B, and Holt, L. Towards Adaptive Scenario Management (ASM), *Proceedings of Inter-service / Industry Training, Simulation, and Education Conference 2012 (I/ITSEC 2012)*, Orlando, Florida, December 2012.

Benjamin, P, Akella, K, Malek, K, Fernandes, F. An Ontology-Driven Framework for Process-Oriented Applications. *Proceedings of the 2005 Winter Simulation Conference*, Orlando, FL, December 2005.

Sottilare, R. Adaptive Training Research: Guided Instruction for Individual and Unit Level Tasks in Support of FHTE-LS, October 2014.

Sottilare, R. Special Report: Adaptive Intelligent Tutoring System (ITS) Research in Support of the Army Learning Model – Research Outline. Army Research Laboratory (ARL-SR-0284), December 2013.

ABOUT THE AUTHORS

Perakath Benjamin *is a Vice President (R&D) at Knowledge Based Systems, Inc. (KBSI), He has over 30 years of professional experience in systems analysis, design, development, testing, documentation, deployment, and training. Dr. Benjamin has been responsible for the development of Artificial Intelligence, Knowledge Discovery, Knowledge Management, Planning and Scheduling, and Simulation Modeling technology applications that are being applied extensively throughout industry and government. Dr. Benjamin's expertise areas include ontology management, adaptive training, training content management, semantic technologies, semantic data and information integration, data analytics, and knowledge discovery.*

Andrew Stephenson *is a programmer analyst at Knowledge Based Systems, Inc. (KBSI), received his Bachelor's degree in Biomedical Engineering in 2012 from Wright State University and a Master's degree in Human Factors Engineering with a focus in Human Systems Integration in December 2015. Mr. Stephenson has been involved with several research and development projects involving adaptive intelligence training, cyber situation awareness measurement, physiological data processing and analysis, sensor integration and data fusion, and interactive data visualization design for the intelligence domain. He has been a key developer of KBSI's Maestro™ tool which is used to rapidly develop and execute pedagogically sound distributed intelligence training.*

Kumar Akella *is a research scientist at KBSI, received his Ph.D. in Mechanical Engineering in 1998 from Texas A&M University. He has 19 years of experience in discrete event simulation modeling, systems dynamic modeling, data mining analysis, and Bayesian analysis. Dr. Akella worked on several projects like adaptive training for warfighters, semantic search using NLP methods, scheduling workloads for depot-level maintenance and operations, and identifying patterns of interest using data mining and fusion methods.*

Rodney Long *is a Science and Technology Manager at the Army Research Laboratory in Orlando, Florida and is currently conducting research in adaptive training technologies. Mr. Long has a wide range of simulation and training experience spanning 28 years in the Department of Defense (DOD) and has a Bachelor's Degree in Computer Engineering from the University of South Carolina and Master's degree in Industrial Engineering from the University of Central Florida.*

THEME II: GIFT AUTHORING TOOLS

The GIFT Authoring Experience: 2018 Update

Rodney A. Long[1], Robert A. Sottilare, Ph.D.[1]
Army Research Laboratory 1

INTRODUCTION

Intelligent tutoring systems (ITSs) are computer-based adaptive instructional systems (AISs) "that guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of domain learning objectives" (Sottilare & Brawner, 2018, p. 25). In general, the more adaptive an ITS is, "the more content it needs to support tailoring and personalization of instruction – which also leads to longer development times and higher costs" (Sottilare & Fletcher, 2018, p. 1-2).

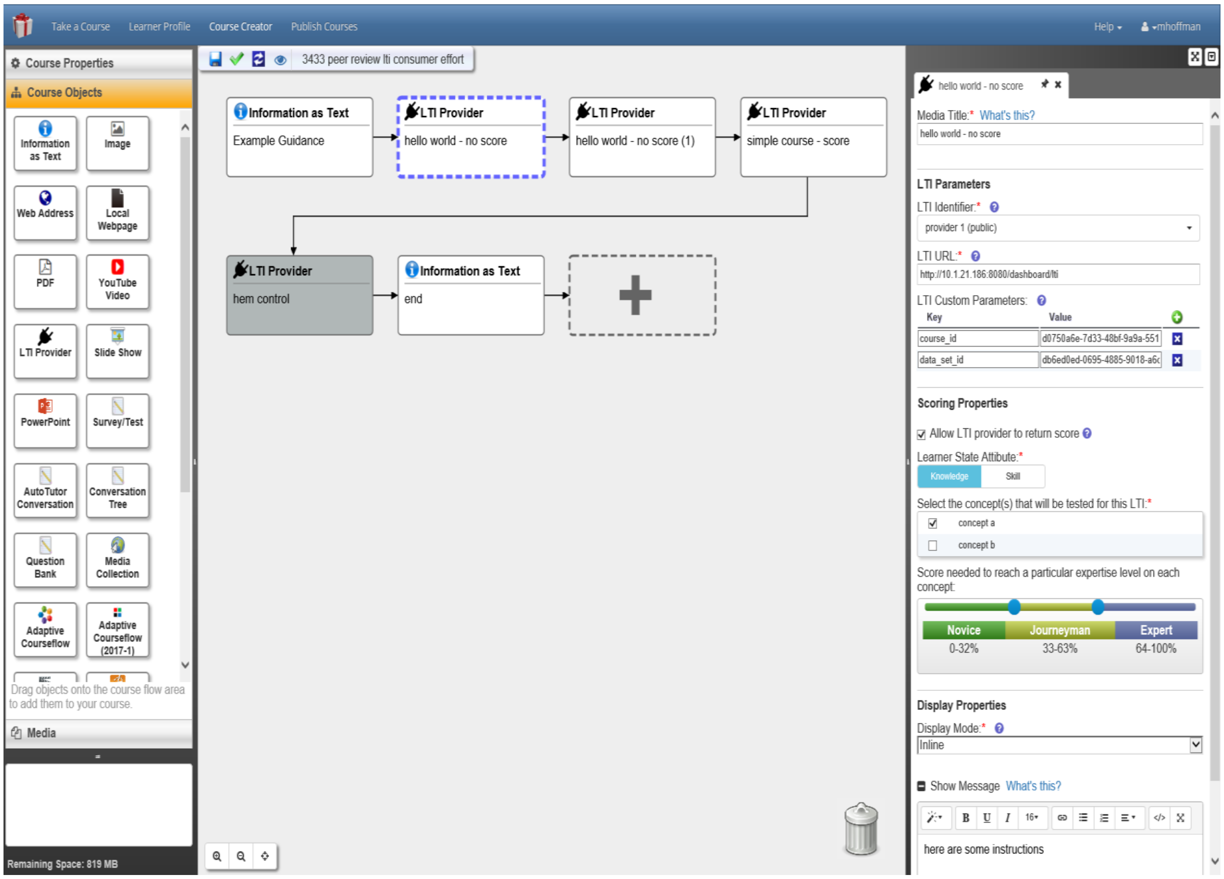
The three primary barriers to the adoption of ITSs are 1) their cost, 2) the specialized skills required to build them, and 3) their lack of standardization and thereby lack of reusability. All of these factors influence cost and therefore reduce the return-on-investment. Their cost is largely driven by the usability (or lack of usability) of authoring tools and the high degree of skill in instructional design, domain knowledge, and computer programming required to construct and ITS that can function without human intervention.

A major goal of the authoring tools for the Generalized Intelligent Framework for Tutoring (GIFT; Ososky, Sottilare, Brawner, Long, and Graesser, 2015) is to ease the development of ITSs in a variety of domains: cognitive, affective, psychomotor and social (Sottilare, Goldberg, Brawner, & Holden, 2012). With this goal in mind, the US Army Research Laboratory (ARL) focused their authoring research and development in 2017-2018 in four primary areas:

- Enhance the user interface to make it easier to develop ITSs
- Enhance user support for authoring tasks
- Expand authoring support for new capabilities
- Identifying opportunities for AIS standards

ENHANCE THE GIFT USER INTERFACE

In 2017, the cloud-based GIFT authoring tool was completely redesigned to allow users, without Instructional Systems Design (ISD) or computer programming expertise, to develop an intelligent tutor. As described in (Ososky, 2017), the redesigned authoring tools provide a graphical view of the adaptive course that is being created or modified. Figure 1 shows the current authoring tool which is made up of three frames. The one on the left shows the Course Properties (top), types of *Course Objects* that the author may choose (middle), and any media that the user has uploaded for use in the course (bottom). The center frame is the *Visual Flow Editor* and shows the flow of the course and allows the author to drag and drop various types of course objects onto the course flow diagram and then configure them to the specific domain of instruction. On the right side of the display is the *Editing Frame* that allows the user to edit/configure the selected course object. The focus of this paper is on new approaches to support GIFT authors to improve the user experience (UX), as well as authoring tool changes resulting from research conducted over the past year to add new capabilities to GIFT.



Media Title:* What's this?
hello world - no score

LTI Parameters

LTI Identifier:* provider 1 (public)

LTI URL:* http://10.1.21.186:8080/dashboard/lti

LTI Custom Parameters:

Key	Value
course_id	d0750a6e-7d33-48bf-9a9a-551
data_set_id	db6ed0ed-0695-4885-9018-a6c

Scoring Properties

☑ Allow LTI provider to return score

Learner State Attibute:* [✓ Knowledge] Skill

Select the concept(s) that will be tested for this LTI:*

☑ concept a
□ concept b

Score needed to reach a particular expertise level on each concept:

Display Properties

Display Mode:* Inline

Show Message What's this?

here are some instructions

Figure 15. GIFT Authoring Tool

ENHANCE GIFT USER SUPPORT

In January 2017, a team of researchers from ARL went to Fort Benning, Georgia to gather feedback on the newly re-designed GIFT Authoring Tool from potential military users. While the collected data showed that there was some improvement over previous versions of the authoring tool, there was still much work to be done. To provide additional support for the GIFT authors, ARL commissioned the development of a *GIFT Summer Camp* to provide detailed instruction for potential ITS authors, the development of several *GIFT instructional videos*, and the production of exemplar tutors to illustrate how GIFT tutors are constructed/configured.

GIFT Summer Camp

The ultimate goal for GIFT is that a domain Subject Matter Expert (SME) should be able to use the GIFT authoring tools to develop adaptive instruction (e.g., training course) without any additional help from instructional systems designers, computer programmers, etc. While we have come very close to reaching that goal, some users may still need some help in using the authoring tools. As a result, we developed a two day course, *GIFT Summer Camp* that teaches participants how to use the GIFT authoring tools to create an adaptive training course. Learning material was provided to the participants to use in making an adaptive training course on human anatomy using GIFT. Participant feedback was collected and may be used to provide a commercial offering of the course in the future.

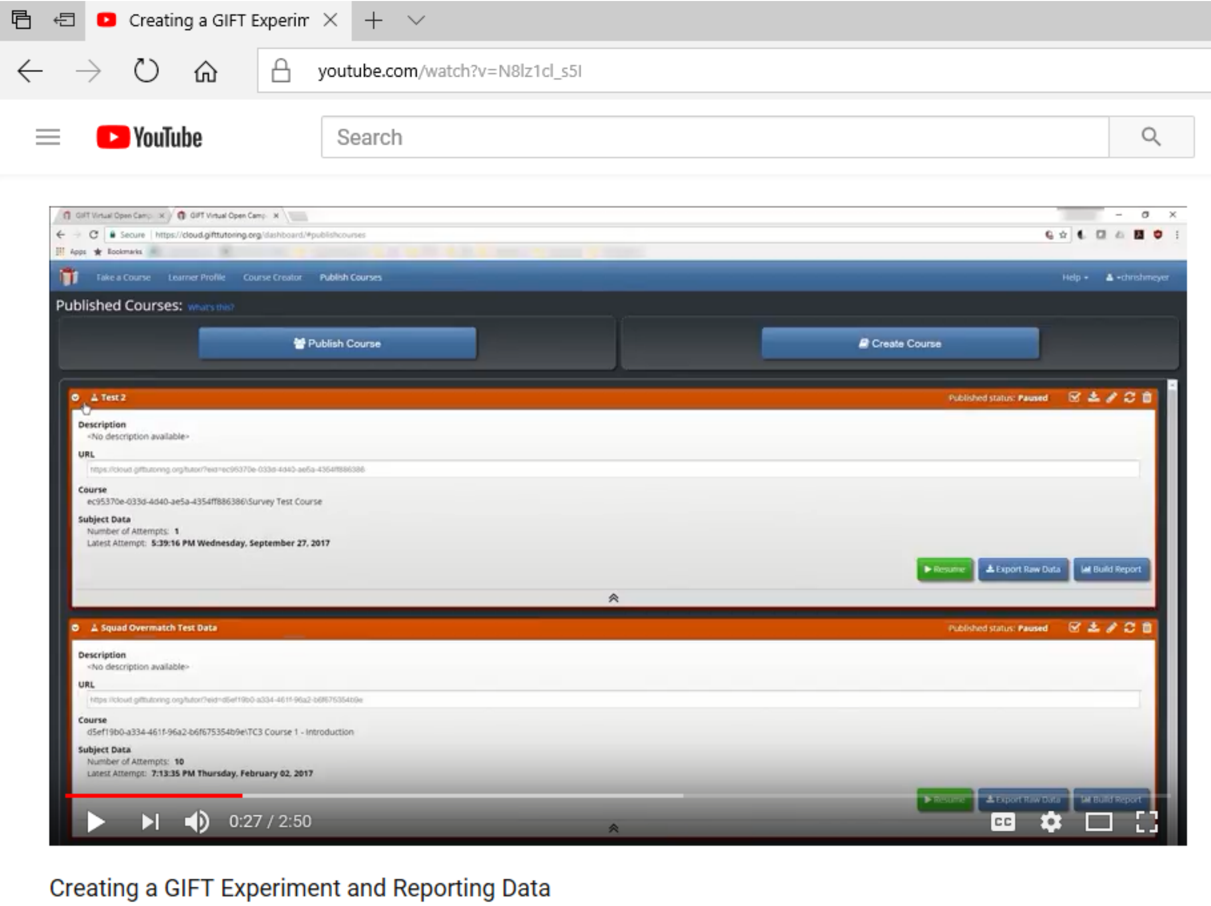
GIFT Instructional Videos

Since not everyone will be able to attend a GIFT Summer Camp session, we have also generated several educational youtube videos that demonstrate how to use the various features of the GIFT authoring tools (https://www.youtube.com/watch?v=nGywC-jf0Mk). Table 1 provides a current list of the videos. More instructional videos will be added in the future to guide GIFT users.

Table 1 – List of Available Gift Instructional Videos

About GIFT	Difference between types of surveys
GIFT Authoring Process	Import Tutor
Cloud vs. Downloadable GIFT	Copy Tutor
Adding a Survey	Metadata Tagging
Importing Media	Course Concepts
Where to find help	Linking to a simulation
Computer-based Training vs Intelligent Tutors	Question bank
GIFTSym and Community	Powerpoint vs Slideshow
Course Objects Overview	Case Study – Excavator Simulator
Export Tutor	Making an experiment

Since new functionality is continually being added to GIFT, the use of instructional videos is a quick and effective method to support GIFT authors in developing tutors. The videos are made in three phases: the script generation, the video demonstration, and the voice/video integration. The script is created and reviewed by the GIFT team to insure accuracy. We then have someone perform the specific function using GIFT while video capture software records the video from the computer screen. Once that is accomplished, the video is played back while the narration is captured. After final quality assurance and security review, the video is posted on youtube (Figure 2).



Published Courses: What's this?

Test 2

Description
<No description available>

URL
https://cloud.gifttutoring.org/tutor/?eid=ec95370e-033d-4d40-ae5a-4354ff886386

Course
ec95370e-033d-4d40-ae5a-4354ff886386\Survey Test Course

Subject Data
Number of Attempts: **1**
Latest Attempt: **5:39:16 PM Wednesday, September 27, 2017**

Squad Overmatch Test Data

Description
<No description available>

URL
https://cloud.gifttutoring.org/tutor/?eid=d5ef19b0-a334-461f-96a2-b6f675354b9e

Course
d5ef19b0-a334-461f-96a2-b6f675354b9e\TC3 Course 1 - Introduction

Subject Data
Number of Attempts: **10**
Latest Attempt: **7:13:35 PM Thursday, February 02, 2017**

Creating a GIFT Experiment and Reporting Data

Figure 2. GIFT Instructional Video on YouTube

Exemplar ITSs

To improve usability, we provided a set of public tutors that authors can use a models or exemplar tutors to guide development of their own tutors. While the exemplar tutors are not full ITSs, they do illustrate how authors are using GIFT to develop tutors in a variety of domains. Starting in FY18, ARL will be developing three additional tutors using the GIFT authoring tools to train knowledge and skills within three military contexts – land navigation, intelligence reporting, and visual signaling. The tutors will be authored to include multimedia components and will incorporate instructional system design principles for adaptive learning environments, to include passive, didactic training as well as interactive practice and rehearsal in prototypical scenarios. The GIFT-based tutors will incorporate external training applications like Unity and Virtual Battle Space (VBS3), and Leap Motion technology to incorporate real-time performance assessments to drive adaptive instruction.

Throughout the development of the tutors, GIFT will be continually assessed and usability issues will be identified. The tutor development process will also be thoroughly documented. The project will provide a systematic evaluation of GIFT usability, with recommendations for areas of improvement in user interface design and user experience. The research will result in high quality, sharable tutor exemplars highlighting the "art of the possible" in developing ITSs with GIFT. The documented process to developing tutors will benefit other authors in the GIFT community, demonstrating GIFT's utility for creating effective adaptive training.

AUTHORING SUPPORT FOR NEW CAPABILITIES

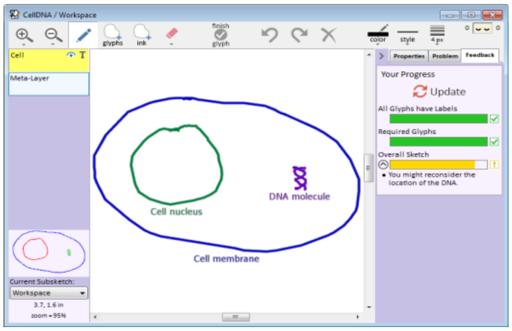
There has been a lot of new functionality added to GIFT over the past year. Most of these capabilities required changes to the GIFT authoring tools. The resulting new capabilities are discussed below.

Learning Tools Interoperability (LTI)

Last year, the LTI protocol (https://imsglobal.org/activity/learning-tools-interoperability) was partially implemented in GIFT to support ongoing research in the use of Massively Open Online Courses (MOOCs; Aleven et al., 2017). MOOCs are typically made up of recorded video lectures and outside learning activities. The main problem with MOOCs, however, is the very high drop-out rate compared to other on-line learning environments. The goal for this project is to provide additional support to the leaners through the use of ITSs, to include CTAT and GIFT, to support the Big Data in Education MOOC. The GIFT authoring tools were modified to support LTI version 1.0 as an LTI provider. As a result, GIFT could send/provide learner performance data from these activities to the LTI compliant Learning Management System (LMS) including edX, Canvas, and Blackboard. This year, the rest of the LTI has been implemented and now makes GIFT an LTI consumer. This new capability allows GIFT to receive data from other educational systems. For example, GIFT can now use LTI to receive learner performance data from a CTAT tutor. This new capability allows GIFT to control the outer loop of a course (e.g., macro-adaptation) while the CTAT tutor supports the inner loop (micro-adaptation.

Sketching Activities

Over the past year, the Army Research Laboratory (ARL) and Northwestern University have been exploring sketching technologies to support spatial learning, as part of on-going cooperative research in adaptive training technologies (Long, Forbus, Hinrichs, & Hill, 2018). Cogsketch and its associated Sketch Worksheets were designed to be general-purpose and use artificial intelligence to provide feedback to the learner performing sketching assignments (Figure 3). Cogsketch also has its own authoring environment for domain experts and instructors, to enable them to create new worksheets. The goal of this on-going cooperative research is to leverage Cogsketch to support the use of a sketching modality in GIFT as a new type of instructional media.



- You might reconsider the location of the DNA.

Figure 3. Cogsketch Authoring Tool

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

From the user feedback we received from GIFT users at Ft Benning, we realize that the GIFT authoring tools still have room for improvement. In the near term, we have provided additional support to authors in the form of instructional videos, on-site training classes (summer camp), and exemplar tutors. Future research should include a study to determine the effectiveness of the instructional videos, as these are fairly easy to make and provide just in time training for users that need support. In addition, we will continue to gather user feedback to improve usability of the GIFT authoring tools.

ACKNOWLEDGEMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Aleven, V., Sewell, J., Popescu, O., Sottilare, R., Long, R., & Baker, R. (2018, June, *in press*). Towards Adapting to Learners at Scale: Integrating MOOC and Intelligent Tutoring Frameworks. In Proceedings of the *Learning @ Scale Conference*, London, England, June 26-28, 2018. DOI: TBD

Long, R., Forbus, K., Hinrichs, T., & Hill, S. (2018, *in press*). Sketching as a Modality in Intelligent Tutoring Systems. In Proceedings of the 2018 Human Computer Interaction International (HCII) Conference, Las Vegas, NV.

Ososky, S., Sottilare, R., Brawner, K., Long, R., and Graesser, A. (2015). Authoring Tools and Methods for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline. US Army Research Laboratory (ARL-SR-0339), October 2015.

Ososky, S. (2017, May). The 2017 Overview of the GIFT Authoring Experience. In Proceedings of the Fifth Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium, Orlando, FL.

Sottilare, R. & Brawner, K. (2018, March). Exploring Standardization Opportunities by Examining Interaction between Common Adaptive Instructional System Components. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Sottilare, R.A. & Fletcher, J.D. (2018). Chapter 1 – Research Task Group 237 Work Plan. NATO Final Report of the Human Factors & Medicine Research Task Group (HFM-RTG-237), Assessment of Intelligent Tutoring System Technologies and Opportunities. NATO Science & Technology Organization. DOI: 10.14339/STO-TR-HFM-237. ISBN 978-92-837-2091-1.

Sottilare, R., Goldberg, B., Brawner, K., & Holden, H. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.

ABOUT THE AUTHORS

Mr. Rodney Long *is a Science and Technology Manager at ARL in Orlando, Florida and is currently conducting research in adaptive training technologies. Mr. Long has a wide range of simulation and training experience spanning over 30 years in the Department of Defense and has a Bachelor's Degree in Computer Engineering from the University of South Carolina and Master's degree in Industrial Engineering from the University of Central Florida.*

Dr. Robert Sottilare *leads adaptive training research within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).*

Incorporating psychomotor skills training into GIFT tutors: "outside the box" authoring support

Debby Brown[1], Benjamin Goldberg, Ph.D.[2], Benjamin Bell, Ph.D.[1], and Elaine Kelsey[1]
Eduworks Corporation[1], U.S. Army Research Laboratory[2]

INTRODUCTION

The Generalized Instructional Framework for Tutoring (GIFT) is enabling training developers to create diverse and effective Intelligent Tutoring Systems (ITS) in support of a broad array of U.S. Army training needs. GIFT-enabled technology initiatives are developing new tools and methods for streamlining ITS development along numerous fronts. However, a general category of performance that is under-represented in ITS is skills falling within the psychomotor domain. Paradoxically, psychomotor skills are foundational to many of the competencies that compose the U.S. Army's vision for 21st Century Soldier Competencies as expressed in the Army Learning Model (ALM).

Although there has been steady improvement in GIFT tools, libraries and methods, development of tutors for skills falling within the psychomotor domain remains a challenge that designers must address with little support from GIFT or its contemporary authoring tools. Despite these challenge, a few examples have illustrated the promise of using ITS for psychomotor skills training in domains including marksmanship and tactical combat casualty care. The success of such demonstrations though has relied on significant investments of time by highly specialized training development and technology experts. In order to scale the production of training that incorporates psychomotor skills, ITS frameworks such as GIFT must support an author not only in creating the conventional elements of an ITS but also in interpreting information arriving from external sensors in a way that productively advances learning objectives.

The Psychomotor Skills Training Agent-based Authoring Tool (PSTAAT) is an agent-assisted ITS authoring tool for the GIFT framework. In this paper we present our approach to helping an author link psychomotor measures from external sensors with performance thresholds and with corresponding instructional feedback. We discuss the use of guided examples and the agent's encapsulated knowledge of psychomotor ITS authoring. We also introduce a new machine learning-based approach that analyzes sensor data to recommend performance ranges. We conclude with an example authoring interaction.

PSYCHOMOTOR SKILLS: INSTRUCTIONAL CHALLENGES

Psychomotor skills have properties that are distinctive from skills training in other domains (cognitive and affective, Bloom, et al., 1956). Psychomotor skills involve movement and coordination but generally de-emphasize verbal processes. Tasks like fast-roping, applying a tourniquet, flying a CH-47, aiming a weapon, or traversing a chasm illustrate the prevalence and military relevance of psychomotor skills.

Psychomotor skills typically include physical movement, coordination, and use of gross, fine, or combined motor-skills. Learning these skills (like all learning) requires practice. Tutoring systems to train psychomotor skills would thus emphasize practice of some kind, opportunities for skill performance with coaching and feedback, and assessed skill demonstration. However, tutoring systems in this domain of learning must accommodate the distinctive metrics for assessing performance of psychomotor skills (e.g., speed, force, precision, distance, technique).

Another differentiating property of psychomotor skills is the process involved in mastery – i.e., the stages of skill acquisition. Adopting a process model is important to authoring because it helps structure the authoring dialogue. Our model draws from multiple researchers who generally follow Bloom's basic tenets. From theories advanced by Dave (1970), Simpson (1972), Harrow (1972), and Romiszowski (1999), we adopted a simplified synthesis of psychomotor skill learning suitable for organizing the PSTAAT authoring process (Brown, Bell & Goldberg, 2017), summarized in Table 1.

Table 1. Psychomotor skill model synthesized by comparing multiple research models.

Level	Definition	Example
Observing	Active mental attending of a physical event.	The learner watches a more experienced person. Other mental activity, such as reading may be a part of the observation process.
Imitating	Attempted copying of a physical behavior.	The first steps in learning a skill. The learner is observed and given direction and feedback on performance. Movement is not automatic or smooth.
Practicing	Trying a specific physical activity over and over.	The skill is repeated over and over. The entire sequence is performed repeatedly. Movement is moving towards becoming automatic and smooth.
Adapting	Fine tuning. Making minor adjustments in the physical activity in order to perfect it.	The skill is perfected. A mentor or a coach is often needed to provide an outside perspective on how to improve or adjust as needed for the situation.

The authoring interactions in PSTAAT thus support creating activities for a learner progressing through observing, imitating, practicing and adapting. This analysis established a foundation for developing an agent to support the authoring of simulation-based ITS focused on psychomotor skills as discussed next.

PSYCHOMOTOR SKILLS: AUTHORING CHALLENGES

While successful development efforts have demonstrated that ITS are an effective approach to training psychomotor skills, developing these systems remains a costly and time-consuming enterprise. ITS authoring tools are limited in scope, capability, and generalizability, so the time, expertise and resources needed to create ITS persist. In contrast to general-purpose authoring tools, however, tools that address the development of a specific kind of ITS can be more powerful because they embody (and help authors adhere to) a set of assumptions about what the authored product will look like and how it will behave. PSTAAT is representative of a more specific tool, supporting authoring with an agent that encapsulates knowledge useful in guiding the authoring process, to include pedagogical knowledge tailored to instruction in, and assessment of psychomotor skills.

Psychomotor skills can be distinguished from cognitive skills because they involve movement and coordination, typically composed of physical movement, coordination, and use of gross, fine, or combined motor-skills. Because psychomotor skills are not inherently suited to be trained in conventional computer-mediated learning environments, developing ITS that incorporate psychomotor skills training presents several distinctive challenges. Motor skill elements of a psychomotor skill must be practiced using a physical device, such as rudder pedals, a firearm, or a tourniquet. Physical devices that capture and digitize motions and actions have demonstrated the ability to replicate, to varying levels of fidelity, user effects in a simulated environment in domains including flying, driving, performing medical procedures, and firefighting.

Training, however (in contrast to simulation), requires the additional capability to interpret performance from the stream of digital data flowing from the physical device. The ITS author must therefore be able to construct ways for the tutor to make sense of the data captured by a sensor. This presents specific challenges to the author, who must: (1) identify which among the specific data points sampled by the physical device should be attended to as indicators of expert versus novice performance; (2) calibrate each data source, in order to associate numerical data with performance markers; (3) define assessment and feedback associated with specific performance tiers; and (4) accommodate variable performance thresholds in cases where context can alter assessment thresholds.

For training psychomotor skills, the primary factor for mastery is practice. Psychomotor skills tutoring should thus emphasize opportunities to practice physical skills with coaching, feedback, and assessment. The author must also consider the nature of performance metrics for psychomotor skills; measures such as speed, precision, distance, or technique might have to be monitored. The ITS author is thus faced with the complex task of correlating data from physical devices with multiple and composite performance metrics.

PSYCHOMOTOR SKILLS: ASSESSMENT CHALLENGES

To help an author apply the appropriate performance ranges associated with the use of a physical device, we introduce a new machine learning (ML) approach that analyzes and classifies sensor data. The ML algorithms automate the detection of sensor thresholds (e.g., detecting the difference between Expert and Novice performance) based on expert feedback. The ML algorithm processes raw sensor data using sensor-appropriate scripts and integration with appropriate machine learning libraries through a Spark instance. Leveraging RapidMiner integrations with GIFT, it is also possible to bypass or adjust automated sensor threshold detection through direct adjustment of ML models. The ML algorithm applies a range of possible models to the test data generated in the performance modeling phase (or provided directly by the author), and attempts to determine the ‘best-fit’ model for a given combination of sensors and a given performance metric outcome or expertise level.

The ML model uses the data imported from cases to learn one or more reward functions that characterize and explain expert behavior, using Inverse Reinforcement Learning (IRL); and to learn to distinguish expert behavior from novice behavior (i.e., clustering). Once the training data set has produced an ML model, we use it to auto-generate the logic model. The logic model then evaluates performance during task execution. If desired, the ML model can also be an additional source of feedback on how to improve performance (e.g., “reduce breathing rate during the latter half of task performance”). Figure **1** illustrates the steps in the creation and use of the ML model.

EXTENDING GIFT WITH PSYCHOMOTOR AUTHORING

PSTAAT is designed to work within the Army Research Laboratory (ARL) Generalized Instructional Framework for Tutoring (GIFT) (e.g., Sottilare, 2012; Sottilare, Goldberg, Brawner & Holden, 2012). PSTAAT is thus an extension to GIFT that supports the authoring of psychomotor skills specifically, and that leaves to the broader GIFT environment support for authoring skills in the cognitive domain.

PSTAAT uses an exemplar ITS to provide relevant illustrations for authoring and to inform the design of the authoring tool itself. This exemplar, the Adaptive Marksmanship Trainer (AMT), was created in the GIFT to enhance an existing Engagement Skills Trainer (EST) that uses instrumented emulators of several types of firearms. AMT enhances this system by incorporating adaptive tutoring and automated performance measures (Goldberg, Amburn, Brawner & Westphal, 2014).

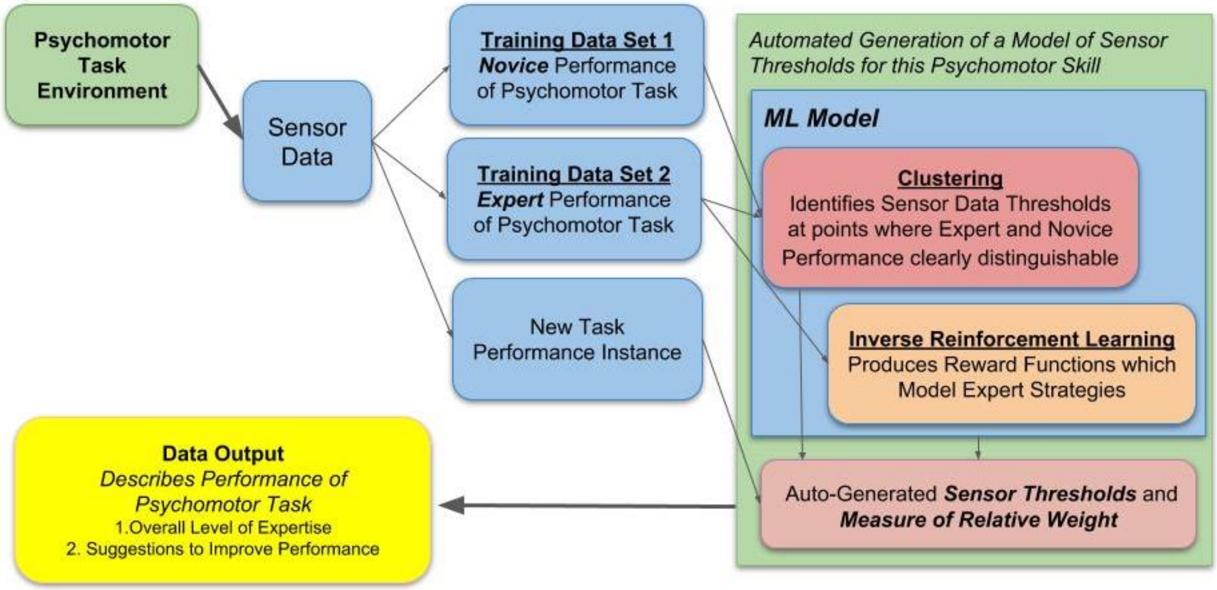

Psychomotor Task Environment
Sensor Data
Training Data Set 1 Novice Performance of Psychomotor Task
Training Data Set 2 Expert Performance of Psychomotor Task
New Task Performance Instance
Automated Generation of a Model of Sensor Thresholds for this Psychomotor Skill
ML Model
Clustering Identifies Sensor Data Thresholds at points where Expert and Novice Performance clearly distinguishable
Inverse Reinforcement Learning Produces Reward Functions which Model Expert Strategies
Auto-Generated Sensor Thresholds and Measure of Relative Weight
Data Output Describes Performance of Psychomotor Task
1.Overall Level of Expertise
2. Suggestions to Improve Performance

Figure 1. Steps in creation and application of machine learning model.

An initial step of the authoring dialogue is to instantiate the instructional model (recall Table 1). The author may incorporate some or all of the phases of the model for the psychomotor instruction being developed. Within each phase the author specifies the psychomotor activities to be performed by linking to a corresponding training application (e.g., a Unity application incorporating a backhoe emulator).

To help the author conceptualize the mapping from device outputs (e.g., a trigger squeeze, an aim point) to performance assessment for any given activity (external simulation), we adapt from AMT a layered mapping to associate sensor outputs with skill metrics, mediated by a middle layer that encapsulates the mechanisms for analyzing input data to determine a performance threshold. Figure 2 shows the layers using the exemplar ITS sensors and skills. This abstraction helps an author focus on mapping sensor data (top layer) to skill performance (bottom layer). The processing of those inputs (done by performance profiles, middle layer) is defined by the author and guided by PSTAAT to create adaptive, contextual feedback specific to the learner’s detected performance (currently, above, at, or below expectation).

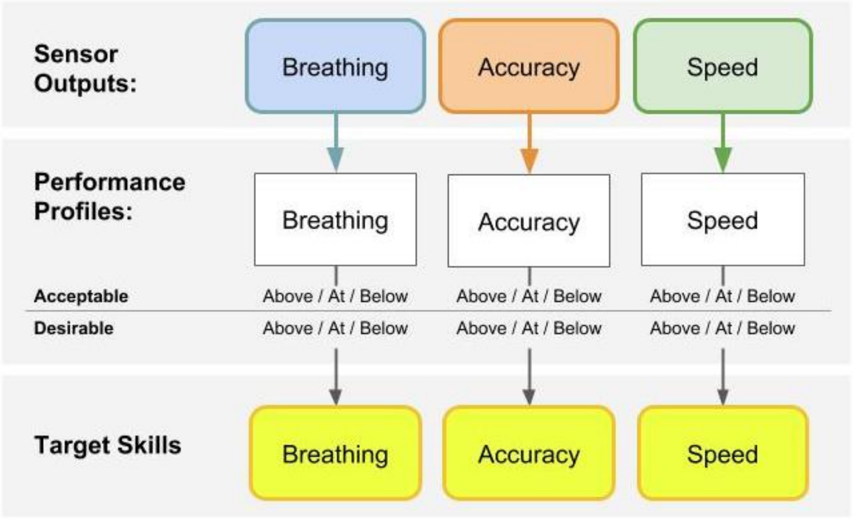


Figure 2. Example of layered mapping of sensors to skills, mediated by performance profiles.

GIFT IMPLEMENTATION

To provide psychomotor domain-specific authoring support, PSTAAT introduces a Psychomotor Activity Course Object to the GIFT Course Creator. A course object is an element that can be selected from a panel of supported types and added via a drag-and-drop authoring interface to a course flow sequence being created in the Course Creator. Each type of course object represents a different method of presentation and/or interaction with the learner and can be combined in any order in a course sequence. The PSTAAT extension, called the Psychomotor Activity Course Object, is depicted in Figure 3.

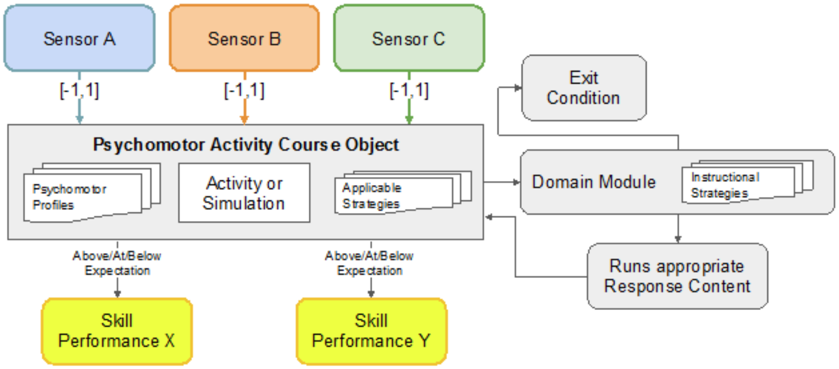


Figure 3. Schematic diagram of PSTAAT course object for mapping sensors to target skills

When a Psychomotor Course Object is added to the Course Creator, PSTAAT auto-generates a GIFT-compliant template, organized by the phases of the psychomotor domain (Observe, Imitate, Practice, Adapt). For each phase, the author selects a performance profile (that related sensor outputs to skill performance). At this point, the author can choose an existing profile, modifying it if desired, or create a new profile, a process discussed later. Once each phase is configured with a selected psychomotor profile, the PSTAAT agent auto-populates the psychomotor activity with placeholder learner states and guides the author through development of instructional strategies to complete the tutor.

To define a psychomotor activity, an author selects configured sensors as inputs and defines adaptive content delivery for the configured target skills. This adaptive behavior is defined by associating tailored feedback with corresponding performance levels calculated by a Psychomotor Profile (Figure 4). A Psychomotor Profile processes data from active sensor feeds to derive measures of performance (using an above/at/below expectation scale). The algorithms driving this assessment are informed by cases – previously generated data captured from subjects performing a task and tagged with performance outcomes. During data capture, data is tagged in one of two ways; either by an objective measure of task performance (e.g., a score generated automatically by the task environment) or by a subjective, human labeling of task performance (e.g., an expert observer determining that a given instance of the task performance was “above” expectation).

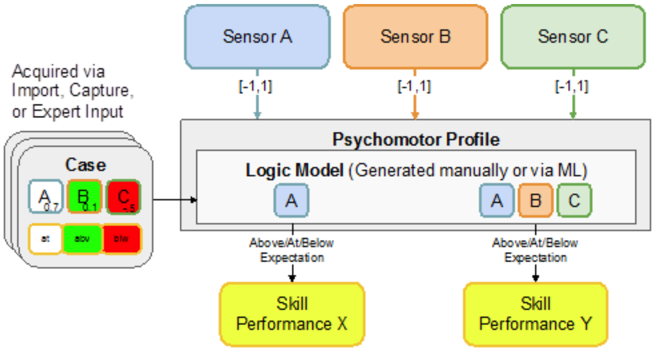
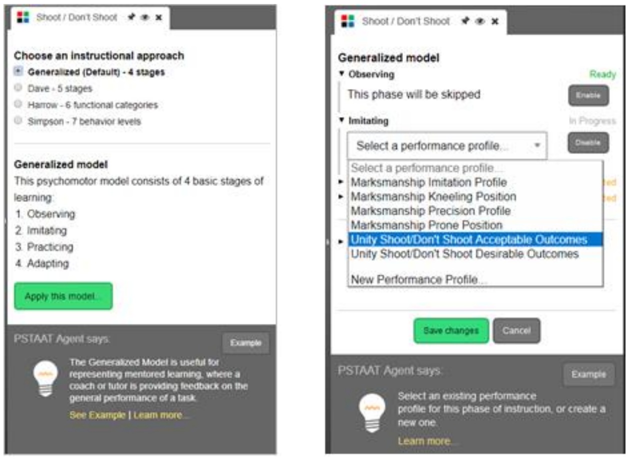

Sensor A
Sensor B
Sensor C
[-1,1]
[-1,1]
[-1,1]
Acquired via Import, Capture, or Expert Input
Case
A 0.7
B 0.1
C -.5
at
abv
blw
Psychomotor Profile
Logic Model (Generated manually or via ML)
A
A
B
C
Above/At/Below Expectation
Above/At/Below Expectation
Skill Performance X
Skill Performance Y

Figure 4. Schematic showing detail of the Psychomotor Profile.

PSTAAT thus manages the authoring dialogue in three segments: skills profiling, sensor mapping, and course object definition (i.e., activities, sequencing). The PSTAAT authoring agent provides contextual authoring support for each of these general-purpose task areas, and recommends the use of psychomotor domain instructional approaches and adaptive feedback strategies in the form of templates and examples.

EXAMPLE INTERACTION

A brief example illustrates an authoring interaction. For brevity we omit preceding steps typical in the GIFT Course Creator unrelated to PSTAAT. The author first chooses a preferred instructional model, skips the Observing phase, and selects an existing performance profile for the Imitating phase (Figure 5).



Shoot / Don't Shoot

Choose an instructional approach

● **Generalized (Default) - 4 stages**
○ Dave - 5 stages
○ Harrow - 6 functional categories
○ Simpson - 7 behavior levels

Generalized model

This psychomotor model consists of 4 basic stages of learning:

1. Observing
2. Imitating
3. Practicing
4. Adapting

PSTAAT Agent says:

The Generalized Model is useful for representing mentored learning, where a coach or tutor is providing feedback on the general performance of a task.

See Example | Learn more...

Shoot / Don't Shoot

Generalized model

▾ **Observing** Ready

This phase will be skipped

▾ **Imitating** In Progress

Select a performance profile...

Select a performance profile...
Marksmanship Imitation Profile
Marksmanship Kneeling Position
Marksmanship Precision Profile
Marksmanship Prone Position
Unity Shoot/Don't Shoot Acceptable Outcomes
Unity Shoot/Don't Shoot Desirable Outcomes
New Performance Profile...

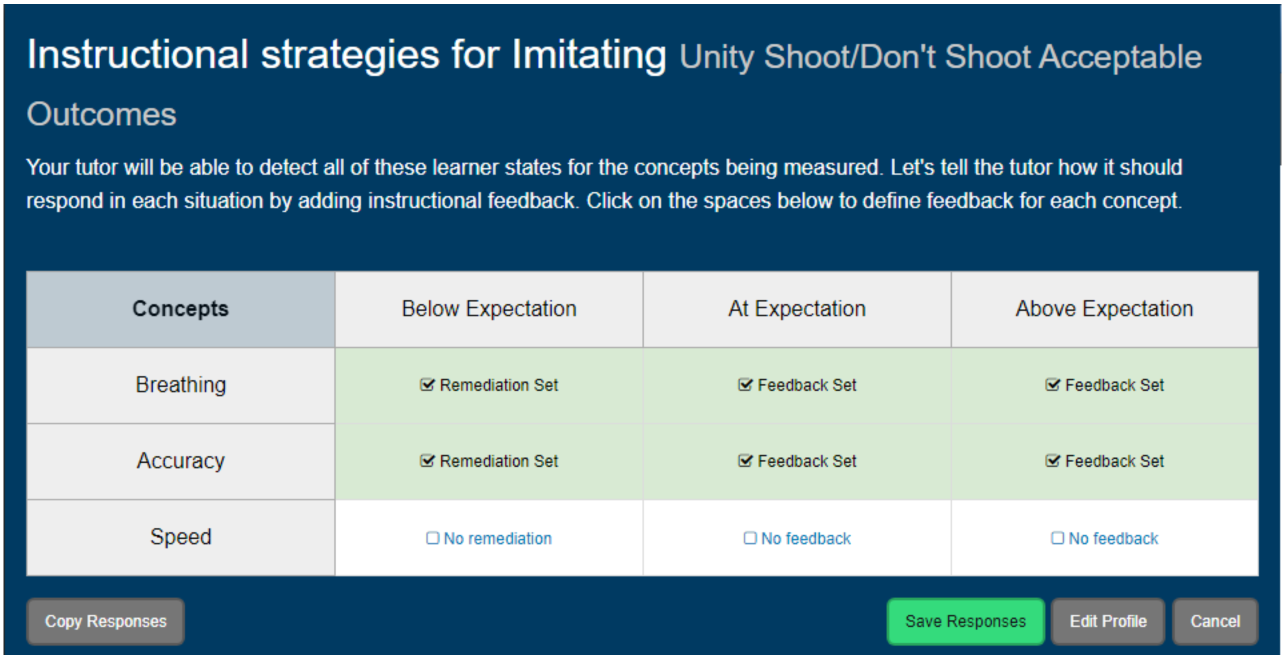
PSTAAT Agent says:

Select an existing performance profile for this phase of instruction, or create a new one.

Learn more...

Figure 5. Selecting an instructional model and assigning performance profiles to each phase.

The author then adds instructional feedback and remediation for the selected performance profile (Figure 6). At any time, the author may edit the threshold values internal to a performance profile. Figure 7 illustrates threshold values for different performance tiers for a given device displayed. From this editor, the author can modify this configuration, add sensors, and link this profile with an external application.



Instructional strategies for Imitating Unity Shoot/Don't Shoot Acceptable Outcomes

Your tutor will be able to detect all of these learner states for the concepts being measured. Let's tell the tutor how it should respond in each situation by adding instructional feedback. Click on the spaces below to define feedback for each concept.

Concepts	Below Expectation	At Expectation	Above Expectation
Breathing	☑ Remediation Set	☑ Feedback Set	☑ Feedback Set
Accuracy	☑ Remediation Set	☑ Feedback Set	☑ Feedback Set
Speed	□ No remediation	□ No feedback	□ No feedback

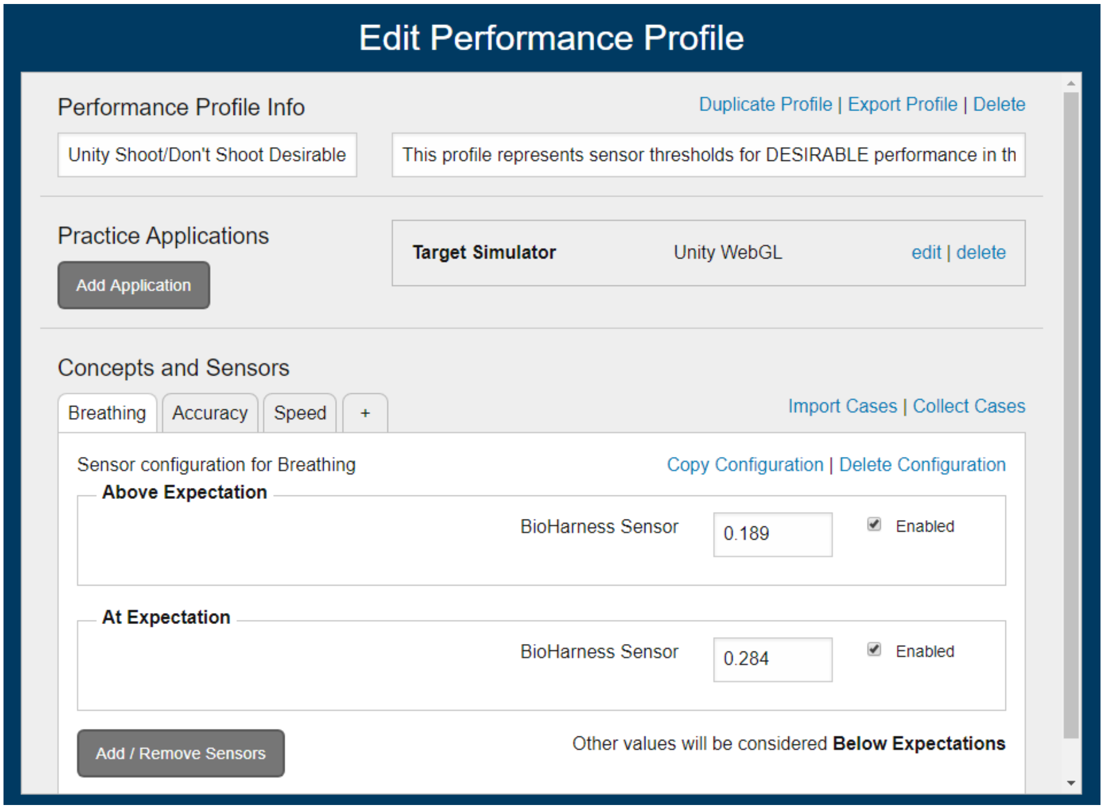
Figure 6. Assigning feedback and remediation for a selected performance profile.

CONCLUSIONS AND FUTURE RESEARCH

Streamlining ITS authoring remains an elusive goal, but steady progress in tools and frameworks such as GIFT are bridging this gap. For ITS that train psychomotor skills, authors face additional challenges. To support the integration of external training simulations and corresponding physical devices with a tutoring system, PSTAAT demonstrates an agent-driven system that employs templates, editors, and sensor data processing via machine learning-derived assessments. When fully integrated, PSTAAT will expand the reach of ITS authors by enabling them to incorporate psychomotor skills training along with cognitive skills training, cultivating a richer diversity of training applications emerging from the GIFT community.

PSTAAT demonstrates an integrated approach to GIFT ITS authoring that uses performance support and agent techniques to provide informative feedback and guidance to the author during the ITS development process. We discuss how psychomotor task performance models and sensor configurations can be abstracted into reusable psychomotor profiles that both simplify and streamline the design of psychomotor activities within GIFT.

The process to develop ITS thus remains time-consuming and costly. For the Army to successfully realize the ALM vision, creating ITS that target psychomotor skills must be an affordable, replicable, and reusable process.



Edit Performance Profile

Performance Profile Info

Duplicate Profile | Export Profile | Delete

Unity Shoot/Don't Shoot Desirable

This profile represents sensor thresholds for DESIRABLE performance in th

Practice Applications

Add Application

Target Simulator Unity WebGL edit | delete

Concepts and Sensors

Breathing | Accuracy | Speed | +

Import Cases | Collect Cases

Sensor configuration for Breathing

Copy Configuration | Delete Configuration

Above Expectation

BioHarness Sensor 0.189 ☑ Enabled

At Expectation

BioHarness Sensor 0.284 ☑ Enabled

Add / Remove Sensors

Other values will be considered **Below Expectations**

Figure 7. Performance Profile editor for viewing/modifying performance thresholds and adding sensors.

REFERENCES

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay.

Brown, D., Bell, B. and Goldberg, B., 2017. Authoring Adaptive Tutors for Simulations in Psychomotor Skills Domains. In *Proceedings of MODSIM World 2017*, Virginia Beach, VA: NTSA.

Dave, R.H. (1970). Psychomotor levels. In R.J. Armstrong (Ed.), *Developing and Writing Behavioral Objectives*. Tucson, Arizona: Educational Innovators Press.

Goldberg, B., Amburn, C., Brawner, K., & Westphal, M. (2014). Developing models of expert performance for support in an adaptive marksmanship trainer. In *Proceedings of I/ITSEC*.

Harrow, A. (1972) A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives. New York: David McKay.

Romiszowski, A. (1999). The development of physical skills: Instruction in the psychomotor domain. In Instructional-design theories and models: a new paradigm of instructional theory (Vol. 2). Mahwah, NJ: Erlbaum.

Simpson E.J. (1972). The Classification of Educational Objectives in the Psychomotor Domain. Washington, DC: Gryphon House.

Sottilare R. (2012). Considerations in the development of an ontology for a Generalized Intelligent Framework for Tutoring. International Defense and Homeland Security Simulation Workshop, in *Proc. of the I3M Conference*. Vienna, Austria, September 2012.

Sottilare, R.A., Goldberg, B.S., Brawner, K.W., & Holden, H.K. (2012). A Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems (CBTS). In *Proc. of Interservice/Industry Training, Simulation, and Education Conference* (I/ITSEC), Dec, 2012.

ABOUT THE AUTHORS

Debbie Brown *is a software engineer and senior learning technologist at Eduworks Corporation focusing on web application implementations that use AI and machine learning to semi-automate user-centered workflows and enable advanced adaptive eLearning tools and experiences. She has been conducting eLearning R&D projects for 20 years, and operating as a software engineer in government, academic, and workforce settings for 30 years. She holds an MS in Instructional Systems and a BS in Computer Engineering from Mississippi State University.*

Benjamin Goldberg *is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL) Human Research and Engineering Directorate (HRED) in Orlando, FL. He has been conducting research in intelligent tutoring for the past eight years with a focus on adaptive learning in simulation-based environments and how to leverage Artificial Intelligence tools and methods to create personalized learning experiences. Dr. Goldberg holds a Ph.D. from the University of Central Florida in Modeling & Simulation.*

Benjamin Bell *is the president of Eduworks Government Solutions and an expert in the application of artificial intelligence to decision support, simulation, training, and human-machine interaction. He has practiced in the field of AI for over twenty years, leading funded research and development in applied settings, largely for military applications. He holds a Ph.D. from Northwestern University and is a graduate of the University of Pennsylvania.*

Elaine Kelsey *is a research engineer and computational linguist with Eduworks Corporation, where she focuses on developing natural language processing and machine learning solutions for a range of government and commercial applications in intelligent tutoring. She most recently led development of Eduworks' automated question generation, assessment, and competency alignment algorithms. She holds multiple bachelor's and master's degrees in computer science, linguistics, molecular biology, and biostatistics, and is currently working towards a combined MS / PhD in machine learning and computational linguistics.*

Toward Automated Scenario Generation with Deep Reinforcement Learning in GIFT

Jonathan Rowe[1], Andy Smith[1], Robert Pokorny[2], Bradford Mott[2], and James Lester[2]
North Carolina State University[1], Intelligent Automation, Inc.[2]

INTRODUCTION

Simulation-based training is an important tool for preparing learners to perform a broad range of complex tasks and skills. A key functionality of simulation-based training environments is delivering scenarios that drive learning interactions that approximate real-world situations. However, simulation-based training scenarios are typically resource-intensive to create. Some simulation environments provide authoring tools that enable new training scenarios to be manually created by subject matter experts, but these tools often require a high degree of specialized knowledge to be utilized effectively. Authored scenarios often cannot be reused in other training environments, and the knowledge associated with particular authoring tools has limited transferability. Further, learners are usually limited to training with a finite set of training scenarios provided by the system's designers. If learners have mastered the learning objectives associated with the available set of training scenarios, there is marginal benefit provided by further training with the simulation. Finally, training simulation scenarios are often delivered following a one-size-fits-all approach: they have limited capacity to dynamically respond to the broad range of individual differences in knowledge or behavior that are typical among learners.

Automated scenario generation offers considerable promise for addressing the needs of simulation-based training. By utilizing automated scenario generation techniques, simulation environments can account for individual differences in how learners respond to different types of scenario events. Further, they can create effective variations on training scenarios without requiring every scenario to be manually authored or managed by human experts. By leveraging generative techniques from interactive narrative technologies, we can dynamically create training scenarios that are configured to address instructors' learning objectives and tailored to the cognitive and behavioral characteristics of individual students (Riedl & Bulitko, 2012; Wang et al., 2017).

Recent advances in machine learning, including artificial neural networks (in general) and deep learning (in particular), have spurred growing interest in data-driven approaches to interactive narrative generation. For example, deep reinforcement learning (deep RL) has begun to show significant promise for personalizing events in narrative-centered learning environments (Wang et al., 2017). However, there are many open questions regarding how we can most effectively leverage machine learning in order to automatically generate training scenarios that are tailored to instructors' and trainees' learning objectives. To begin to address these questions, we are launching a new collaborative effort between North Carolina State

University, Intelligent Automation, Inc., and the U.S. Army Research Laboratory to investigate the design and development of a deep RL framework for automated scenario generation in GIFT. To serve as a testbed environment, we are generating training scenarios for Virtual Battlespace 3 (VBS3), a widely used simulation platform for small unit training within the Army, with an initial focus on Call for Fire (CFF) training.

In this paper, we provide an overview of the deep RL framework for automatic scenario generation. We describe how to formalize automatic scenario generation as a deep RL task. We discuss several key components of the framework, including the scenario adaptation library, simulated learners, and a deep neural network model of multi-objective rewards. We describe the VBS3 training simulation that we are utilizing as an initial testbed environment. Next, we describe preliminary findings from a proof-of-concept implementation of a reinforcement learning-based scenario generator that centers on generating initial scenario conditions for CFF training using multi-armed bandits (i.e., a stochastic scheduling technique that is related to deep RL).

AUTOMATED SCENARIO GENERATION WITH DEEP REINFORCEMENT LEARNING

We approach the task of automated scenario generation from the perspective of interactive narrative technologies. Automated scenario generation and interactive narrative generation share several key characteristics. First, in both automated scenario generation and interactive narrative generation, users are active participants in virtual worlds that dynamically respond to users' actions. Second, both automated scenario generation and interactive narrative generation center on generating sequences of events that achieve author-specified objectives to produce scenarios that are effective and engaging. Third, both automated scenario generation and interactive narrative generation produce scenarios that are realized in immersive simulation environments.

We formulate automated scenario generation as an instance of data-driven interactive narrative generation using deep RL (Wang et al., 2017). Deep RL is a computational framework that integrates two complementary families of machine learning techniques: reinforcement learning methods for training models for sequential decision-making under uncertainty, and deep neural networks for pattern recognition and representation learning with big data. Reinforcement learning is the task of a software agent inducing a control policy for selecting actions in an uncertain environment with delayed rewards (Sutton & Barto, 1998). Deep neural networks combine weighted summations of non-linear functions to extract and model multi-layer hierarchical representations of data using supervised, semi-supervised, and unsupervised machine learning techniques (Goodfellow, Bengio, & Courville, 2016). By integrating reinforcement learning and deep neural networks, deep RL provides a formalized framework for sequential decision-making in complex environments.

Deep reinforcement learning provides a natural computational framework for formalizing dynamic scenario generation: the generator is tasked with making a series of decisions about how specific scenario events should unfold at runtime to optimize student performance on a pre-specified set of learning objectives. Dynamic scenario generation can be modeled as a sequential decision-making task in which a scenario generator introduces successive adaptations to scenario events over discrete time steps. A time step represents the time point when an adaptable event, such as the introduction of an obstacle or elimination of a resource, is triggered in the scenario. Using this formalization, deep RL can be utilized to dynamically generate adaptive “child” training scenarios from a canonical (i.e., “parent”) scenario that explicitly optimizes for both author-specified objectives and trainee learning outcomes. By inducing multi-objective reward models for controlling run-time decisions about training scenario events, we intend to enable authors to specify learning objectives that generate personalized training scenarios in immersive simulation environments integrated with GIFT.

The deep RL framework for automated scenario generation consists of several key components: (1) a deep Q-Network model for controlling run-time scenario adaptation decisions that optimize multiple scenario objectives, (2) a scenario adaptation library that specifies possible transformations of "parent" scenarios to generate "child" scenarios, and (3) a simulated learner framework for generating synthetic data to train an initial version of the scenario generator. In addition, the framework requires a software infrastructure for integrating automated scenario generation functionalities with GIFT's modular software architecture.

Deep reinforcement learning leverages a Q-Network, a type of deep neural network, to model the estimated Q values of state-action pairs gathered from past observations of student interactions with a scenario during reinforcement learning. Q-networks encode the expected benefits of specific scenario adaptations in terms of a "reward function," an explicit mathematical expression of optimization criteria that guide automated scenario generation. In the original work on deep reinforcement learning for Atari game playing, the Q-network was organized as a convolutional neural network, which is a natural choice for processing image data from 2D games. For automated scenario generation, we will investigate deep architectures that utilize ***long-short term memory networks*** (LSTMs), a type of recurrent neural network, for modeling sequential data as typically expected in simulation scenarios. LSTMs are specifically designed for processing sequences of temporal data. LSTMs have achieved high predictive performance in many sequence labeling tasks, often outperforming standard recurrent neural networks by using a longer-term memory than standard RNNs, preserving short-term lag capabilities, and effectively addressing the vanishing gradient problem. We anticipate that utilizing LSTMs will enable reinforcement learning-based scenario generators to extract complex nonlinear interaction patterns between observed events and scenario adaptation decisions. LSTMs will be utilized to implement multi-objective deep Q-networks for automated scenario generation, as well as to implement machine learning-based simulated students to generate synthetic training data in future work.

Multi-objective reward functions will enable the automated scenario generator to consider tradeoffs between competing authorial goals, learning objectives, and learner engagement. This builds upon prior research by the NCSU team on multi-objective RL for interactive narrative generation (Sawyer, Rowe, & Lester, 2017), and it involves incorporating a vector-based representation of reward in the output layer of a deep Q-network, where vector indices correspond to different reward sources. A multi-objective Q-Network is induced at training time, and it yields a run-time scenario adaptation model after the course author has specified relative preferences among competing reward sources at course creation-time.

State representations for driving deep RL-based scenario generation decisions will consist of several complementary sources of information. First, state vectors will include domain-independent features encoding learner knowledge, traits, and performance characteristics. A key requirement for automated scenario generation is devising generalized assessment rules that can be applied to a broad range of generated scenarios within a given task domain; it would be prohibitive for a system designer to devise custom assessment logic for every automatically generated scenario. Second, state vectors will include several features that summarize the history of past scenario adaptation decisions performed by the scenario generator. Third, state vectors will include a one-hot encoding of the category of scenario adaptation decision under consideration in order to leverage modularity and maintain tractability of the reinforcement learning task. These state features are consistent with the Adaptive Tutoring Learning Effect Chain, and they are consistent with our project's vision for investigating how scenario generation functionalities should most effectively be integrated with the Pedagogical and Domain Modules of GIFT.

Scenario Adaptation Library

A key component of the deep RL-based scenario generation framework is devising a scenario adaptation library, which enumerates the range of possible transformations to a "parent" scenario that can be applied to generate "child" scenarios. By investigating different combinations of prospective scenario adaptations, the deep RL framework can generate a vast range of possible training scenarios that can be deployed with simulated or human learners, evaluated for their effectiveness in terms of trainee learning outcomes, and utilized to refine the scenario adaptation model for adaptive personalized scenario generation.

Integrating deep RL-based scenario generation into GIFT is a key aspect of the project. A key interest is exploring potential extensions to GIFT that support domain-independent specifications of scenario adaptations—these would be specified with GIFT's Pedagogical Module—in line with project objectives of generalizability of deep RL-based scenario generation. We envision a generalized taxonomy of scenario adaptations that includes several hierarchical domain-independent categories, including (1) inserting or removing obstacles; (2) constraining or increasing resources; (3) reconfiguring key objects; (4) adding, modifying, or removing sub-tasks; and (5) providing or removing embedded scaffolding. These categories will characterize a range of candidate adaptations that can be applied to a parent scenario in order to generate a set of "child" scenarios. These domain-independent scenario adaptations could be instantiated within GIFT's Domain Module, which will configure and launch scenario events at runtime via a GIFT gateway to be realized in the simulation-based virtual training environment.

Simulated Learners

In order to train deep reinforcement learning-based models of dynamic scenario generation, we will utilize synthetic training data produced by simulated students created for each of the virtual training environments. The design of simulated students is informed by related work in artificial intelligence in education (McCalla & Champaign, 2013) and spoken dialog systems (Schatzmann, Weilhammer, Stuttle, & Young, 2006). We investigate how simulation parameters related to model granularity and model complexity influence synthetic data generation for deep reinforcement learning-based scenario generation (Rowe et al., 2017).

Dynamic Scenario Generation User Experience

The user experience of automated adaptive scenario generation functionalities in GIFT is likely to be different based on whether the user is a course designer, a student, or a software developer. For a pre-integrated training environment, a course designer will select the training objectives that he is targeting in the GIFT Course Creator, and he can specify constraints on specific scenario adaptations that he would like to avoid in the generated run-time scenario. As long as the deep RL-trained scenario generator can produce a scenario that is consistent with the objectives and constraints provided by the author, the course will validate and it can be tested with live students. For a student, automated scenario generation will be invisible, and training events will be tailored based on the student's individual traits, knowledge, and performance in the simulation environment.

For a software developer seeking to integrate a new domain or training application, she will need to (1) have a deep knowledge of the "parent" scenarios supported by the training environment, (2) create a specification of possible "child" scenario adaptations that can be realized in the training environment, (3) develop a gateway module that mediates communication between scenario generation functionalities in GIFT and the training application, (4) have access to training data for inducing deep reinforcement learning-based scenario generation models if existing domain-independent models cannot be reused, and (5) integrate trained scenario generation models into run-time GIFT courses. Given these resources, a software developer will be able to use deep reinforcement learning-based scenario generation functionalities to create a new scenario generator for a novel domain or simulation environment.

GIFT Integration

Integrating deep RL-based scenario generation into GIFT is a key objective of the project, and supporting automated scenario generation has several implications for the GIFT architecture, authoring tools, and software. For example, supporting automated scenario generation in GIFT will likely involve extensions to the GIFT Course Creator to enable instructors to identify relevant learning objectives that should guide the generation of relevant training scenarios. Further, devising tools for ranking and visualizing automatically generated training scenarios will be essential for instructors to configure scenario generation functionalities for use in training courses that they create. Devising generalized assessment logic that can operate across multiple scenarios, and be specified in GIFT DKF files, will be critical for ensuring that course creators do not need to hand-specify custom assessment rules for every generated scenario. Finally, the project seeks to investigate support for domain-independent specifications of scenario adaptations—these would be specified by GIFT's Pedagogical Module—in line with project objectives of generalizability of deep RL-based scenario generation. This generalized taxonomy of scenario adaptations will include hierarchical domain-independent categories, such as (1) inserting or removing obstacles; (2) constraining or increasing resources; (3) reconfiguring key objects; (4) adding, modifying, or removing sub-tasks; and (5) providing or removing embedded scaffolding. These categories characterize a scenario adaptation library that defines the space of possible scenarios in a manner that holds potential for portability and reuse.

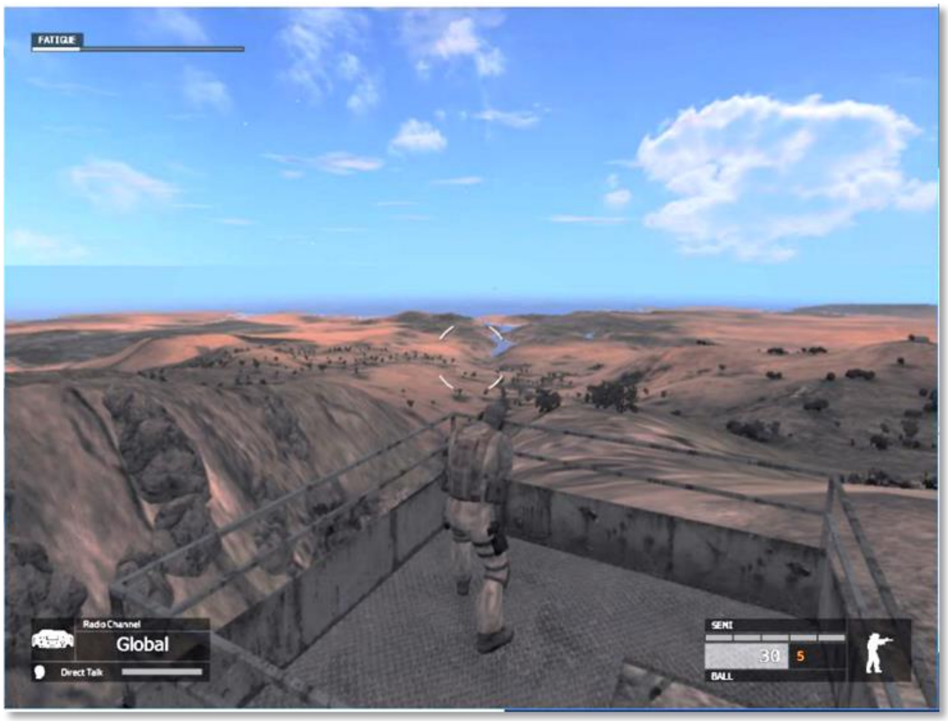


Figure 1. Screenshot of Virtual Battlespace 3 simulation environment.

VIRTUAL BATTLESPACE 3 TESTBED ENVIRONMENT

The selection of testbed simulation-based virtual training environments is guided by two key requirements. First, the simulation environment should either be open-source, or include APIs or tools for generating novel training scenarios, as well as models for specifying adaptations to scenarios at run-time. By enabling close integration between GIFT and the simulation-based training environment, it is possible to engage in rapid iteration cycles for designing, developing, and testing directions for dynamic scenario generation. Second, the selection of the simulation-based virtual training environment testbeds should prioritize environments that support scenario generation that enable the scenario generator to produce thousands (or more) of "child" scenarios from a single "parent" scenario. To fully exercise the deep RL framework's generative capabilities (i.e., its ability to broadly explore a given scenario space) and fully stress test its computational capabilities, testbeds should include a broad range of event types, actors, and trainee interactions.

The primary testbed simulation environment for the first phase of this project will be Virtual Battlespace 3 (VBS3). Built by Bohemia Interactive Simulations, VBS3 is the Army's most widely used simulation platform for small unit training (Figure 1). Designed as a flexible simulation tool for tactics training and mission rehearsal, VBS3 provides realistic physics, high-fidelity 3D graphics, expansive geo-specific terrains, and a large content library of 3D digital assets. VBS3 can be used for a broad range of training purposes, including training for cordon and search of specific structures, breaching obstacles, defense of territory with machine gun and mortars, and clearing highways of IEDs. VBS3 also includes features that enable dynamic modifications to training scenarios, as well as features for observation of the environment by instructors, and an After Action Review playback capability.

Although it is a closed-source training simulation, VBS3 provides several developer tools that can facilitate research on automated scenario generation, including a real-time scenario editor, an offline mission editor, tools for importing new 3D assets, and flexible terrain creation functionalities. VBS3 is used widely in the U.S. Army, and it is integrated with GIFT 2017-1 through a previously developed gateway module. Further, our work with VBS3 will build upon prior research by IAI to devise low-cost assessment frameworks for intelligent tutoring systems through feedback from subject matter experts.

During the first year of the project, we will focus on automated scenario generation in the task domain of Call for Fire training. The CFF task domain in VBS3 will encompass scenarios in which an infantry soldier requests indirect fire from supporting artillery (e.g., unmanned aircraft) on an identified target. The steps of this task include identifying the target, waiting for the artillery to move into position, calling for artillery fire using an established communication protocol, adjusting artillery fire, and providing a damage assessment. "Child" scenarios in the Call for Fire task domain will modify the type, visibility, and movement of the target; augment surrounding terrain and vegetation; change the weather and time-of-day; impact radio communications with artillery operators; augment the type of artillery fire (e.g., smoke, explosive); and influence the accuracy and damage of the artillery fire.

PRELIMINARY FINDINGS ON AUTOMATED GENERATION OF INITIAL SCENARIO CONDITIONS

As a starting point, we developed a prototype system using a multi-armed bandit (MAB) computational formalism, consisting of several components of the proposed deep RL pipeline. The MAB implementation utilizes initial versions of a scenario adaptation library, a simulated learner, and a reward function.

A multi-armed bandit is a class of sequential decision problem in which a set of resources must be allocated between competing choices. MABs are related to reinforcement learning, but they do not account for stochastic effects of sequential decisions on environment states. In an MAB, each choice, or arm, has a defined reward unknown to the system, thus requiring it to explore different choices to learn which of the choices provides optimal expected reward over a finite series of trials. Typically, bandit algorithms are designed to minimize regret, which is the difference between the reward accumulated by the system and the reward the system would have received if it had pulled the optimal arm at every trial. Depending on a variety of factors such as the type of rewards (stochastic, non-stochastic) or the type of regret being minimized (instantaneous or cumulative), different algorithms have been shown to obtain near optimal solutions (Vermorel, 2005). Variants of MABs have been shown to be an effective solution for a variety of tasks such as sequencing learning activities (Liu, 2014) and playing real-time strategy games (Ontanón, 2017).

The first step in formalizing scenario generation as a MAB is defining a scenario adaptation library for the Call for Fires task. As MABs do not have a concept of state, and thus do not capture changes in the simulation environment, we focus our scenario adaptation library on initial conditions of Call for Fires scenarios. In this prototype we focus on 3 categories of initial conditions: weather, time of day, and target mobility. These categories were chosen because of they affect the difficulty of a Call for Fires scenario, and they can also be realized in the VBS3 environment. We defined three possible values for weather (clear, cloudy, rain), three possible values for time of day (day, dusk, night), and two possible values for target mobility (still, moving). This corresponds to 18 possible scenarios that could be generated and evaluated by the MAB.

Next, to provide data to train the system we created a set simulated learners. Each simulated learner consists of a competency score from 0 to 1, representing their ability for a Call for Fire task. To generate rewards for each scenario, we crafted a reward function that takes into account both the difficulty of the scenario and the skill level of the student. Difficulty levels were authored for each type and value of initial condition, with values being averaged to determine the difficulty of each generated scenario. The difficulty level was then combined with the learner's competency score to generate the probability that the learner would increase their competency level from creating the exercise. A 0 or 1 reward was then generated for the trial by sampling from a Bernoulli distribution that was parameterized using the combination of scenario difficulty and learner competency level.

We ran MAB simulations for two populations of simulated learners. For the first simulation, a learner was selected for each trial from a Gaussian distribution centered around a "low" competency score (*M* = .2, *SD* =.1). For the second simulation, learners were selected from a distribution centered around a high competency score (*M* = .8, *SD* = .1). For each simulation, we ran 50,000 trials of an 18-armed bandit using the UCB1 algorithm to manage exploitation/exploration of the arms. Figure 2 shows the average rewards of the top-5 arms (i.e., generated scenarios) for both types of simulated learners. We observe that after some shuffling, each arm begins to stabilize around the "true" reward for that given scenario. For the Low Competency learner group, the scenarios recommended are all "easier" scenarios with non-moving targets and high visibility, which is to be expected given that our reward formulation does not expect low-competency learners to benefit significantly from difficult scenarios. Similarly, the High Competency learner group favors more difficult scenarios featuring moving targets and poor weather/visibility.

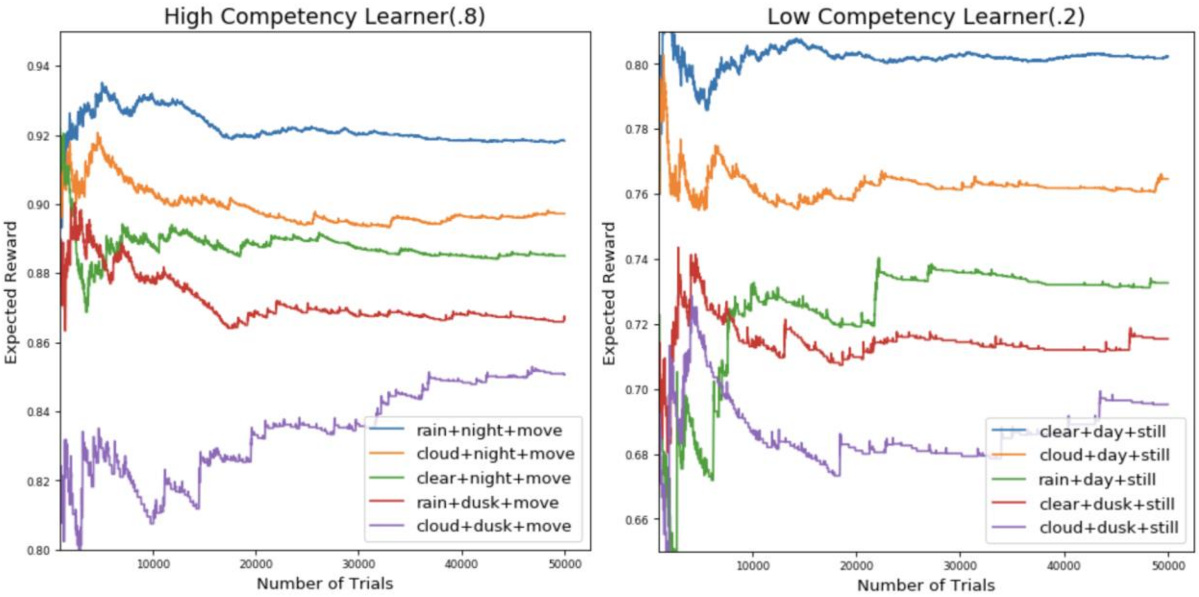


Figure 2. Average rewards over MAB trials for top-5 generated CFF scenarios for high- and low-competency simulated learners.

This prototype highlights key design considerations for deep RL-based scenario generation, but it also has several limitations. First, since MABs have no concept of state, they are not necessarily the ideal formalism for generating and evaluating dynamic, adaptive training scenarios required by more complex CFF tasks; MABs are well suited for generating the initial conditions of simple training scenarios but not sequential events. In future iterations, we will utilize reinforcement learning techniques that account for sequential decisions in order to address this additional source of complexity in scenario generation.

A second limitation is that our current simulated learner and reward models only consider one competency and reward source. As we move forward, the system will need to consider multiple learning objectives and trade-offs between them. Additionally synthesized data will eventually need to be replaced or validated with data from real human learners.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Automated scenario generation is likely to serve a key role in the future of simulation-based training because of its significant potential to reduce the cost of creating novel scenarios and expand access to high-quality simulation-based training. Data-driven approaches to automated scenario generation hold promise for enhancing trainee learning experiences by leveraging recent advances in both machine learning and interactive narrative technologies. We have presented an overview of a deep RL framework for data-driven automated scenario generation, which formalizes the task in terms of sequential adaptations to a canonical "parent" scenario in order to generate "child" scenarios that can be evaluated with simulated or human learners to assess learning outcomes. Automated assessments of trainee learning outcomes drive the generator to iteratively refine its scenario generation policies and tailor scenario generation to individual learners and instructor training objectives. During the initial stages of the project, we are investigating deep RL-based scenario generation in the context of CFF training using the VBS3 simulation environment. To serve as an initial proof-of-concept for data-driven automated scenario generation, we conducted a preliminary investigation of multi-armed bandit techniques for generating initial conditions of CFF training scenarios. Preliminary results indicate that multi-armed bandits, combined with a simple simulated learner model and scenario adaptation library, can produce a ranked ordering of automatically generated training scenarios that are tailored to learners' individual differences.

In future work, we plan to significantly expand the scenario adaptation library to capture a broader range of possible transformations to "parent" training scenarios, including sequential adaptations that can be performed dynamically over the course of a scenario. This will allow us to expand our formulation of automated scenario generation beyond initial scenario conditions and begin exploring deep RL techniques. Further, we plan to investigate richer simulated learner models that can serve as a bootstrapping mechanism for automated scenario generation, as well as multi-objective rewards to enable scenario generation that accounts for complex tradeoffs between complementary and competing learning objectives. Finally, we plan to investigate manual, semi-automated, and automated techniques for realizing generated scenarios in VBS3, enabling human learners to interact with adaptive training scenarios that have been generated using deep RL.

ACKNOWLEDGMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory under cooperative agreement W911NF-18-2-0020. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep Learning (Vol. 1)*. Cambridge: MIT press.

Liu, Y. E., Mandel, T., Brunskill, E., & Popovic, Z. (2014). Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. *Proceedings of the 7th International Conference on Educational Data Mining* (EDM-2014), London, UK, pp. 161-168.

McCalla, G. & Champaign, J. (2013). Simulated Learners. *IEEE Intelligent Systems*, 28(4), 67-71.

Ontanón, S. (2017). Combinatorial multi-armed bandits for real-time strategy games. *Journal of Artificial Intelligence Research*, 58, 665-702.

Riedl, M. O., & Bulitko, V. (2012). Interactive narrative: An intelligent systems approach. *AI Magazine*, 34(1), 67.

Rowe, J., Pokorny, B., Goldberg, B., Mott, B., and Lester, J. (2017). Toward Simulated Students for Reinforcement Learning-Driven Tutorial Planning in GIFT. *Proceedings of the 5th Annual GIFT User Symposium (GIFTSym5)*. Orlando, Florida

Sawyer, R., Rowe, J., & Lester, J. (2017). Balancing Learning and Engagement in Game-Based Learning Environments with Multi-Objective Reinforcement Learning. *Proceedings of the 18th International Conference on Artificial Intelligence in Education (AIED-2017)*, Wuhan, China, pp. 323-334.

Schatzmann, J., Weilhammer, K., Stuttle, M., & Young, S. (2006). A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The Knowledge Engineering Review*, 21(2), 97-126.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Vermorel, J., & Mohri, M. (2005). Multi-Armed Bandit Algorithms and Empirical Evaluation. *Proceedings of the 16th European Conference on Machine Learning (ECML-2005)*, Porto, Portugal, pp. 437-448.

Wang, P., Rowe, J., Min, W., Mott, B., & Lester, J. (2017). Interactive Narrative Personalization with Deep Reinforcement Learning. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-2017), Melbourne, Australia, pp. 3852-3858.

ABOUT THE AUTHORS

***Dr. Jonathan Rowe** is a Research Scientist in the Center for Educational Informatics at North Carolina State University. He received the Ph.D. and M.S. degrees in Computer Science from North Carolina State University, and the B.S. degree in Computer Science from Lafayette College. His research is in the areas of artificial intelligence and human-computer interaction for advanced learning technologies, with an emphasis on game-based learning environments, intelligent tutoring systems, user modeling, educational data mining, and computational models of interactive narrative generation.*

***Mr. Andy Smith** is a Research Scientist in the Center for Educational Informatics at North Carolina State University. He received his M.S. in Computer Science from North Carolina State University, and his B.S. degrees in Computer Science and Electrical and Computer engineering from Duke University. Prior to graduate school Andy worked as an Underwater Robotics Engineer at SPAWAR SSC Pacific in San Diego, CA. His research is focused on the intersection of artificial intelligence and education, with emphasis on user modeling, game-based learning, and educational data mining.*

***Dr. Robert Pokorny** is Principal of the Education and Training Technologies Division at Intelligent Automation, Inc. He earned his Ph.D. in Experimental Psychology at University of Oregon in 1985, and completed a postdoctoral appointment at University of Texas at Austin in Artificial Intelligence. Bob's first position after completing graduate school was at the Air Force Research Laboratory, where he developed methodologies to efficiently create intelligent tutoring systems for a wide variety of Air Force jobs. At Intelligent Automation, Bob has led many cognitive science projects, including adaptive visualization training for equipment maintainers, and an expert system approach for scoring trainee performance in complex simulations.*

***Dr. Bradford Mott** is a Senior Research Scientist in the Center for Educational Informatics at North Carolina State University. Prior to joining North Carolina State University, he served as Technical Director at Emergent Game Technologies where he created cross-platform middleware solutions for video game development, including solutions for the PlayStation 3, Wii, and Xbox 360. Dr. Mott received his Ph.D. in Computer Science from North Carolina State University in 2006, where his research focused on intelligent game-based learning environments. His current research interests include computer games, computational models of interactive narrative, and intelligent game-based learning environments.*

***Dr. James Lester** is Distinguished Professor of Computer Science at North Carolina State University, where he is Director of the Center for Educational Informatics. His research centers on transforming education with technology-rich learning environments. With a focus on adaptive learning technologies, his research spans intelligent tutoring systems, game-based learning environments, affective computing, and tutorial dialogue. The adaptive learning environments he and his colleagues develop have been used by thousands of students in K-12 classrooms. He received his Ph.D. in Computer Science from the University of Texas at Austin in 1994. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI).*

Automating Variation in Training Content for Domain-general Pedagogical Tailoring

J. T. Folsom-Kovarik[1], Keith Brawner[2]
Soar Technology, Inc.[1], U. S. Army Research Laboratory[2]

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Goldberg, & Holden, 2012) is able to tailor training content selection and presentation in order to give individual learners the support or challenge they need (Goldberg, Sottilare, Brawner, & Holden, 2011). However, the effectiveness of tailoring is always limited by the choices available. Automated scenario generation (ASG) offers promise to create many more variations on training content than human experts can create alone. A proof of concept ASG implementation is being researched and developed. The ASG can create variants on training scenarios that encompass multiple types and parts of scenario content to include simulation events and narrative, entity location and size, and feedback or framing text. Combining several such variants will let GIFT simultaneously support and challenge different learning objectives in one training scenario.

A key insight in the work is that generated variants are labeled in a domain-general way according to their predicted impact across learning objectives. As a result, all the details of variations are expressed to GIFT in a manner easily processed by general pedagogical algorithms. Furthermore, the domain-general dimensions of support and challenge for each learning objective define a space within which the ASG components should search. That is, ASG does not simply find random variations on the scenario factors it can control directly, such as entity location and size. Apparent variation in those factors could easily turn out to be different only at a surface level – not different at a level that adds new kinds of support or challenge to the library of choices available to GIFT pedagogical tailoring. A key portion of this work is not the simple generation of many superficially different scenarios, but the generation of scenarios which are sufficiently different from one another while still being pedagogically valid.

Two steps enable ASG to create variants that are valuable for GIFT to tailor learning content. First, ASG must perform search in a space where movements are possible to directly control, such as entity size and location. Second, ASG must translate the variants it made in that space into the space defined by multidimensional impact on learning support and challenge. The first step is carried out by a form of evolutionary algorithm called novelty search. The second step is carried out by defining a cognitive science-based understanding of learning factors and creating domain-specific rules that translate expert knowledge into the terms of the generalized framework.

In the initial days of research and development, ASG is being prototyped in a specific training system for decision-making in the employment of small unmanned air systems (SUAS). The training system contains expert-authored content such as maps with geographical features, tactical objectives and constraints, friendly and hostile entities; text briefing materials and initial conditions; and learner decision prompts and feedback. In the current state of research, an example is presented using map variation. The example illustrates how novelty search programmatically creates candidate arrangements of elements on the training scenario map, then labels each one according to its predicted impact on learning. Specifically those map variations that change the scenario challenge level are stored in a library for the GIFT pedagogical module to select, using the domain-general labels on each variant. As research and development continues, more elements of training such as text content will also be varied using ASG. The result of having many variants is that GIFT may individualize the training experience in order to simultaneously support one learning objective and challenge another, according to learners' needs.

BACKGROUND AND RELATED WORK

ASG has combined relevance across multiple research fields. This section discusses the current state of the art in (1) evolving scenario content, (2) novelty search as an approach to addressing issues of traditional evolutionary algorithms, and (3) computational accounting of pedagogical impact.

First, there have been several successes in procedurally generated content for games or training scenarios. Evolutionary algorithms have generated content such as scenario terrain, behaviors, events, and narrative (Luo, Yin, Cai, Zhong, & Lees, 2016; Stanley, Bryant, & Miikkulainen, 2005; Zook, Lee-Urban, Riedl, Holden, Sottilare et al., 2012). Evolution is well-suited to spaces where there are too many possible variations to explore them all or randomly choose variants to evaluate. In generating training content, effective search is needed because it may not be easy to predict how the changes that are easy to make, such as terrain or unit locations, will affect the desired outcomes, which are to change how instruction works for learners. As a concrete example, changing the position of an enemy unit from "left of the hill" to "right of the hill" may have no training effect for a ground-based tactical movement scenario, but significant training effect for an indirect fire scenario, where accounting for wind velocity and smoke effects is a training objective.

Second, evolutionary algorithms as a class suffer from certain shortcomings that are broad and practical in their importance. Evolution can require days to complete, or may demand high-performance server clusters. A key reason is that evolution typically needs to be carefully tuned to avoid premature convergence on one local optimum, finding the same variants repeatedly instead of new ones. However, a recent advance in evolutionary algorithms suggests that an entirely different approach both yields better results and reduces required computer resources. This approach is novelty search (Lehman & Stanley, 2008, 2011). Novelty search replaces evolution toward a higher fitness with evolution toward increasingly different individuals. Novelty search has been used with success to evolve content similar to training scenarios, such as game levels (Liapis, Yannakakis, & Togelius, 2015).

Third, there is the question of letting novelty search predict the training impact of generated variants. Instructional designers, educators, and cognitive psychologists are among those who have created frameworks for predicting the effect of training scenarios, interventions, and other content on individual learners and in different training contexts (Campbell, Ford, Campbell, & Quinkert, 1998). Two factors that have been recently studied are scenario helpfulness and complexity. Helpfulness describes the *explicit* interventions that can be part of training, such as help messages, hints, or formative feedback—are they specific or broad, immediate or delayed, and so on (Shute, 2008). Complexity gives a good complement by measuring *implicit* interventions which vary scenario content in order to support or challenge learners, such as tailoring the number of enemies or the amount of time remaining to carry out a task (Dunne, Sivo, & Jones, 2015). Measuring different dimensions or categories of helpfulness and complexity has driven tailoring in past work, but has been manually defined for each variant (Folsom-Kovarik, Newton, Haley, & Wray, 2014; Folsom-Kovarik, Sukthankar, & Schatz, 2013; Graffeo, Benoit, Wray, & Folsom-Kovarik, 2015).

In summary, evolutionary approaches may be able to generate meaningful scenarios from the infinite set of possibilities, to do so quickly when using a metric such as novelty search, and objectively measure instructional relevance. The natural divisions of the "genotype" and "phenotype" space within an evolutionary approach lend themselves to representing the literal scenario content (genotype) and its learning impacts (phenotype). An evolutionary content generation method that would also let end users such as instructors and subject-matter experts understand and control the content evaluation in an objective manner would help to improve usability and user acceptance of the approach (Folsom-Kovarik, Wray, & Hamel, 2013).

NOVELTY SEARCH AND APPLICATION TO ASG

Evolutionary algorithms are appropriate methods to search when a space is too high-dimensional, unevenly gradiated, or otherwise inappropriate for simpler enumeration or gradiant descent methods. Evolution typically maintains a notional population of points in the search space which are evaluated to find their fitness for the purpose at hand. The points in the population are then combined and varied with operators that aim to increase fitness of the next generation and remove points that have lower fitness. Novelty search addresses some limitations of evolutionary algorithms. Instead of working to increase fitness, the aim is to increase novelty and explore points that are as different as possible from what has been seen before. In this way, novelty search attempts to remove premature convergence concerns typical in evolutionary algorithms and produce many variants that can be filtered for fitness to a specific need. This is specifically an advantage in the training domain space where differences among the scenarios is an explicitly stated goal.

This section describes the current state of the novelty search algorithm under development. Novelty search efficiently finds variants that are new in a domain-general sense. That is, the variants provide a different manner of support or challenge than any variant already available. As novel variants are created offline, they can be stored for human review and access by GIFT. The novelty search is an "anytime" process meaning that it can provide results immediately or continue to improve the results as more compute time is available when not actively training. GIFT is then able to select in real time during training between variants using its existing, domain-general Pedagogical Module. Instructors will be able to see what variants are available and identify any gaps that still need to be evolved. This allows for both the generation of content that the instructor can approve and for the further development of training exercises for students if the amount of approved content is exhausted.

The current implementation of novelty search is built on the open-source library Distributed Evolutionary Algorithms in Python (DEAP) (Fortin, Rainville, Gardner, Parizeau, & Gagné, 2012). DEAP offers a combination of fast prototyping now with fast computation and variant generation in future deployment. Like other evolutionary algorithms, novelty search depends on effective design of (1) genotype representation, (2) genetic operators, and (3) fitness function or in this case novelty evaluation function. In the current state of research, these are domain-specific. However, future research may be able to identify opportunities for reuse across broad domains such as the spaces of all images or all text documents.

First, ASG differentiates genotypes from phenotypes in this way. Genotypes are those objects, such as strings of digits that can be easily changed during evolution, while phenotypes are the objects that can be evaluated for their novelty. The phenotype is the scenario variant itself that which the learner experiences during training and which the instructor must agree provides appropriate support and challenge. As a result of this difference, evolution is not needed when it is possible to jump straight to the desired phenotype. Instead, the separation of genotype and phenotype is necessary because the phenotype can be measured on dimensions that matter to instruction, like complexity and helpfulness, but the genotype cannot be measured until it is transformed into a phenotype. Conversely, at the genotype level there are a set of changes which are easy to apply to generate new variants, but it turns out to be difficult to make changes at will in the training experience phenotype, because human creativity would be required.

The genetic representation currently used throughout the rest of this work in ASG is a direct representation. More complex representations such as neural networks or hypercubes (Kocmánek, 2015) have also been used in novelty search but were deemed unnecessary at this stage of development. The genetic representation encodes each element of the evolved training scenario one-to-one in a vector of descriptive values. For example, in order to evolve locations of objects in a two-dimensional space, the genetic representation would describe each object with its type, x-coordinate, y-coordinate, and perhaps scale or rotation values. The current representation describes points, lines, and regions in two-dimensional space, which is hypothesized to be extensible to multiple domains.

Second, genetic operators in ASG are designed to make changes to genotypes. The changes are not guaranteed to produce a better genotype, but they should be designed to build on what has already been evolved and create new genotypes that have a reasonable possibility to be viable. The genetic operators used are element insertion, single-point mutation, and single-point crossover. Element insertion increases the complexity of lines and regions by adding a new point at random to the genotype. Single-point mutation chooses a numeric value uniformly at random and changes it by adding a random perturbation with Gaussian distribution about zero. The crossover operator combines two existing genotypes by choosing a point in the vector and taking all elements to the left of that point from one genotype, all elements to the right of that point from the other. Since direct representations have been well-studied in other evolutionary algorithms, these operators are standard in the field and do not introduce additional domain specificity.

Third, the novelty evaluation function in ASG is the tool that determines when evolution has produced a variant that is new in an interesting way, as opposed to a variant that appears to be new on the surface but does not provide any difference in training support or challenge. The terms "support" and "challenge" are considered to be opposite ends of a single continuous scale for the purposes of this work. The evaluation phase consists of applying domain-specific rules to each phenotype (training variant) in order to find its value on domain-general dimensions. Four domain-specific rules were created as a proof of concept and are described in the ASG Example section below. The domain-general measures that result from these rules describe facts about the training such as complexity of meeting one learning objective or another.

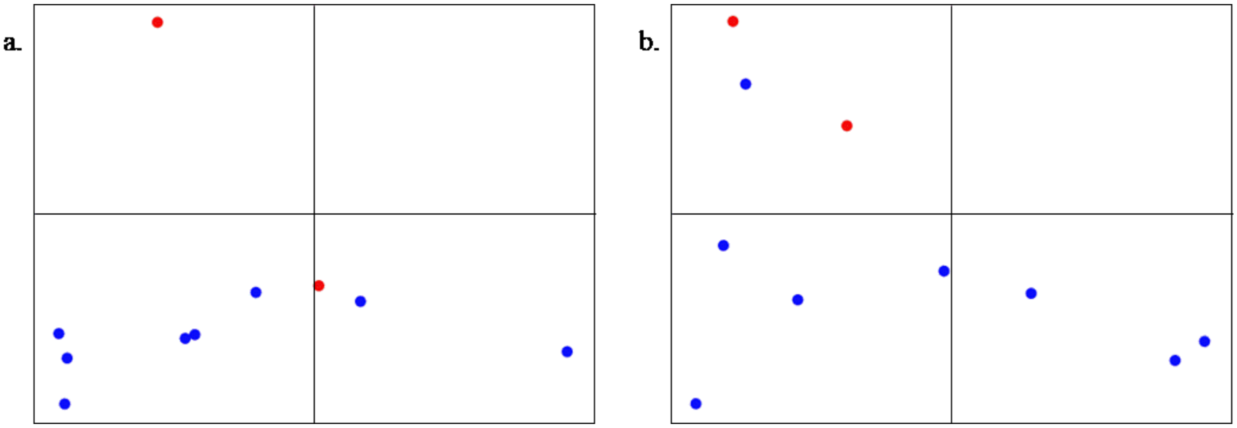


Figure 1: Training complexity (a) in the first generation and (b) after 20 generations. The different point colors and x,y locations (spread) visualize the diversity of training options that the evolved variants offer.

ASG determines which variants are novel in the training challenge sense by clustering variants on the training complexity values, finding the k nearest neighbors (k=2 for efficiency), and selecting the variants which maximize Euclidean distance from their respective nearest neighbors. A hall of fame was maintained to provide persistence across generations of the current maximally novel individuals. In this scheme, different factors that separately affect complexity formed additional dimensions in the complexity measure. Examples were number of distractors or time constraints. As such, evaluation was found to require a scaling step in order to make different dimensions comparable and prevent one dimension from outweighing others.

The outcome of the overall algorithm for novelty search is an increasingly diverse collection of training variants. Figure 1 demonstrates the difference between an initial generation and the variation after 20 generations. The example complexity measures described below are projected into two-dimensional space. The increasing distance between the individuals after evolution indicates that novelty search produced variants which provide GIFT with more choices between noticeably different levels of challenge and complexity.

A DOMAIN-GENERAL REPRESENTATION OF SCENARIO CONTENT

The creation of a domain-general representation of learning impact enables contributions from many instructors, authors, and researchers to work together to increase GIFT tailoring options. A computational representation of factors that impact learning also enables automatically evaluating what is novel in ASG and what will help learners at scenario runtime. GIFT has previously conducted a literature review to support the selection of domain-general factors, including complexity. This proof of concept adds to that review a high degree of precision that breaks down support and challenge into multiple contributing factors that can be separately measured, varied, and objectively compared across learning domains and systems.

Diverse instructional theories suggest categorizing or sequencing learning tasks based on a continuum of complexity. Gagné (1965) organized learning tasks into a broad hierarchy consisting of stimulus recognition at the least complex through application and problem solving as the most complex tasks for any particular identified skill. A similar concept of task complexity has been used in past research with a Dynamic Tailoring System (DTS) that could choose in real time between training variants that were labeled by human experts with scalar complexity values (Wray & Woods, 2013).

In the context of ASG, a single scale for task complexity is not expressive enough to capture the different ways in which the same task can be varied to be more or less complex. Many generated variants are likely to have the same complexity when measured on a single scale – an example might be a hypothetical GIFT question bank that contains a hundred different multiple choice items. Empirically some challenge learners more than others and are answered correctly less often. However, they might all be evaluated as equal in complexity because they all require simple recognition (of the correct choice). In the multiple choice situation, which seems not atypical, two possible approaches could let a computer system differentiate the available variants in advance without human expert labeling. First, GIFT could have an extremely fine-grained hierarchy of sub-concepts. In this case, GIFT could differentiate and sequence the available variants based on hierarchical relationships between the sub-concepts such as prerequisites. This approach is not attempted in ASG. The second approach, which is explored here, is to increase the dimensions by which variants may be described. Complexity itelf must be analyzed in more detail.

Dunne, Cooley, and Gordon (2014) conducted an initial analysis of factors that contribute to learning complexity. These factors included task complexity factors such as number of actions required and number of interdependent actions, as well as learning context factors such as number of possible ways to complete a task and number of distractors. These factors appear in Table 1. On the other hand, Table 1 also introduces a notional definition of helpfulness. As a complement to complexity, helpfulness has not yet been operationalized to provide concrete measures and will be discussed here at an early stage of exploration.

First, Dunne and colleagues suggest theory-based, countable measures that help provide a multidimensional framework for objectively measuring complexity. Complexity increases with each of the factors in Table 1, although possibly nonlinearly. Current work with ASG is working to build rules that predict the impact of scenario variants by counting factors such as the number of cues, actions, and distractors in each variant. Each dimension in the framework is furthermore related to one equation that calculates scenario complexity and has been initially validated with empirical study of a military training sequence in the same citation. The rules that carry out counting the complexity factors are domain-specific, but they result in domain-general measurements. The domain-general measures let the GIFT pedagogical module work without domain-specific knowledge and enable objective comparison across variants.

Table 1: Domain-general dimensions describing challenge and support for each task or learning objective.

Measuring Complexity		Measuring Helpfulness
Number of cues		Attention via perceptual arousal
Number of actions		Attention via inquiry arousal
Number of subtasks across actions		Relevance via previous link
Number of interdependent subtasks		Relevance via needs link
Number of possible paths		Confidence via evaluation link

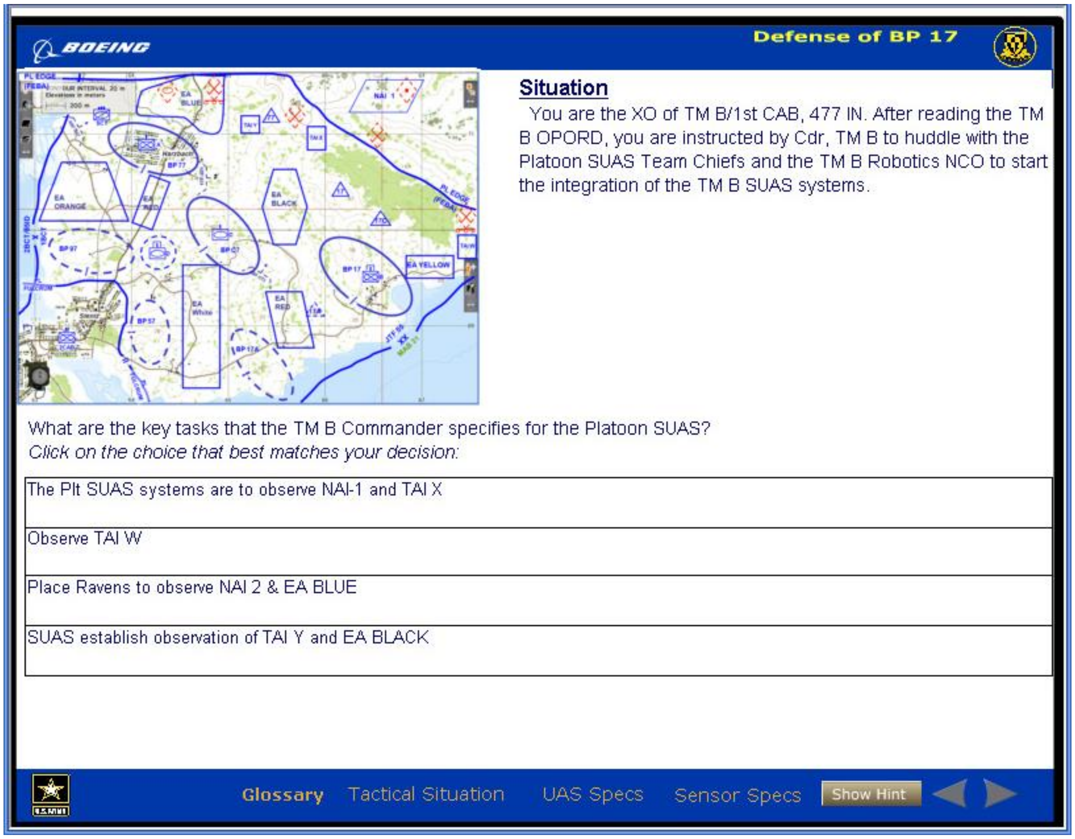
Number of criteria to satisfy		Confidence via learner control
Number of conflicting paths		Satisfaction via feedback positivity
Number of distractors		Satisfaction via future link

Second, a measure is needed which describes the dimensions on which extrinsic or direct interventions can be measured. Interventions such as hints, help messages, and text documents that deliver remediation vary in their helpfulness with respect to specific learning objectives. As an intuitive example, a help message delivered inside a scenario by a character over the radio can be clear, concise, and on point to provide support, or it can deliberately challenge learners by being ambiguous, wordy, or distracting. A cognitive science basis for enumerating the possible differences in how helpful scenario components are lies in the ARCS model of instructional design (Keller, 1987). This model describes factors of attention, relevance, confidence, and satisfaction that motivate an adult learner to engage with learning content. Unlike the Dunne model, research is still needed to produce an accounting of how a computer system can see factors in this model. One example that moves the ARCS model toward countable dimensions might be a heuristic measure of inquiry arousal from counting the number or frequency of keywords like “how” and “why.”

The combination of complexity and helpfulness is hypothesized to provide ASG with multiple objective measures to describe and differentiate the impact of every variant on different learning objectives. In this way, a domain-general representation of scenario content is hypothesized to increase the opportunities to apply learning theory in GIFT’s automated design and selection of content that is tailored for learners. ASG can augment a hierarchical analysis or a fragmented, parts-to-whole sequencing with recommendations that reflect how adults learn material of increasing complexity in context (Reigeluth & Stein, 1983).

ASG EXAMPLE FOR SCENARIO LAYDOWN

ASG is being developed and evaluated in the context of existing training for proper use of small unmanned air systems (SUASs). The training consists of sequential problem presentations in the context of a narrative supported by mission briefings and maps depicting the area of operations (Figure 2). Learners are also presented with adaptive hints and texts for remediation depending on their performance. The system has been designed to teach nine terminal learning objectives and 48 enabling learning objectives, a huge number of dimensions for evolution to explore if all combinations of learning objectives can eventually be varied in complexity and helpfulness. In the present research and development, a subset of three learning objectives was chosen for initial exploration. The first target for evolution was the mission map. Future work is planned to evolve text-based content such as briefings, pop-up events, and hints or remediation documents.



Situation

You are the XO of TM B/1st CAB, 477 IN. After reading the TM B OPORD, you are instructed by Cdr, TM B to huddle with the Platoon SUAS Team Chiefs and the TM B Robotics NCO to start the integration of the TM B SUAS systems.

What are the key tasks that the TM B Commander specifies for the Platoon SUAS?

Click on the choice that best matches your decision:

- The Plt SUAS systems are to observe NAI-1 and TAI X
- Observe TAI W
- Place Ravens to observe NAI 2 & EA BLUE
- SUAS establish observation of TAI Y and EA BLACK

Figure 2: The SUAS training domain for developing and evaluating ASG.

The ASG example seeks to evolve variants on the mission map depicted in the top left of Figure 2. In this example, the elements that can be evolved are shown in Figures 3 and 4. These include the locations of friendly and hostile units, a no-fly zone (red oval), and the shapes of roads, water, and forest terrain features. According to the ASG algorithm described above, these elements could be moved easily on the generated map variants. The next step was to demonstrate how the variants could be measured in training complexity space and selected as being more or less novel from a training impact perspective.

Three training complexity measures were created for the example implementation. The measures were simple rules reflecting three of the learning objectives in the real training system: enemy air defense avoidance, recon and surveillance, and airspace coordination procedures. The simplified rules showed examples of three dimension types: continuous scalar, discontinuous scalar, and categorical. (1) For enemy air defense, one rule was created that stated training complexity increases with proximity to an enemy unit. The enemy was considered to have air defense capabilities that made it difficult to operate when near the enemy, (2) for recon and surveillance, two rules defined one complexity dimension. If an enemy unit was located within a forest region, the complexity of the training increased with the size of the forest. If the enemy was outside a forest, the complexity decreased in proportion to the distance from the enemy to the nearest edge of a forest region, and (3) For airspace coordination, the rule was that complexity was high when a no-fly zone lay between the enemy and friendly units, while complexity was low otherwise. In Figures 3 and 4, red dots indicate scenarios with high complexity in this dimension while blue dots have low complexity.

During novelty search, the first generation maps (Figure 3) typically did not even contain both a friendly and a hostile unit. This helps explain why they are all rated as similar in the complexity of air defense avoidance (Figure 1 above).

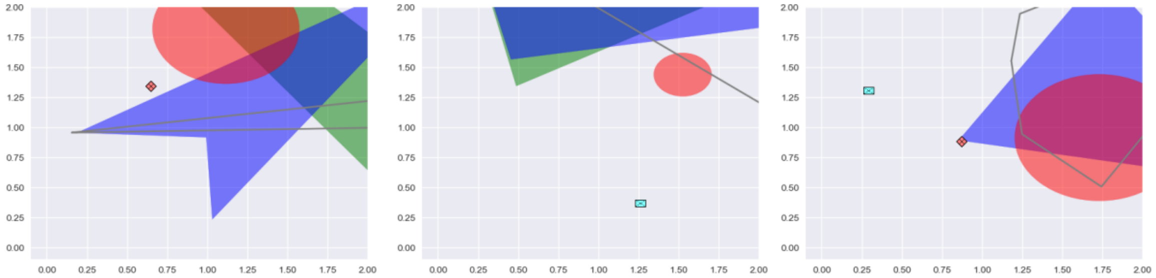


Figure 3: First generation of evolving scenario variants.

The last-generation maps (Figure 4) have evolved a greater frequency of having one friendly and one hostile unit on the map, which probably helped to explore more possible values of air defense avoidance complexity and let ASG provide variants at more places on this scale.

The last-generation maps also display increased complexity of the contours around water and forest regions. This is an example of a difference that is visually very apparent but makes no difference for the purposes of measuring training complexity. The value of novelty search using domain-general measures of the variants is that the simplicity or complexity of the scenarios are evaluated without regard to surface details except where a rule tells ASG that those will change the learning experience.

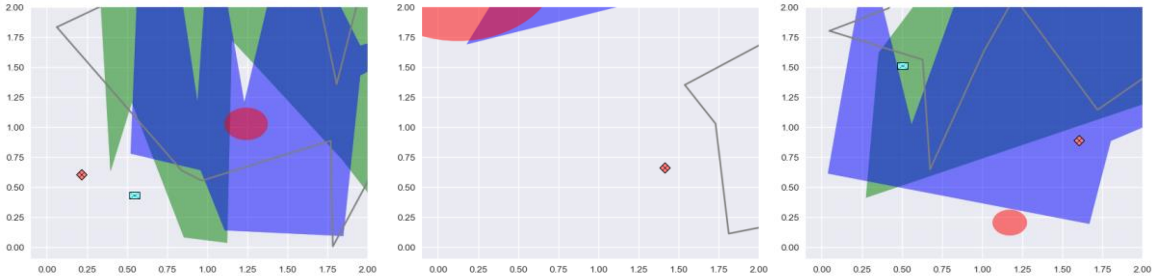


Figure 4: Last generation scenario variants.

In summary, the work of developing an example of ASG in a GIFT-enabled training domain has helped to develop some of the proposals and surface findings in this paper, as well as considerations that will be addressed in ongoing research and development. The initial novelty search examples presented here used only a small fraction of the potential dimensions that could be measured to describe the SUAS training. As a result, there is great potential for novelty search to create a large library of scenario variants that offer GIFT any desired combination of support and challenge for delivering tailored training.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Presented above are the first steps in the investigation of Automated Scenario Generation. This research divides the problem up into three problems – downselecting from an infinite number of possible scenarios, doing so in reasonable time, and making scenario variants which are instructionally relevant. The proposed approach uses genetic algorithms, with a novelty search fitness function and a domain-general representation of scenario content, to enable variant selection without a specialized pedagogical module and to present to instructors and students. Generally, the above research is incomplete – it did not include an analysis on whether the generalized scenarios are useful to SMEs or what dimensions of changes are needed or desired. This work is yet to be performed.

This work, thus far, is able to make use of the existing GIFT logic, structure, and modules. The scenario selection logic within the pedagogical module should be written in a general enough manner so as to be able to be applied to a large number of generated scenarios. Performance assessment within the generated variants is also needed. Fundamentally, GIFT will have to provide the same tutoring to the new scenarios as it does to the old – the functionality is built into GIFT and the DKF structure. Integration at the end of the project may be as simple as a pointer to the optimal fit in a library of generated variants when the student reaches the appropriate experience at runtime. In the ideal case, performance assessment could be dynamic and follow rules drawn from or similar to the novelty evaluation rules, in the same manner as the DKF currently allows for pointers to external assessment engines.

Next steps in the near term will be to replace the illustrative ASG example presented in this paper with more realistic rules for complexity. Instructor and SME interviews are planned to determine which learning objectives and aspects of the real training system are likely to be most impactful to vary. Understanding what dimensions human expert instructors actually want and need to vary will inform a more comprehensive set of rules that will help test and improve the efficiency and effectiveness of the ASG approach.

In regards to future work, scenarios represent the most complex piece of content represented within GIFT. The example ASG could easily be extended to other two-dimensional content like images or VBS terrain. Simpler pieces of content, such as prompted, hints, feedback strings, webpages and other items are also shown to the user in GIFT training but are not procedurally generated. Technology to procedurally generate these types of items may have to be implemented differently, as these items may rely on text or image processing techniques rather than modification and generation techniques, but there may be another class of ASG representations and operators that is effective for many types of text content.

REFERENCES

Campbell, C.H., Ford, L.A., Campbell, R.C., & Quinkert, K.A. (1998). *A Procedure for Development of Structured Vignette Training Exercises for Small Groups* Alexandria, VA: Human Resources Research Organization.

Dunne, R., Cooley, T., & Gordon, S. (2014). *Proficiency Evaluation and Cost-Avoidance Proof of Concept M1A1 Study Results.* Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando, FL.

Dunne, R., Sivo, S.A., & Jones, N. (2015). *Validating scenario-based training sequencing: The scenario complexity tool.* Paper presented at the Interservice/ Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL.

Folsom-Kovarik, J.T., Newton, C., Haley, J., & Wray, R.E. (2014). *Modeling Proficiency in a Tailored, Situated Training Environment*. Paper presented at the 23rd Conference on Behavior Representation in Modeling and Simulation (BRIMS), Washington, DC.

Folsom-Kovarik, J.T., Sukthankar, G., & Schatz, S. (2013). Tractable POMDP representations for intelligent tutoring systems. *ACM Transactions on Intelligent Systems and Technology (TIST), 4*(2), 29.

Folsom-Kovarik, J.T., Wray, R.E., & Hamel, L. (2013, July 9-13). *Adaptive assessment in an instructor-mediated system.* Paper presented at the 16th International Conference on Artificial Intelligence in Education (AIED), Memphis, TN.

Fortin, F.-A., Rainville, F.-M.D., Gardner, M.-A., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research, 13*(Jul), 2171-2175.

Gagné, R.M. (1965). *Conditions of Learning*. New York, NY: Holt, Rinehart, and Winston.

Goldberg, B., Sottilare, R., Brawner, K., & Holden, H.K. (2011). *Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions*. Paper presented at the HUMAINE Association on Affective Computing and Intelligent Interaction, Memphis, TN.

Graffeo, C., Benoit, T.S., Wray, R.E., & Folsom-Kovarik, J.T. (2015). Creating a Scenario Design Workflow for Dynamically Tailored Training in Socio-Cultural Perception. *Procedia Manufacturing, 3*, 1486-1493.

Keller, J.M. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development, 10*(3), 2.

Kocmánek, T. (2015). *HyperNEAT and Novelty Search for Image Recognition.* Master's thesis, Czech Technical University in Prague.

Lehman, J., & Stanley, K.O. (2008, August 5-8). *Exploiting Open-Endedness to Solve Problems Through the Search for Novelty*. Paper presented at the 11th International Conference on the Synthesis and Simulation of Living Systems (ALIFE), Winchester, UK.

Lehman, J., & Stanley, K.O. (2011). Novelty search and the problem with objectives *Genetic Programming Theory and Practice IX* (pp. 37-56): Springer.

Liapis, A., Yannakakis, G.N., & Togelius, J. (2015). Constrained novelty search: A study on game content generation. *Evolutionary computation, 23*(1), 101-129.

Luo, L., Yin, H., Cai, W., Zhong, J., & Lees, M. (2016). Design and evaluation of a data-driven scenario generation framework for game-based training. *IEEE Transactions on Computational Intelligence and AI in Games*.

Reigeluth, C., & Stein, R. (1983). *Elaboration theory*.

Shute, V.J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S., & Holden, H.K. (2012). The generalized intelligent framework for tutoring (GIFT). Orlando, FL: US Army Research Laboratory Human Research & Engineering Directorate.

Stanley, K.O., Bryant, B.D., & Miikkulainen, R. (2005). Real-time neuroevolution in the NERO video game. *IEEE transactions on evolutionary computation, 9*(6), 653-668.

Wray, R.E., & Woods, A. (2013). *A Cognitive Systems Approach to Tailoring Learner Practice.* Paper presented at the 2nd Advances in Cognitive Systems Conference, Baltimore, MD.

Zook, A., Lee-Urban, S., Riedl, M.O., Holden, H.K., Sottilare, R.A., & Brawner, K.W. (2012, May 29-June 1). *Automated scenario generation: toward tailored and optimized military training in virtual environments.* Paper presented at the International conference on the foundations of digital games, Raleigh, NC.

ABOUT THE AUTHORS

***J.T. Folsom-Kovarik, PhD** is the lead scientist at Soar Technology, Inc. for adaptation and assessment within intelligent training. He earned a PhD in computer science from the University of Central Florida in 2012. His research combines modern data science and machine learning approaches with SoarTech's long experience in modeling expert knowledge and human experience. When expert knowledge guides machine learning and data analytics algorithms, they become more applicable and useful in real-world training settings. The combination of approaches can remain feasible when available data is small, concepts evolve over time, or nontechnical users need to control the training.*

***Keith Brawner, PhD** is a researcher for the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED), and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He has 12 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current research is in 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.*

Integrating Sketch Worksheets into GIFT

Kenneth D. Forbus, Thomas Hinrichs, Samuel Hill, Madeline Usher
Qualitative Reasoning Group, Northwestern University

INTRODUCTION

While there is evidence that sketching can improve student learning (e.g. Ainsworth et al. 2011; Jee et al. 2014; Scheiter et al. 2017), sketching has rarely used in intelligent tutoring systems because it has been difficult for software to understand what a student's sketch means. To tackle this problem, our group developed CogSketch (Forbus et al. 2011), which provides a robust model of high-level human visual processing and representation. It has been used to model a variety of human visual reasoning and STEM problem-solving (Forbus et al. 2017), providing evidence that its representations and reasoning can provide a solid basis for creating new kinds of sketch-based educational software.

One such new kind of sketch-based educational software are Sketch Worksheets (Yin et al. 2010; Forbus et al. 2017). In a Sketch Worksheet, students tackle problems by drawing and modifying sketches. At any point, they can request feedback, and then improve their sketch. Gradebook software built into CogSketch enables instructors to rapidly grade sketching assignments. CogSketch also provides instructors with detailed assessment data as to the student's process as well as their final product. Importantly, Sketch Worksheets can be authored by instructors, after learning some basics of CogSketch, without programming. This improves dissemnination, by broadening participation in authoring. Sketch Worksheets have now been deployed in several classrooms and subjects (Garnier et al. 2017; Forbus et al. 2018).

While Sketch Worksheets are useful, we believe there is much untapped potential to be explored for using sketching in new kinds of educational software. This paper describes our next step in exploring sketching in intelligent tutoring systems more broadly, by integrating Sketch Worksheets as a medium in GIFT, to benefit from the adaptivity that GIFT provides, and to provide a new capability for GIFT tutors. We describe the basic ideas of sketch worksheets, how we are integrating them into GIFT, and the prototype Simple Machines tutor we are building as an experimental vehicle. Planned experiments are discussed. While this integration is still in progress, we plan to demo a version of the Simple Machines tutor during the symposium.

SKETCH WORKSHEETS: A BRIEF REVIEW

Here we summarize the basics of Sketch Worksheets, more technical details can be found in [Forbus, et al. 2017]. A student tackling a Sketch Worksheet is trying to solve a problem, whose solution is expressed by them drawing or modifying a drawing. For example, in geoscience, they may be asked to mark up a photograph, indicating the properties of the geological strata it illustrates. In engineering graphics, a student might have to redraw a design shown in perspective projection in orthogonal projection. In cognitive science, a student might have to draw a concept map representing the semantic content of a sentence.

Being able to do this range of tasks with the same software requires a fundamentally different approach than the usual view that identifies sketch understanding with sketch recognition. The mapping from concepts in STEM education to visual shapes is many to many: Recognition typically isn't an option. Instead, people talk when they sketch with each other. CogSketch provides a simple interface that enables students to identify how they are considering their ink as partitioned into objects, and give them a label in terms of concepts from the underlying knowledge base (which, to the student, look like natural language words or phrases). CogSketch computes visual relationships between the ink entities that students draw, including a rich vocabulary of qualitative relationships that can be used to connect spatial concepts to language. When an instructor authors a worksheet, they draw their solution using CogSketch, which analyzes their ink. The instructor marks some subset of the facts CogSketch computes as important, assigning points to each such fact and providing text to be provided if the analog of that fact is not found in the student's sketch. When a student tackles a worksheet, they draw (or modify existing ink, depending on the worksheet) their solution. When they ask for feedback, CogSketch performs the same analysis as it did on the instructor's sketch, and uses analogy (Forbus et al. 2016) to compare the facts computed about the two sketches. Any differences that correspond to important facts lead to the appropriate advice being produced for the student, or an indication that they've successfully finished the worksheet. They are free to continue working on it as long as they like.

For assessment purposes, CogSketch records timestamps for all of the ink, as well as what order entities were drawn in. The state of the sketch at every time the student asked for help is also recorded, so the instructor (or educational data mining software) can examine their performance in detail and look for patterns across students.

INTEGRATING SKETCH WORKSHEETS INTO GIFT

Our approach is to integrate Sketch Worksheets as a new kind of media that can be used in GIFT tutors. Since GIFT is implemented via an Amazon-based cloud, we are building a cloud-based version of CogSketch to support these experiments.

The cloud-based version of CogSketch is called WebSketch. The services are implemented as Docker containers grouped together in a stack that can be deployed on various cloud services. In order to integrate with GIFT (as well as other educational software infrastructures) our WebSketch stack also contains services to support the LTI protocol (Learning Tools Interoperability, https://www.imsglobal.org/activity/learning-tools-interoperability). Figure 1 shows how this would work with GIFT.

GIFT communicates with WebSketch through LTI. When a GIFT course makes use of a Sketch Worksheet, GIFT uses LTI to handoff control to WebSketch. The student works through the worksheet and a score is returned to GIFT. WebSketch is functioning as an LTI Tool Provider and GIFT is an LTI Tool Consumer in this setup.

The LTI Authorization service in the WebSketch container stack handles the initial communications from a Tool Consumer (GIFT in this case). This includes confirming that the request is coming from a valid Tool Consumer that has permissions to use WebSketch and starting an LTI session. The initial communications from GIFT include a unique and consistent identification of the student (anonymized), which worksheet should be used, and a URI to which the student's score should be returned.

If the LTI request is valid and authorized, control is passed to the WebSketch Node Management service, which chooses an available WebSketch node from a pool of nodes. The selected node is used for the student's session with WebSketch. Each time a student requests tutoring advice from WebSketch, the student's score is updated and conveyed to GIFT. When a student is finished working, their sketch is saved in our Sketch Repository. The saved sketch can be accessed later as needed for assessment and aggregate data collection. If a student revisits a given worksheet through GIFT, the worksheet can be retrieved in the state they last left it.

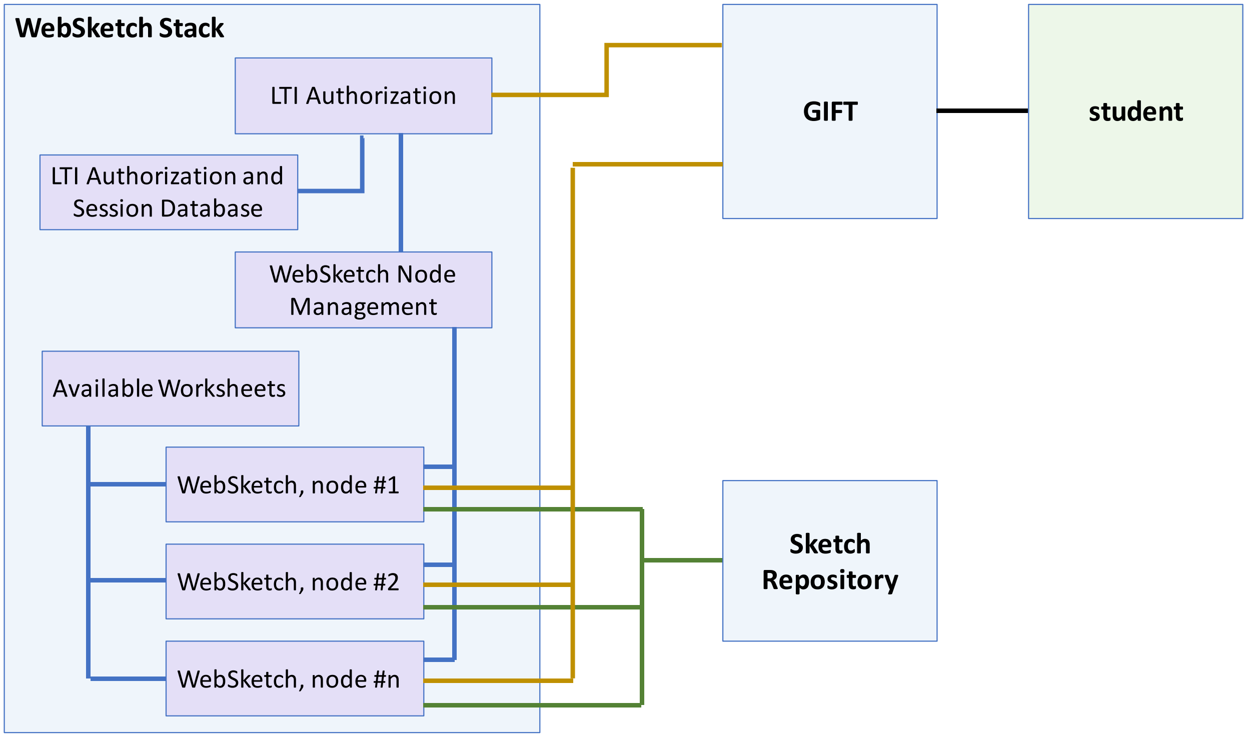


Figure 16. WebSketch/GIFT integration

There are several steps remaining before our initial implementation is finished. The first is a Sketch Worksheet service, which needs to have a repository of blank worksheets, and a registry that connects an ID used in the GIFT tutor with a sketch file. We have implemented a Sketch Repository to store student work, but it is currently running outside the Docker container, whereas for portability it needs to be part of the WebSketch Docker Swarm. There are also a variety of WebSketch UI improvements to be made, including support for CogSketch annotations. We are planning to have these improvements finished before the Symposium.

EXPERIMENT IN PROGRESS: A SIMPLE MACHINES TUTOR

A common topic in STEM instruction for K-12 students, and relevant to understanding and maintaining many kinds of Army equipment, are *simple machines*: Levers, pulleys, screws, and so on. Aside from their practical importance, simple machines provide an interesting application of scientific principles, and provides a bridge between intuition and qualitative understanding to mathematical models that support design and predication. They are also inherently spatial, which makes them a natural for sketching activities. Consequently, we are using GIFT and Sketch Worksheets to create a Simple Machines Tutor.

The learning goals for our curriculum are that, after working through it, a student should be able to

1. Understand the kinematics and force dynamics of simple machines.
2. Recognize structural components, salient relations, quantities and ratios relevant to their operation.
3. Recognize simple machines in the everyday world

4. Understand the tradeoffs between distance, force, and work and how these tradeoffs are manifested in physical systems

5. Have an improved physical intuition for how mechanisms can or will behave and be able to use calculations to verify that intuition

6. Understand the design space of alternative ways to achieve a given effect.

Simple Machines Curriculum Design

The medium of sketching is not limited to representational drawing. It also includes annotating existing sketches or photographs, labeling, and re-arranging constituent components of sketches. With these interactions, it becomes possible to go beyond simple presentation and multiple choice questions to tease out more subtle misconceptions and knowledge gaps.

The curriculum will initially focus on recognition and qualitative analysis of each type of simple machine. Starting with an informal definition and exposure to multiple examples of a simple machine in the context of common everyday devices and situations, the learner is encouraged to compare and make his own analogies to induce a general concept. That generalization can be tested with additional classification examples and near-misses.

Next, more subtle relations can be conveyed in a generative fashion by having the student modify a sketch to alter critical relationships. For example, they might be asked to move the fulcrum in a lever to change it from a first-class to a second-class lever. By not providing explicit choices, it is possible to detect a broader range of misconceptions (e.g., can they even recognize the fulcrum in this context?)

The next activity involves qualitative comparative analysis in which two machines of the same type but different quantitative relations are presented side-by-side. Here, the task is causal reasoning about differences, e.g. which machine would apply greater output force given the same input force. The learner must annotate the depictions to identify which quantity is larger (or smaller) in the selected machine, and also which visual property gave rise to this conclusion, giving a window into their reasoning.

With a solid qualitative foundation, the formal notion of mechanical advantage can be introduced. The three quantities involved (distance, force, and work) will be presented in the context of one kind of simple machine (e.g., inclined plane) and then by analogy those concepts are extended to other machines. So if distance travelled is straightforward in the context of an inclined plane, what does distance travelled correspond to in a screw? (translational distance? distance along the helix?) How about in a block and tackle?

Once correspondences between quantities across different types are established, it becomes possible to draw more abstract analogies between different types of machines. For example, a screw can be conceptually unrolled into an inclined plane. What activities might support comparing the mechanical advantage of one to the other?

As quantities are introduced, it becomes possible to present simple parametric synthesis tasks, in which the student labels a machine’s lengths, angles, and ratios with numerical values to achieve a desired performance. For example, a problem might specify the desired mechanical advantage and one structural parameter, leaving the last parameter open, to be added as an annotation. Finally, when exploring mechanical advantage, we want to avoid functional fixity, in which all machines are seen as force amplifiers, by illustrating the design tradeoffs in other directions. So for example, sometimes the problem will be to attain greater precision rather than force amplification. Vernier calipers exploit the ratio of rotational distance to translational distance to attain high precision. An exercise will have the student modify a sketch by swapping out one or more components of a simple machine (e.g., the pitch of a screw or the diameter) to achieve different kinds of goals, such as minimize displacement, or reduce overall physical size of the machine.

The last set of exercises will addresses structural synthesis. Here, we have to consider simple machines in the context of more complex compound machines. The first kind of synthesis exercise is the sketching analog of fill-in the blank questions. The learner will be presented with an incomplete kinematic chain and a desired global property. They must fill in the missing element by sketching and labeling it, along with its relevant parts and quantities. For example, if the direction of force needs to be reversed, a first-class lever could be used, or a pulley. If rotational to translational conversion is required, either a wheel and axle or a screw could be used.

Another synthesis exercise would be to arrange a fixed set of simple machines into a configuration that achieves a goal. Here, the machines are provided as building blocks and put together, although there could be more than one right answer.

A capstone challenge problem will be to assemble a complex machine in such a way as to demonstrate an understanding of the design space and tradeoffs. Rather than focus on practical quantitative design (which is beyond the scope of this curriculum), the problem may be presented more as the design of a "Rube Goldberg" type machine. The goal could be to translate one displacement into another (or one force to another) with particular inputs and outputs, but using as many types of simple machine as possible. Or it might be to use as few machines as possible. It is not yet clear whether this can be achieved with purely open-ended drawing or whether it would be more feasible to construct a solution from prototypical building blocks that can be stretched, flipped, scaled and positioned. In either case, there is no single right answer, but the ability to compare solutions to a generative grammar of compound machines and analogically compare kinematic pairs to teacher-authored prototypes should allow this exercise to be evaluated and scored.

EXPERIMENT DESIGN AND MATERIALS PREPARATION

In experimenting with the Simple Machines curriculum, we plan on using a two by two design. The first factor will be whether or not sketching is used, the second is whether or not GIFT’s adaptive tutoring capabilities are used. In the non-sketching conditions, additional examples presented via text and diagrams will be used to provide balance, to reduce time at task differences as being a source of confounds. Our qualitative predictions for these conditions are shown in Table 1.

Table 1. Qualitative Predictions for Student Learning

	No Sketching	Sketching
Non-Adaptive	Least learning	In between
Adaptive	In between	Most learning

We will measure learning by using a pre-test and post-test, both administered within the GIFT tutor, so that we can recruit participants on-line. We have created a bank of just over 90 questions, focusing on true/false and multiple choice questions for simplicity. The questions are drawn from open-license materials (e.g. the CK-12 Physical Science for Middle School textbook) or made up ourselves. We estimate that 20 questions for each test will provide enough statistical power to measure learning. We have already selected two sets of 20 questions, balanced in terms of difficulty by ensuring that for every question in the pre-test, there is a roughly equivalent, but not identical, question in the post-test. The pre/post tests will be identical for every participant. We will use different questions in the adaptive conditions from either the pre/post-tests.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This paper summarizes work in progress on integrating Sketch Worksheets with GIFT, to explore how sketching can be combined with adaptive tutoring to hopefully improve student learning better than either could alone. The Simple Machines curriculum we are developing as a testbed will cover each type of simple machine, and include recognition, analysis and synthesis activities, qualitative and quantitative concepts and relations, and includes parametric and structural synthesis tasks. The key to supporting these activities is the ability of CogSketch to permit open-ended sketch input and to extract meaningful relationships from that input. In particular, this allows a student to use annotations to show her work and justify answers – not just say what will happen, but also why. Another advantage of open-ended input over multiple choice is that it can permit vastly more possible answers than would be practical to enumerate explicitly, as in tasks such as unscrambling a shuffled machine or filling in gaps in a kinematic chain with missing elements.

Given where we are on this project, the conclusions and recommmendations we have only concern technology development, rather than tutor effectiveness. First, we suggest that finer-grained granularity on saving be supported, i.e. even when questions are incompletely filled out during authoring, and during long quizzes when taking a course. We recommend that future versions of GIFT consider introducing stronger relationships between test items, so that balanced pre/post tests can be automatically generated from a large question bank. We also recommend that development of the LTI interface continue, expanding as that protocol is fleshed out, to provide a richer channel between Sketch Worksheets (and other extensions) and GIFT.

REFERENCES

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science*, 333(6046), 1096-1097.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. (2011). CogSketch: Sketch understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*, 3(4), pp 648-666..

Forbus, K., Ferguson, R., Lovett, A., & Gentner, D. (2016) Extending SME to Handle Large-Scale Cognitive Modeling. *Cognitive Science*. DOI: 10.1111/cogs.12377, pp 1-50.

Forbus, K., Chang, M., McLure, M., and Usher, M. (2017) The Cognitive Science of Sketch Worksheets. Topics in *Cognitive Science*, DOI:10.1111/tops.12262.

Forbus, K.D., Garnier, B., Tikoff, B., Marko, W., Usher, M. & Mclure, M. (2018). Sketch Worksheets in STEM Classrooms: Two Deployments. Deployed Application Prize paper. *Proceedings of AAAI 2018*.

Garnier, B., Chang, M., Ormand, C., Matlen, B., Tikoff, B., & Shipley, T. (2017) Testing the efficacy of CogSketch geoscience worksheets as a spatial learning and sketching tool in introductory Geoscience courses. *Topics in Cognitive Science*

Jee, B., Gentner, D., Uttal, D., Sageman, B., Forbus, K., Manduca, C., Ormand, C., Shipley, T., and Tikoff, B. (2014). Drawing on experience: How domain knowledge is reflected in sketches of scientific structures and processes. *Research in Science Education*. 44(6), 859-883.

Scheiter, K., Schleinschok, K, & Ainsworth, S. (2017) Why Sketching May Aid Learning from Science Texts: Contrasting Sketching with Written Explanations. *Topics in Cognitive Science*, DOI: 10.1111/tops.12261.

ABOUT THE AUTHORS

***Kenneth D. Forbus** is the Walter P. Murphy Professor of Computer Science and Professor of Education at Northwestern University. His research interests include qualitative reasoning, analogical reasoning and learning, natural langauge understanding, sketch understanding, and cognitive architecture.*

***Thomas Hinrichs** is a Research Associate Professor of Computer Science at Northwestern University. His research interests include commonsense reasoning, cognitive architectures and machine learning.*

***Samuel Hill** is a graduate student at Northwestern University. His research interests include AI for assisting education and sketch understanding.*

***Madeline Usher** is a computer programmer / researcher at Northwestern University. Her research interests include sketch understanding and analogy-based tutoring.*

Iterative Development of the GIFT Wrap Authoring Tool

Fleet C. Davis[1], Jennifer M. Riley[2], and Benjamin S. Goldberg[3]
Humanproof LLC[1], Design Interactive Inc.[2], U.S. Army Research Laboratory[3]

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012; Sottilare, Brawner, Sinatra & Johnston, 2017) is an empirically-based, service-oriented framework of tools, methods, and standards aimed at overcoming the challenges associated with authoring computer-based tutoring systems (CBTS), managing instruction and assessing the effect of CBTS, components and methodologies ("Generalized Intelligent," n.d.). One of the primary developmental objectives for GIFT is the creation of an integrated, user-friendly authoring experience that can be used across training applications. Humanproof, with teammate Design Interactive, is working to fulfill this objective via continued development of the GIFT Wrap prototype. This prototype, currently in its third generation, allows training developers to configure the real-time, automated delivery of instructional content triggered by assessing state changes within the training application's environment (e.g., entity location) and/or learner (e.g., concept mastery). This ongoing research and development effort is focused on the design and implementation of the user interface (UI) that guides users through the configuration of tutoring events driven by real-time assessments within a training application. Integration with the LandNavHD Unity game, a computer-based land navigation trainer used as a practice environment for dead reckoning procedures, served as the most recent use case for this ongoing effort. The third generation of GIFT Wrap's development focused on building new integrated, user-friendly tools for authoring real-time assessments within the context LandNavHD training environment. This effort also included the continued integration of legacy authoring functionality into the GIFT Wrap design. The following sections briefly describe the previous GIFT Wrap development efforts, provide an overview of the third generation of GIFT Wrap, present usability findings, and discuss concepts for extending GIFT Wrap to live training environments.

BACKGROUND

From a conceptual level, GIFT manages interaction within a training environment through the Learning Effect Model (LEM; Sottilare, Ragusa, Hoffman & Goldberg, 2013). The LEM outlines the inference processes captured in GIFT that leads to the selection of an instructional strategy based on observed performance. In this model, raw data is consumed by GIFT and routed to the domain module for assessment purposes. In this instance, the domain module uses the raw data to compute a performance state on a set of defined concepts, where Condition Classes designate performance as at-, above-, and below-expectation for the associated concept being assessed. This performance state is combined with learner relevant information (i.e., individual differences) to inform the pedagogical model for a strategy selection. The challenge here is establishing the necessary assessments required to capture appropriate performance states that associate with the objectives of the training event. To meet this challenge, user-centered design approaches are being applied to current architectural components with the intent of providing training developers and subject matter experts with intuitive tools to configure these assessments themselves.

Authoring Challenges - Real-time Assessments

In previous versions of GIFT, there were two major challenges for users authoring the real-time assessment component of a course. First, authoring the Domain Knowledge File (DKF) using the DKF Authoring Tool

(DAT) proved to be too complex for the average user and much better suited for power users that would be more likely to take full advantage of the DKF's extensive functionality. Second, users were required to author using both the DAT as well as any content creation tools for the training application (e.g., the Virtual Battlespace mission editor) in order to configure real-time assessments and other elements of adaptive training. Without real-time communication between GIFT and the training application, direct integration was not possible, making the authoring experience disjointed and cumbersome for users (Davis, Riley, & Goldberg, 2017).

Overcoming Authoring Challenges

GIFT Wrap was purposely designed to overcome the challenges associated with authoring real-time assessments by providing users with an integrated, user-friendly authoring tool. The first generation of GIFT Wrap took an initial step towards addressing integration with training applications by providing users with a tool that allowed them to author tutoring content (i.e., a check on learning (COL)) while simultaneously interacting with the training application's content creation tools (i.e., the Augmented Reality Sandtable (ARES) terrain map) (Hoffman, Markuck, & Goldberg, 2016).

The first generation of GIFT Wrap served as proof-of-concept that led to the development of the second generation. The second generation of GIFT Wrap advanced the tool's functionality by (1) providing a redesigned UI for creating, configuring, and managing a DKF that would eventually replace the DAT, and (2) creating a "blended authoring environment" that allowed users to author real-time assessments (e.g., COLs) directly within the context of a training application's content creation tools via an "Overlay UI" with the flexibility to rapidly switch back to the main GIFT Wrap UI and configure the rest of the DKF (Davis, et al., 2017).

THIRD GENERATION GIFT WRAP

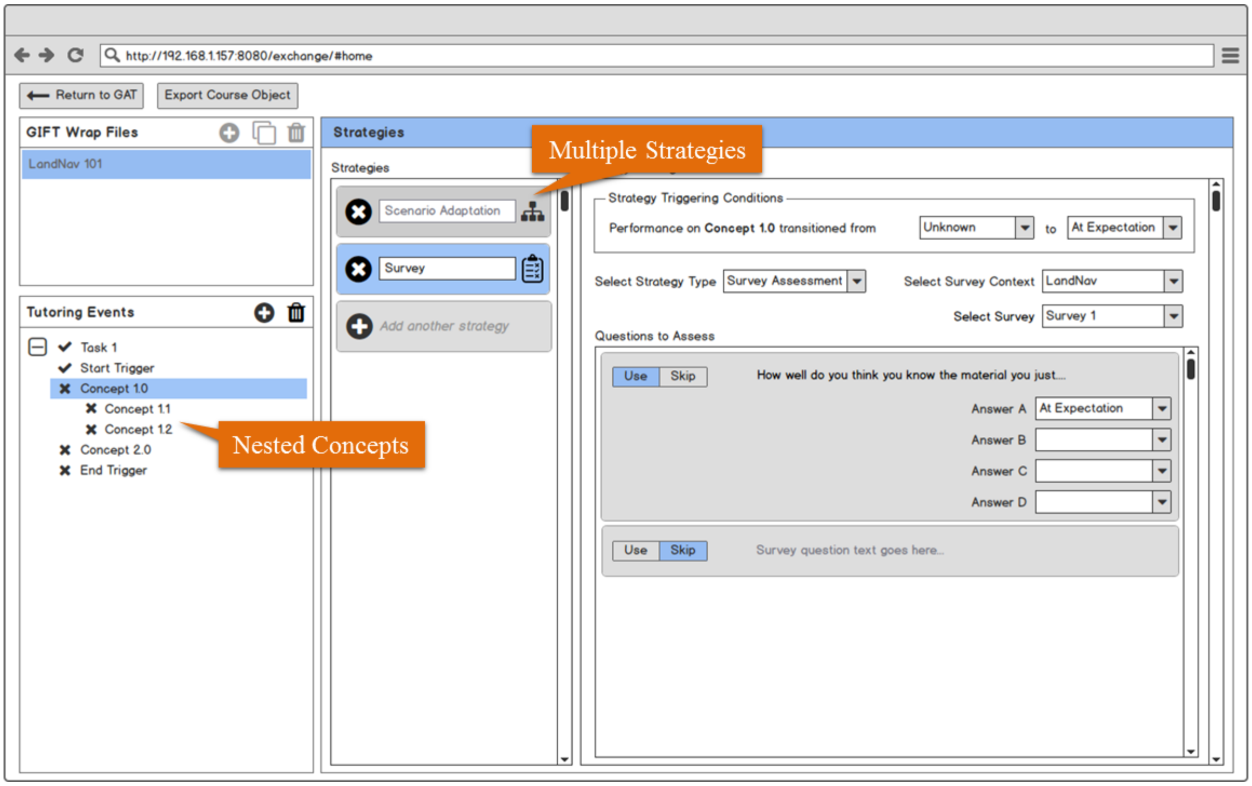
Incorporating DAT Functionality

The second generation of GIFT Wrap was designed to be flexible enough to incorporate all existing DAT functionality into a new, more user-friendly UI that could support both novice GIFT users as well as more experienced GIFT training authors (Davis, et al., 2017). The third generation of GIFT Wrap contains several new features (see Figure 1) that previously only existed in the DAT including, but not limited to, the following:

- Users may now create child Concepts nested up to three layers deep allowing training developers the flexibility to assess Concepts at different levels of granularity.

- User may now create multiple strategies for state transitions and/or assessment levels for a given Condition Class.

- Users may now add time delays for Task triggers to better control the pace and timing of tutoring events.



Performance on **Concept 1.0** transitioned from Unknown to At Expectation

Select Strategy Type: Survey Assessment

Select Survey Context: LandNav

Select Survey: Survey 1

How well do you think you know the material you just...

Answer A: At Expectation

Answer B:

Answer C:

Answer D:

Survey question text goes here...

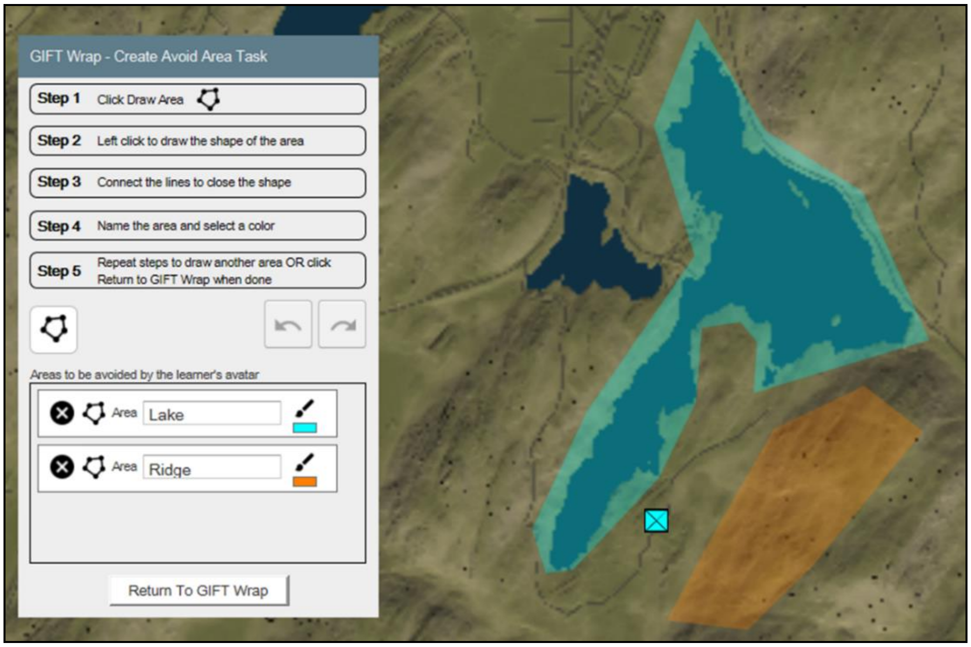
Figure 17. GIFT Wrap New Features

Extending the Blended Authoring Experience

Beyond incorporating additional DAT functionality into the new GIFT Wrap design, the blended authoring experience was extended outside the ARES training application to include the LandNavHD Unity game. In order to accomplish this, GIFT Wrap was integrated with the GIFT Unity plugin to establish communication between GIFT Wrap and the LandNavHD. Also, two new event handlers were created in the LandNavHD Unity project that send messages to GIFT providing information used for real-time assessment. Once GIFT Wrap and the LandNavHD were fully integrated, new real-time assessments were created specifically for the LandNavHD. Carrying forward the land navigation training use case used with the second generation of GIFT Wrap, the following Condition Classes were created to support the training tasks used in the LandNavHD: Avoid Area, Follow Path, and Locate Navigation Points. Next, the GIFT Wrap Overlay UI was updated to accommodate authoring these new real-time assessments within context of the LandNavHD environment. The current version of the LandNavHD does not include content creation tools that would allow users to create or edit new scenarios. To account for this, a top-down image of the terrain was extracted and a new layer was created in the GIFT Wrap UI to simulate the functionality of authoring within the training application's virtual environment. Each of the new LandNavHD real-time assessments and corresponding Overlay UIs are described below.

Avoid Area

This Condition Class checks whether or not a specific entity avoided an area in the virtual environment. This is used to assess the learner's ability to move by terrain association and/or dead reckoning while avoiding certain obstacles, areas, terrain features, etc. GIFT Wrap allows users to easily draw areas to avoid directly on the LandNavHD terrain (see Figure 2) rather than requiring manual entry of a set of coordinates. Users may also adjust the positioning of the area, name it, change its color, and set a tolerance (e.g., entity entered area for more than 30 seconds). While this assessment was created for land navigation, it is generalizable to numerous scenarios relating to zones of interest and trainee location within that interacting space.



GIFT Wrap - Create Avoid Area Task

Step 1 Click Draw Area

Step 2 Left click to draw the shape of the area

Step 3 Connect the lines to close the shape

Step 4 Name the area and select a color

Step 5 Repeat steps to draw another area OR click Return to GIFT Wrap when done

Areas to be avoided by the learner's avatar

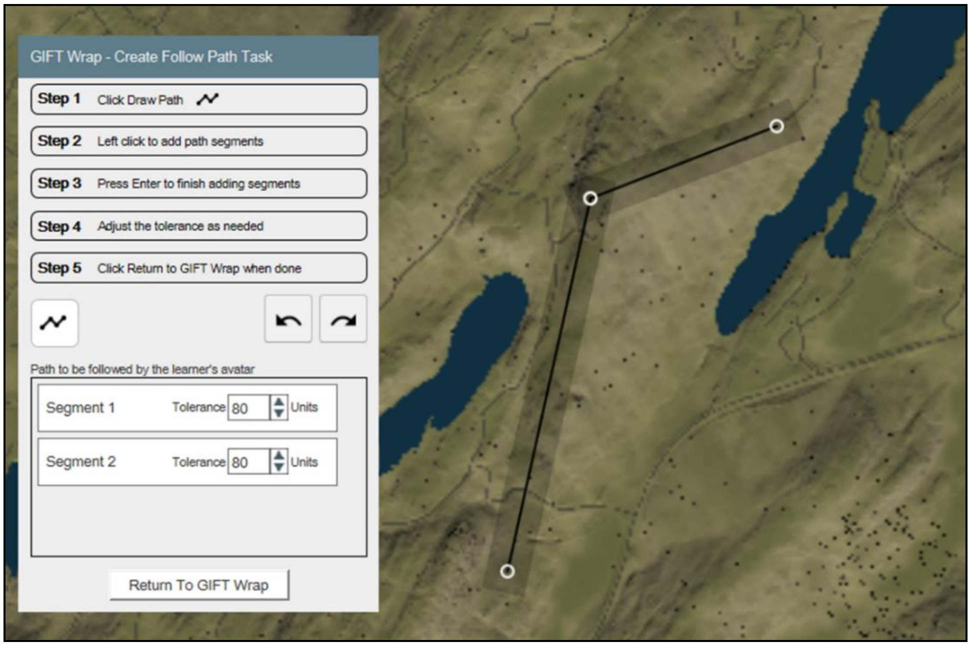
Area Lake

Area Ridge

Figure 18. Avoid Area Overlay UI

Follow Path

This Condition Class checks whether an entity traveled along a series of connected straight line paths in the virtual environment within a set of thresholds for deviation. This is used to assess a learner's ability to move by dead reckoning, point-to-point land navigation. GIFT Wrap allows users to easily draw paths/routes to follow directly on the LandNavHD terrain (see Figure 3) rather than requiring manual entry of a set of coordinates. Users may also adjust the positioning of the end points and set a tolerance (e.g., entity may deviate no more than 30 meters from the path).



GIFT Wrap - Create Follow Path Task

Step 1 Click Draw Path

Step 2 Left click to add path segments

Step 3 Press Enter to finish adding segments

Step 4 Adjust the tolerance as needed

Step 5 Click Return to GIFT Wrap when done

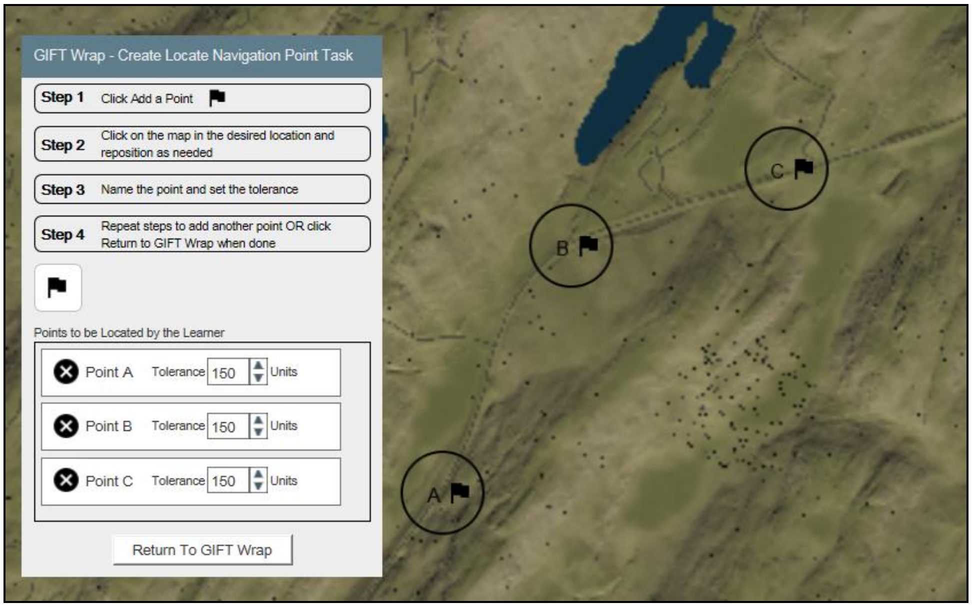
Path to be followed by the learner's avatar

Segment 1	Tolerance	80	Units
Segment 2	Tolerance	80	Units

Figure 19. Follow Path Overlay UI

Locate Navigation Points

This Condition Class checks whether or not an entity reached the location of a specific location (coordinate) in the virtual environment within a set threshold. This is used to assess the learner's ability to navigate to specified locations in the virtual environment. GIFT Wrap allows users to easily drop points directly on the LandNavHD terrain (see Figure 4) rather than requiring manual entry of a set of coordinates. Users may also adjust the positioning of the point and set a tolerance (e.g., entity must be within 30 meters of the point).



GIFT Wrap - Create Locate Navigation Point Task

Step 1 Click Add a Point

Step 2 Click on the map in the desired location and reposition as needed

Step 3 Name the point and set the tolerance

Step 4 Repeat steps to add another point OR click Return to GIFT Wrap when done

Points to be Located by the Learner

Point A	Tolerance	150	Units
Point B	Tolerance	150	Units
Point C	Tolerance	150	Units

Figure 20. Locate Navigation Point Overlay UI

VALIDATING THE DESIGN

The third generation of GIFT Wrap represents the most recent attempt to develop user-friendly authoring tools aimed at configuring real-time assessments that occur during training. However, user testing is always needed to validate claims that the most recent design iteration is indeed an improvement over previous versions. Therefore, a small scale usability test was conducted to compare and contrast authoring a DKF using the DAT and the third generation of GIFT Wrap. A total of seven of participants were asked to complete a comparable set of tasks with both interfaces, in a counter-balanced manner, in order to gather user feedback on their perceived ease of use as well as compare system performance. The results (i.e., descriptive statistics) from each survey and performance measure, findings from the user interviews, and test facilitators' observations are reported in the following sections.

Subjective Measures

Subjective Workload

All participants reported experiencing higher workload with the DAT (*M* = 62.71, *SD* = 8.34) than with GIFT Wrap (*M* = 37.86, *SD* = 9.21) on the NASA-Task Load Index (NASA-TLX) (Hart & Staveland, 1988) (see Figure 5). The subscales that appear to have contributed the most to differences in the overall score were Mental Demand, Performance, and Frustration (see Figure 6). That is, the participants reported higher Mental Demand and Frustration and poorer Performance associated with the DAT than GIFT Wrap.

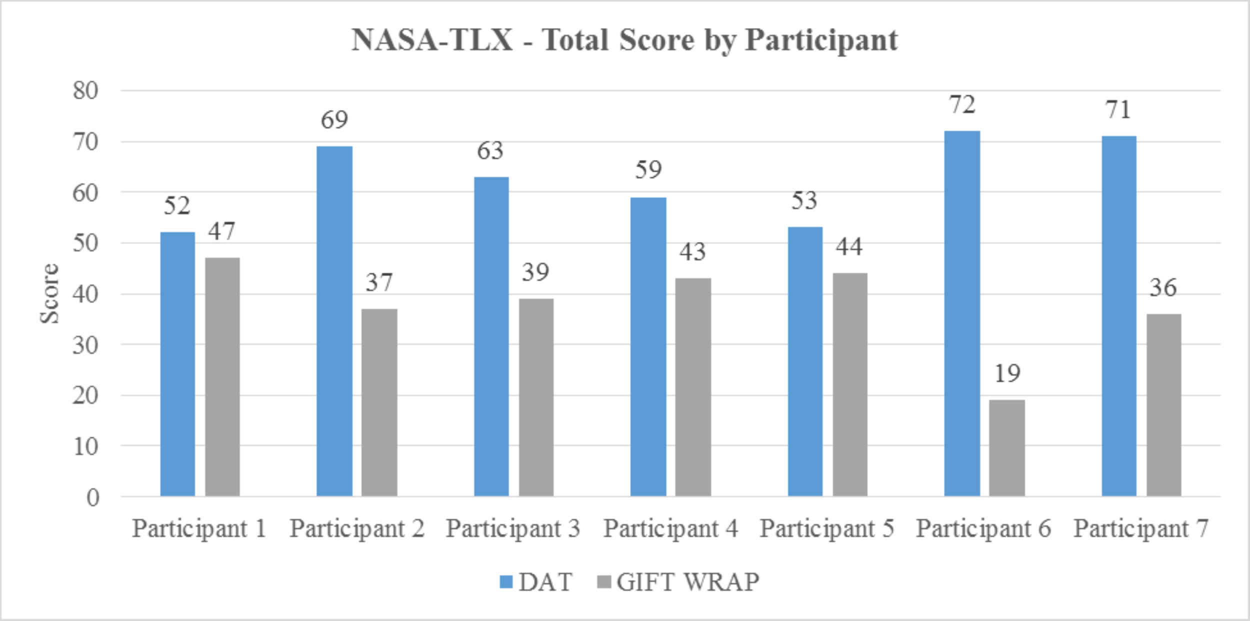


Figure 21. NASA-TLX Total Scores by Participant by Tool

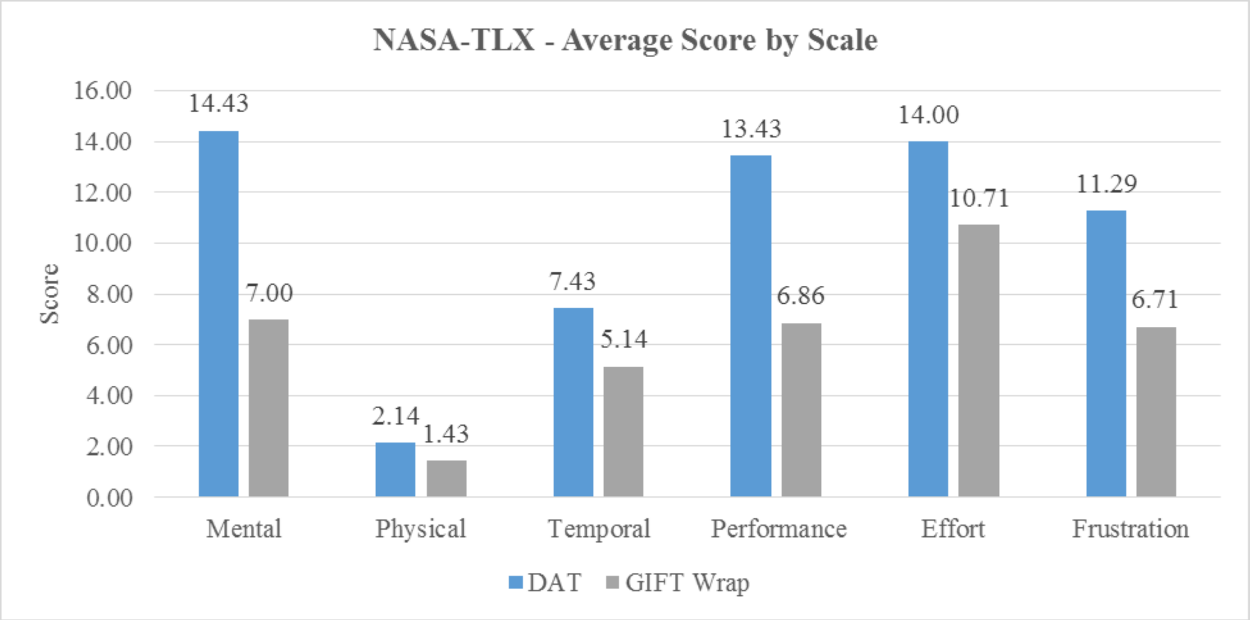


Figure 22. Average Score by Scale

System Usability Scale

All but one participant reported better perceived usability for GIFT Wrap (*M* = 67.86, *SD* = 17.76) than for the DAT (*M* = 36.79, *SD* = 24.01) on the System Usability Scale (SUS) (Brooke, 1996) (see Figure 7). In a review of 500 studies, a score of 68 was found to be the SUS national average (Sauro, 2011). GIFT Wrap received a score roughly equivalent to C while the DAT received a score equivalent to an F.

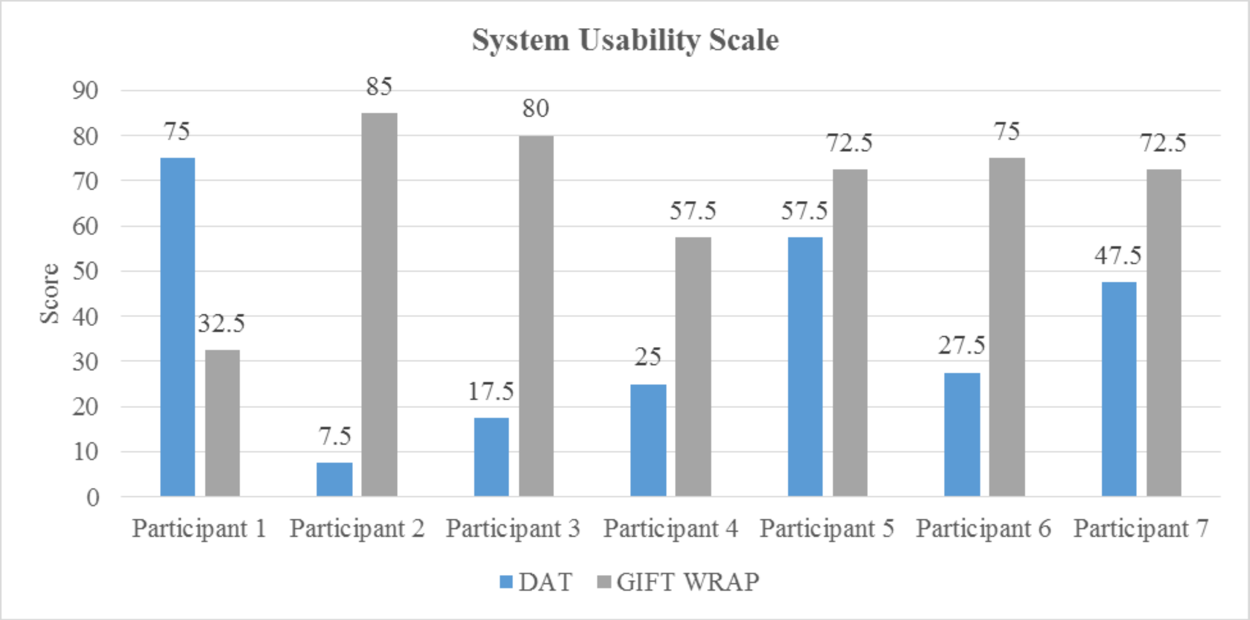


Figure 23. SUS Scores by Participant by Tool

Objective Performance Measures

All participants required more time to complete the test tasks with the DAT (*M* = 1309.00 (21min 49s), *SD* = 353.92) than with GIFT Wrap (*M* = 592.00 (9min 52s), *SD* = 89.74) (see Figure 8). Furthermore, participants required more prompting to complete the test tasks with the DAT (*M* = 16.00, *SD* = 7.02) than with GIFT Wrap (*M* = 5.71, *SD* = 2.75) (see Figure 9).

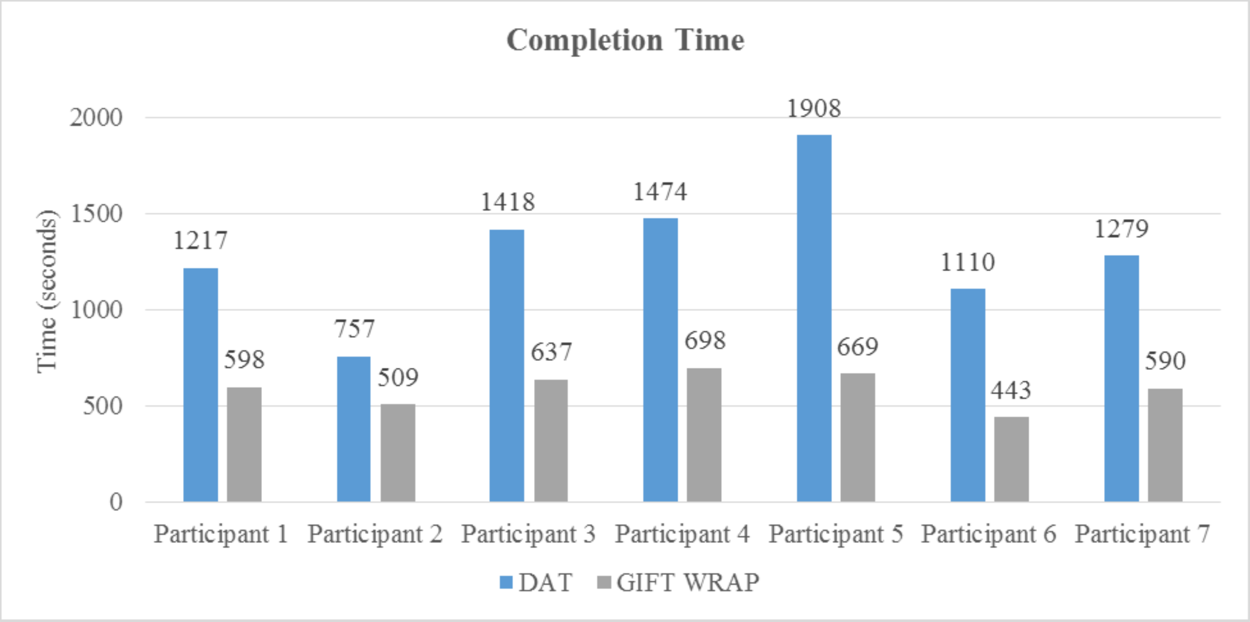


Figure 24. Completion Times by Participant by Tool

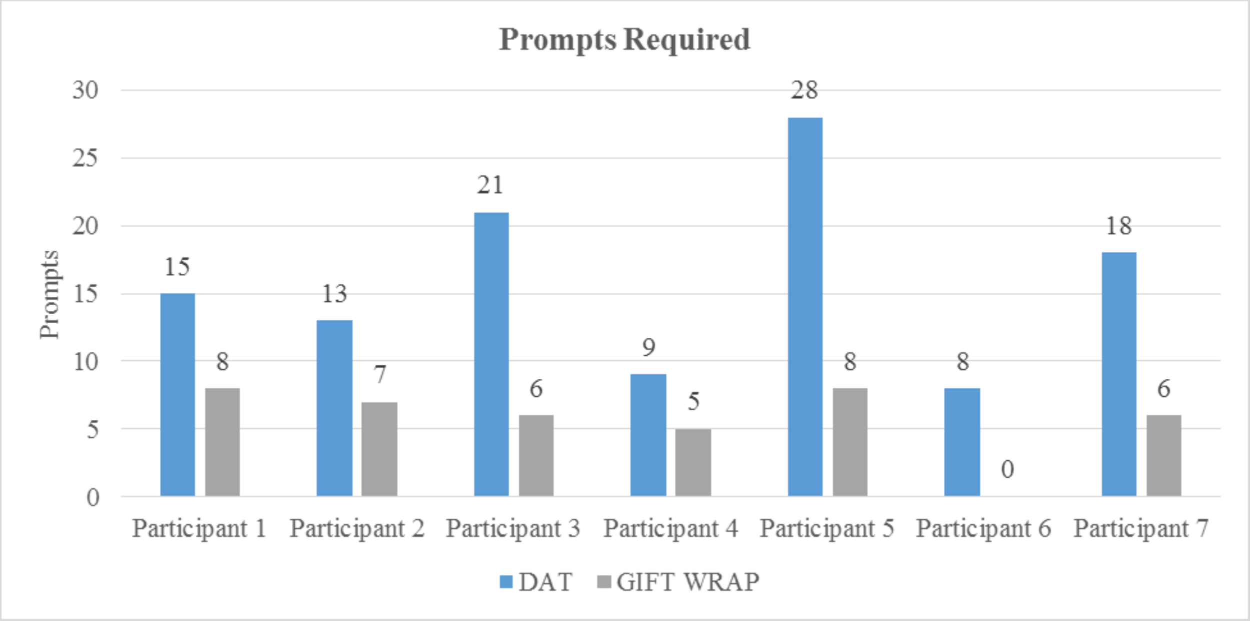


Figure 25. Prompt Count by Participant by Tool

Participant Feedback & Other Observations

Table 1 below summarizes the participant feedback collected immediately following each test session as well as other observations captured by the test facilitators during the usability testing.

Table 2. Participant Feedback & Other Observations

	GIFT Wrap	DAT
Common Usability Issues	• Determining how to add a new concept • Remembering to complete the end trigger • Determining how to rename items (e.g., Concepts) • Recognizing horizontal panels/tabs (e.g., Strategy panel)	• Save and exit errors (i.e., accidental close out of DAT with the intent of saving) • Determining how to set-up and assign waypoints • Determining how to set-up and complete strategies and/or state transitions • Determining how to add sub-concepts • Confusion about end trigger at start of authoring a task, prompted with need to return to it later
Users Liked *Best* about the Tool	• Layout • Intuitiveness, Simplicity • Process flow (i.e., tree menu structure) • Only relevant info presented to user	• More features and options apparent • Descriptive (e.g., tool-tip-text, instructions) • UI “Style” (e.g., colors)
Users Liked *Least* about the Tool	• Fewer instructions at interface • Fewer apparent options	• Confusing, Not intuitive • Frustrating flow • Not user friendly, hard for soldiers to use • Lots of clutter and/or information on interface

Taken together, the results of this usability test indicate that users perceive GIFT Wrap to require less effort and to be more user friendly than the DAT, legacy GIFT authoring tool. Furthermore, the participants were able to complete the tasks much quicker and with less assistance with GIFT Wrap than the DAT.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The third generation of GIFT Wrap successfully incorporated additional DAT functionality into a more user-friendly design and extended GIFT's authoring capabilities to a new training application with the LandNavHD Unity game. Furthermore, usability testing demonstrated that GIFT Wrap is much more user-friendly than legacy authoring tools making GIFT more accessible to the average user without eliminating the important features power users need. However, while GIFT Wrap's design outscored and outperformed the DAT, the test results showed that many design features could be improved. Future developers of GIFT Wrap should take these findings into account as they strive to iteratively improve the design.

GIFT Wrap is now capable of supporting the authoring of land navigation training across multiple training applications (i.e., ARES, LandNavHD). These authoring tools and real-time assessment capabilities are easily extendable to new applications including training in live environments via integration with mobile devices. Efforts are currently underway to determine the "back-end" functionality necessary for GIFT to communicate with mobile devices to retrieve real-time assessment data and to push instructional interventions to learners via a mobile tutor UI. This initial proof-of-concept will aim to layer GIFT's tutoring capabilities on top of an existing live terrain walk exercises conducted at the United States Military Academy at West Point.

The lessons learned from the first three generations of GIFT Wrap will be used to inform and guide the development of the fourth generation of GIFT Wrap. Near term GIFT Wrap research and development efforts will focus on developing new, user-friendly authoring capabilities that will be integrated with web mapping services (e.g., Google Maps) to create a new authoring layer. Work will also be done to apply existing capabilities to this new environment and to develop authoring tools for terrain walk specific real-time assessments (e.g., pace count, plotting routes). This fourth generation of GIFT Wrap will eventually provide training developers with the tools they need to easily create land navigation training using the GIFT ITS to scaffold the learner's phased skill development across three complimentary training environments

REFERENCES

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4-7.

Davis, F., Riley, J.M., & & Goldberg, B. (2017, July). Development of an Integrated, User-Friendly Authoring Tool for Intelligent Tutoring Systems. In *Proceedings of the Fifth Annual GIFT Users Symposium (GIFTSym5), Orlando, Florida.*

Generalized Intelligent Framework for Tutoring (GIFT). (n.d.). Retrieved from https://gifttutoring.org/projects/gift/wiki/Overview

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Hoffman, M., Markuck, C., & Goldberg, B. (2016, July). Using GIFT Wrap to Author Domain Assessment Models with Native Training Applications. In *Proceedings of the Fourth Annual GIFT Users Symposium (GIFTSym4), Orlando, Florida.*

Sauro, J. (2011). A practical guide to the system usability scale: *Background, Benchmarks & Best Practices.* Denver, CO: Measuring Usability LLC.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Concept paper released as part of GIFT software documentation. Orlando, FL: US Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED). Retrieved from: https://gifttutoring.org/attachments/152/GIFTDescription_0.pdf

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: US Army Research Laboratory. May 2017. DOI: 10.13140/RG.2.2.12941.54244.

Sottilare, R., Ragusa, C., Hoffman, M., & Goldberg, B. (2013, December). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In *Proceedings of the Interservice/Industry Training Simulation & Education Conference (I/ITSEC), Orlando, Florida.*

ABOUT THE AUTHORS

***Mr. Fleet Davis** is a Senior Human Factors Engineer at Humanproof, LLC. He is the Principal Investigator for the GIFT Wrap project.*

***Dr. Jennifer Riley** is the Performance Augmentation Division Head at Design Interactive, Inc. She is the Co-Principal investigator for the GIFT Wrap project.*

***Dr. Benjamin Goldberg** is an adaptive training scientist at the Army Research Laboratory (ARL) Human Research and Engineering Directorate (HRED). He leads research focused on instructional management within the Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of GIFT.*

THEME III: INDIVIDUAL LEARNER MODELING

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 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 experience 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 psychomotor 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 highlights 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 motivate.

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 pedagogical 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 re-administer 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 remediation 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 remediation 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, Niehaus, 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-based 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-adaptive 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). How- ever, 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 do- mains, 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 assessments. *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.*

Workload-Adaptive Training Scenarios for Synthetic Training Environments

Nathan D. Smith[1], Ezekiel D. Gunnink[1], Thomas Schnell[1], Christopher Reuter[1], Jason D. Moss[2]
Operator Performance Laboratory (OPL)[1], U.S. Army Research Laboratory (ARL)[2]

INTRODUCTION

The Army is working toward enhancements in soldier training systems using synthetic training environments (STEs) and mission rehearsal capabilities. This capability will augment live training on the range and in many cases will be self-guided, meaning that the trainee will not require a tutor or instructor to administer the training. The STE enables single user training and training of teams in small local groups and across operational networks involving large groups. Adapting the training scenarios to the capabilities and training needs of individual trainees is a proven way to enhance individual training effectiveness.

In distributed training evolutions, proper adaptation of the training scenarios takes on an even more important role, as such exercises involve many trainees at various stages of their training maturity and skill. Problems arise when less experienced, or lower-skilled, trainees are exposed to training scenarios that are too advanced, or complex, for their level of experience. This can easily happen if the STE does not consider the capabilities and limitations of the individual trainees. Such unprepared trainees are more likely to fail their training mission and thereby reduce the benefits of training, further exacerbating frustration in the trainee. In addition, the failure may jeopardize the success of other trainees who depended on a reasonably successful outcome of a mission task element in the scenario. Failure of a single trainee to accomplish his/her mission may result in a chain reaction of adverse events in the training evolution that may reduce the value of the training exercise or increase cost. Conversely, trainees exposed to missions that are not sufficiently challenging may experience boredom, or even apathy, resulting in a negative training benefit. Adaptive scenario administration is needed in STEs to avoid such breakdowns and to enhance individual training effectiveness.

There are many STEs and tools available such as VBS 3 (Virtual Battle Space 3). These tools often allow the creation and storing of scenarios that contain the starting conditions of a training module but the scenarios themselves are usually administered on a brute-force lesson plan. The structure of the simulation tools actually encourage such lesson based administration as it is very easy to create and save static scenarios. What is needed, however, is a mechanism to continually adapt the scenarios to match them to student abilities at their respective stages in the training program. Additionally, students need rich feedback on their performance and guidance on ways to modify behaviors to increase performance, if it is not at or above expectation.

In our work, we have created an adaptive training framework from three separate systems, (1) the Generalized Intelligent Framework for Tutoring (GIFT), (2) the VBS 3 simulation framework, and (3) the Cognitive Assessment Tool Set (CATS) workload quantification library. We developed a generic method to incorporate the GIFT performance grading scheme into VBS 3. This allows for on-the-fly configuration of adaptive VBS 3 training scenarios. Additionally, through CATS, this script can take into consideration the workload exhibited by the trainee and adapt the scenario to avoid over or underload conditions. This adaptive training framework is governed by student performance, workload, and task difficulty. Performance and workload were incorporated as aggregated scores. Workload is assessed using the CATS workload library that is attached to GIFT. Both performance and workload drive the selection of upcoming training scenarios by modulating task difficulty such that the trainee is challenged at an optimum level. A script in VBS 3 uses a decision tree on the basis of performance (below, at, above expectation) to manipulate the level of task difficulty to maximize training effectiveness.

In the current development cycle, we will perform a human factors study to assess the efficacy of adaptive training using two distinct adaptation schemes, one based on performance only, and one based on a combination of performance and workload. The results of the study will be used to determine if workload-adaptive training scenarios are more effective than training scenarios that only consider performance.

BACKGROUND

The value of adaptive training and its positive effect on training effectiveness has been well documented. The idea is of course not new (Lintern G. & Gopher D., 1978). The underlying principle is based on two hypotheses, (1) the learning of a complex task is best accomplished using less difficult versions of the task and increasing levels of difficulty until the whole task can be mastered, (2) learning of a task is better when transition from one level of difficulty to the next is guided by the performance of the student rather than brute-force administration of a rigid training regimen.

In one-on-one training settings, expert instructors use this principle almost instinctively to keep students motivated throughout the building of critical skills. For example, flight instructors may teach the difficult skill of auto-rotating a helicopter using increasing levels of difficulty by gradually increasing the complexity of the maneuver. In the example of autorotation, adaptation is not only representative of good training didactics, it is essential for survival of both the instructor and the student, as poor performance can lead to mechanical damage to an expensive helicopter, such as through over speeding the rotor system, or it could lead to a fatal crash such as allowing the rotor RPM to drop below an allowable minimum or initiating the landing flare too late. Control of this task requires manipulation of four inter-dependent controls (collective, lateral cyclic, longitudinal cyclic, and tail rotor pedals) as well as at least four inter-dependent performance parameters (airspeed, flight path, rotor RPM, aircraft attitude). To an uninitiated person, this maneuver is extremely scary and cognitive workload will be very high. It makes no sense to scare a student on each and every repetition of that maneuver as this will only increase the possibility that the student will never master it and be unable to use it as a needed emergency skill.

Expert instructors will ease their students into auto-rotations through adaptive training principles by giving the student only one control axis at a time (e.g. the collective) or through adjustment of the flight path (straight in path instead of curved). As the student gains confidence in his/her ability to master this skill at a given difficulty level, performance will improve and workload will go down. As is typical in the acquisition of many critically important skills, the decrease in workload is highly indicative of autonomous mastery. In the early stages, students may be able to master the skill at an acceptable technical level but only with the highest levels of cognitive workload expenditure. This is usually sufficient for passing a check-ride or to graduate with a certificate but it is hardly a proper level of training for critical skills in warfighters. Instructors and instructional systems owe it to the warfighter to train them to a higher standard. High levels of cognitive demand causes significant draw on limited attentional resources (see Figure 26) which adversely affects the performance of perception, memory, decision making, and response execution. Trainees who master the skill to a point of automaticity will expend less cognitive workload and thus retain more attention resource capacity. This will afford them to devote those resources to mission critical task elements, which is essential in the projection of military power and for the self-protection of the warfighter.

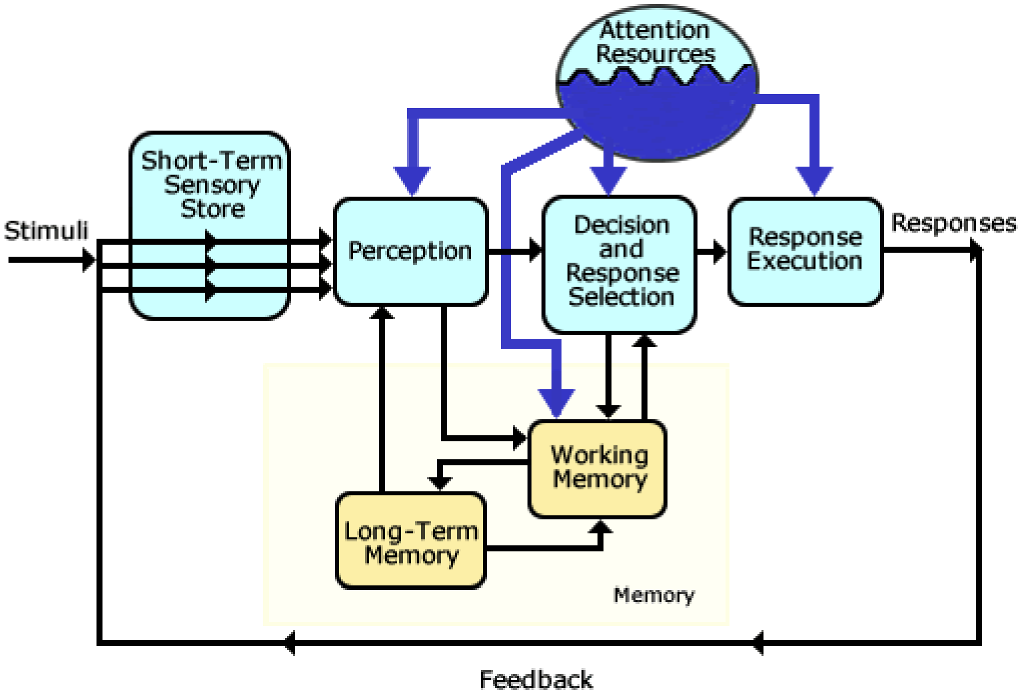


Figure 26. Wickens Information Processing Model (Wickens C. D., 1992, 2008)

In our research, we are working on building the adaptive expertise that good instructors apply almost instinctively into automated Synthetic Training Environments (STEs).

Even as far back as the Seventies, adaptive training was conceived of as a closed-loop controller system (Lintern G. & Gopher D., 1978). Such a system depends on measurement of task performance. Unfortunately, automatically generating performance measures is not always easy and in many training tasks has eluded us to this day. Additionally, the most optimal way to make the training scenario adaptive is not always easily evident. Much research has been devoted to the question of how to make training tasks adaptive. Part-whole training is an adaptation scheme where essential subtasks are learned as building blocks to enable mastery of the whole task. Part Task (PT) training was found to lead to significantly faster convergence of a tactical skill in a video game when compared to Full Task (FT) training (Mané, Adams, & Donchin, 1989). An interesting observation of their work is that the part tasks were not fully representative as fractions of the whole task but when learned in sequence lead to better performance than if the full task is learned at once. An additional observation is that the PT training took longer than FT training. However, the skill transfer rates from the PT were 100% and the overall performance of demonstrating the full task was much better. Thus, while PT may not yield net time savings, the fact that better performance is achieved may mean that less remedial training will be needed.

Mane and Wickens (Mane A. & Wickens C., 1986) studied the effects of task difficulty and workload on training. They noted that training systems should adapt to maintain high levels of workload as otherwise, trainees will learn short-term resource preserving strategies that are counterproductive toward mastering of the long-term skill. Rigid (i.e. non-adaptive) training methods allow such maladaptive resource preservation strategies to take hold. In our work, we use real-time measures of cognitive workload to quickly close that short-term loophole for the trainee by adjusting training difficulty to maintain high levels of workload while at the same time preventing overload or defeat of the student through scenarios that are too difficult (e.g. auto-rotations).

Gerjets et al. (Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014) describe the relationship between cognitive load theory (CLT) and training outcomes through optimal loading of working memory load (WML). The main challenge is a continuous classification of cognitive workload to allow adaptation of the training scenario to modulate WML. They describe methods such as subjective probes or secondary tasks measures. Both methods of workload estimation are disruptive and hinder training effectiveness. They used EEG as a means to estimate workload with some success. The use of EEG signals for classification of workload is well represented in the literature, two additional examples of which are presented here. Wilson and Russell (Wilson & Russell, 2003) attempted to classify workload using a combination of sensors, including six channels of brain electrical activity, eye, heart, and respiration measures. Those authors were able to achieve classification accuracies around 82%. However, their tasks consisted of only two variants of the same test. Additionally, the high number of sensors used to collect the data, is sub-optimal for many scenarios including in flight measurements. Matthews *et al* (Matthews et al., 2008) used a wireless EEG sensor helmet to classify workload in real-time. Those authors achieve classification accuracies on an average of 80.5%. In well over a decade of workload estimation research at the University of Iowa Operator Performance Laboratory (OPL), we have come to the conclusion that the technical readiness level and diagnostic capabilities of EEG based workload probes is very low and unsuitable for a real-world training environment outside a highly controlled laboratory.

A much simpler sensor montage is possible through a three-lead electrocardiogram (ECG). At OPL, we have used discrete deterministic nonlinear models of the full ECG waveform to obtain reliable and highly diagnostic real-time measure of cognitive workload. It is important to note that our method of ECG based workload estimation is NOT a heart rate based method or a time-series based analysis. Rather, we continually transform the entire ECG signal into an embedded phase space and classify workload on the basis of the dynamic representation of the heart though an ergodicity map of the electrical heart signal. We start with the realization that the heart is a chaotic system that is under control of the nervous system. Chaotic systems are often not well represented via the normal scalar time series. Instead, the dynamics of the system are obfuscated in the single dimension whereas they become apparent when a transform of the data is made. This transform moves the data from the single dimensional scalar space into a multi-dimensional embedded phase space (Richter & Schreiber, 1998). In our method, the ECG time series data is transformed into phase space using the CATS software tool (OPL, 2014). This step established the Ergodicity Transition Matrices (ETMs) (Engler & Schnell, 2013) that represented the dynamics of the ECG signal in phase space for the different workload conditions. To generate a real-time workload estimation, we can either use the ETMs directly through lookup of model ETMs using nearest neighbor classifiers or through models of statistical transitions within the ETM called the Transition Probability Variance (TPV). TPV calculates the variance of the probabilities of transition from one cell to another different cell of the course-grained ETM. The TPV therefore captures the variability in the dynamics of the ECG signal as the trainee undergoes different levels of cognitive loading. TPV varies inversely to the degree of workload with higher TPV numbers seen under low workload conditions and low TPV numbers seen under high workload conditions. The benefit of the direct ETM based discrete classifier is its very high accuracy level (near 100%). The downside of this method is that model ETMs need to be established for each participant and each desired level of workload. The TPV method is less accurate (around 85-90% classification accuracy) but it does not require a model. The TPV method provides a continuous measure of workload no more than three heartbeats after the ECG system has been turned on. The TPV system has excellent cross-person and cross-task validity and is easily deployed in complex real-world environments (Schnell T et al., 2017; Schnell T., Hoke J., & Romeas T., 2017; Schnell T., Reichlen C., & C., 2017).

Another trainee specific dimension that may be applied in the context of adaptive training systems is that of trainee affect and engagement. As with performance based adaptation, expert instructors generate a motivating and interesting training experience and they have the ability to detect affectual cues from the trainee such as frustration, fear, boredom, or anger. Effective instructors can interpret affectual cues as levers that affect learning. The affective domain of training provides a framework for instruction that includes student awareness, response, value perception, organization, and integration (FAA, 2008). A trainee has to be aware of the material being taught. It is the responsibility of the instructor or instructional system to raise the awareness level in the trainee through immersive and interesting content. The student responds through active participation, decides on the value of the training, organizes the training into his/her belief system, and finally, internalizes it. Motivation and enthusiasm are important enabling components of the affect domain.

Ocumpaugh et al. (2017) provided a thorough review of the role of emotions in training. A quantitative understanding of affect dynamics allows not only for an understanding of a learner's current affective state but also enables prediction of future affective states. Ocumpaugh et al. leverage data of the trainee's affect dynamics toward making better adaptive training transitions. A proposed approach for incorporating affective state assessment into the GIFT training system draws from the observed model of affect dynamics presented by D'Mello and Graesser (D'Mello & Graesser, 2012).

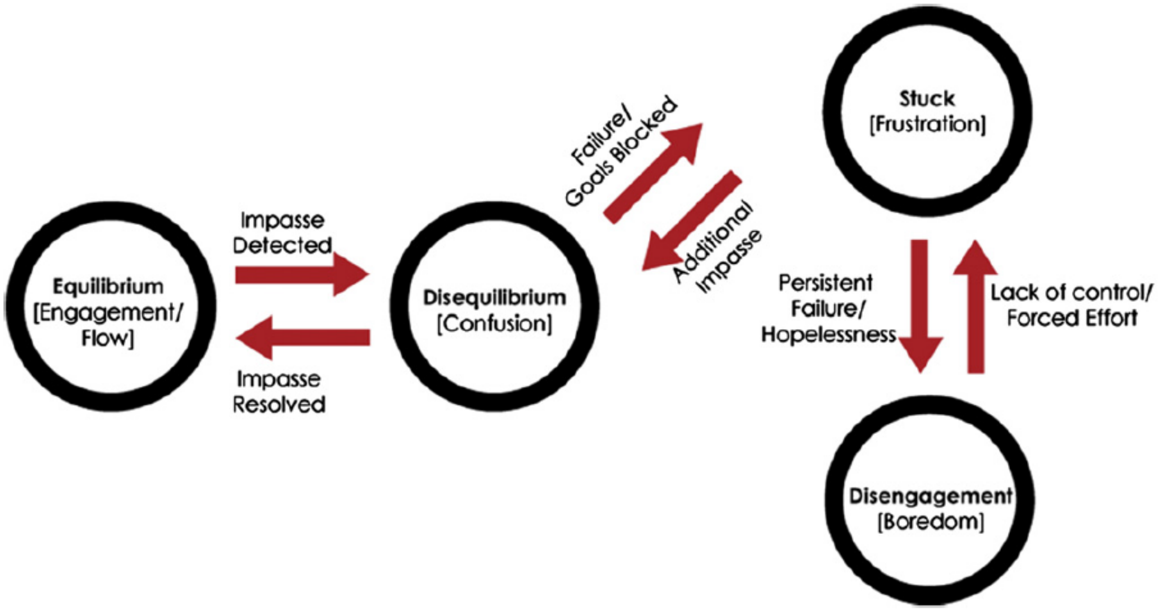

Equilibrium [Engagement/ Flow]
Impasse Detected
Impasse Resolved
Disequilibrium [Confusion]
Failure/ Goals Blocked
Additional Impasse
Stuck [Frustration]
Persistent Failure/ Hopelessness
Lack of control/ Forced Effort
Disengagement [Boredom]

Figure 27. D'Mello & Graesser Model of Affect Dynamics

We identify the following states and definitions from the referenced work

- s_1 – Engagement/Flow: A state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete. This is of course a desired state in a training system.

- s_2 – Confusion: A state experienced while encountering the "cognitive disequilibrium" that occurs when confronted with obstacles to goals, interruptions of organized action sequences, impasses, contradictions, anomalous events, dissonance, incongruities, unexpected feedback, uncertainty, deviations from norms, and novelty. In the context of a training system, this is a "productive" form of confusion as the resolution of the impasse provides a sense of accomplishment.

- s_3 – Frustration: A state experienced while encountering the "hopeless confusion" that occurs when an impasse cannot be resolved, the learner gets stuck, there is no available plan, and important goals are blocked

- s_4 – Boredom: A state experienced when a learner disengages from the learning process

The model depicts six primary state transitions, but the design focuses on four (4) transitions that have pedagogical implications to an adaptive training system

- $s_1 \rightarrow s_2$: Caused when an impasse is detected and the learner engages is effortful problem solving

- $s_2 \rightarrow s_1$: Caused when an impasse is resolved. Additional positive affective states, such as delight, may occur as a result of achieving goals or receiving positive feedback

- $s_2 \rightarrow s_3$: Caused when an impasse cannot be resolved, the learner is stuck, or important goals are blocked

- $s_3 \rightarrow s_4$: Caused when persistent frustration prompts the learner to disengage from the learning process

While it is not documented in this model, a direct $s_1 \rightarrow s_4$ transition may also occur if the learner is under-tasked, or when concentration or attention is broken. Proposed training adaptations are presented in two specific contexts: affective state alone and affective state coupled with physiological workload. If the trainee is in a prolonged state of equilibrium, scenario complexity should be increased to trigger the $s_1 \rightarrow s_2$ transition and cause the learner to engage in effortful problem solving. Sustained equilibrium should be managed to prevent an $s_1 \rightarrow s_4$ transition. The $s_2 \rightarrow s_1$ transition back into equilibrium does not require immediate intervention, as it indicates problem solving has been applied to successfully achieve a goal or resolve an impasse. However, the transition should trigger the system to monitor for a prolonged state of equilibrium. The $s_2 \rightarrow s_3$ transition into frustration does not require an immediate intervention; however, it should trigger the system to monitor for a prolonged state of frustration. Sustained frustration should be managed by reducing scenario complexity to prevent the $s_3 \rightarrow s_4$ transition. If the $s_3 \rightarrow s_4$ transition occurs, the scenario complexity should be reduced to present the learner with a more simplified problem, but the complexity of the problem must also increase the learner's interest in re-engaging with the training session. If an $s_1 \rightarrow s_4$ transition occurs, the scenario complexity should be increased to present the learner with a more complex problem that also increases the learner's interest in re-engaging with the training session.

Physiological workload assessment techniques can reinforce, or modify, the adaptations based solely on affective state. For brevity, the differences to the list above are included here. Stable or decreasing workload reinforces the adaptation that increases scenario complexity and triggers the $s_1 \rightarrow s_2$ transition during a prolonged state of equilibrium. A decreasing workload trend should immediately trigger the adaptation to prevent an $s_1 \rightarrow s_4$ transition. Workload provides an added dimension to the $s_2 \rightarrow s_3$ transition into frustration. The transition to frustration, paired with a stable or moderate increase in workload, does not require an immediate intervention, but should trigger the system to monitor for prolonged frustration. The $s_2 \rightarrow s_3$ transition accompanied with a dramatic increase in workload should result in a reduction in scenario complexity to prevent a rapid $s_3 \rightarrow s_4$ transition. Ideally, the coupling of affective and physiological state should allow for early detection and prevent the $s_3 \rightarrow s_4$ transition from occurring.

STUDY: PHYSIOLOGICAL BASED ADAPTIVE TRAINING

This paper describes a study that we are preparing to conduct over the next few months. Unfortunately, we cannot present any results art this time. However, we feel that there is value in conveying our test plan to the scientific community.

The present study is intended to assess the value of adaptive training systems that use measures of subject workload. We intend to test the hypothesis that adaptation using performance and workload (P+WL) will lead to better training outcomes than adaptations using performance only (P). Stated as a testable hypothesis EH_1:

- H_0: performance only based adaptive training score = performance with workload adaptive training score

- H_1: performance only based adaptive training score < performance with workload adaptive training score

In this experiment, both groups (A and B) will receive task training using their respective P+WL or P only adaption scheme. The effectiveness of that training will then be assessed in a graded capstone check-ride. Throughout the training, we will periodically administer subjective workload probes to allow us an independent validation of the accuracy of the OPL workload algorithm.

Each subject will wear a NeXus 4 channel wireless ECG system that collects raw data used by the UPCAT system to assess workload of the participants. Performance metrics from within the virtual environment along with workload are used to adapt the scenario. Figure 28 shows the system architecture used to collect and assess subjects' performance and workload. All audio and video from the HMI, as well as audio and video of the subject is recorded and synchronized. Figure 29 shows the system architecture used to collect and synchronize audio and video data.

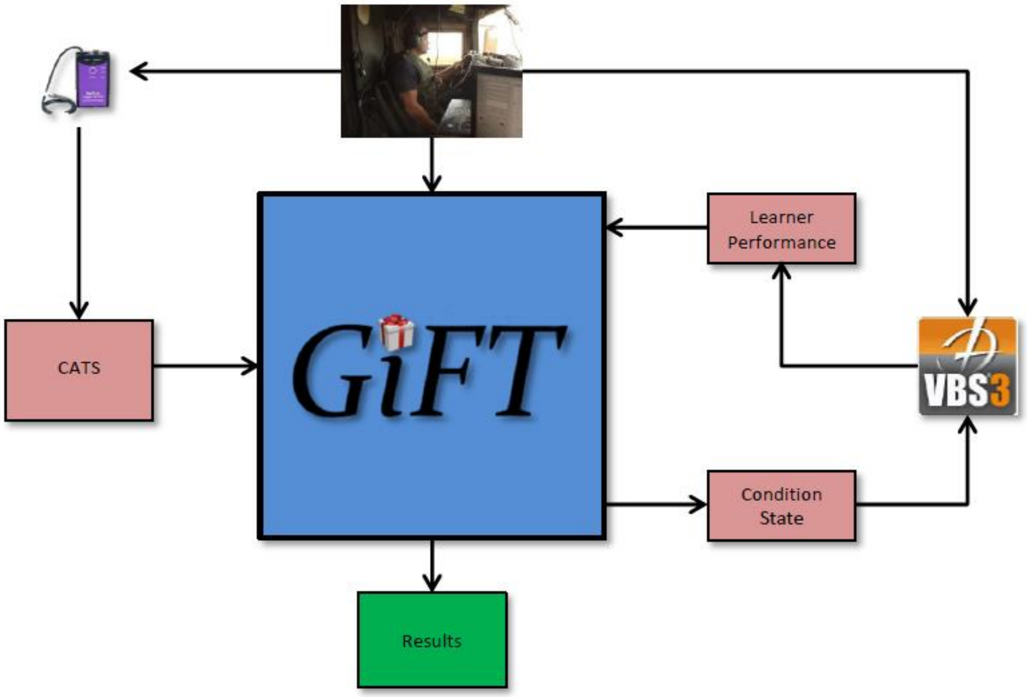


Figure 28. UPCAT System Architecture

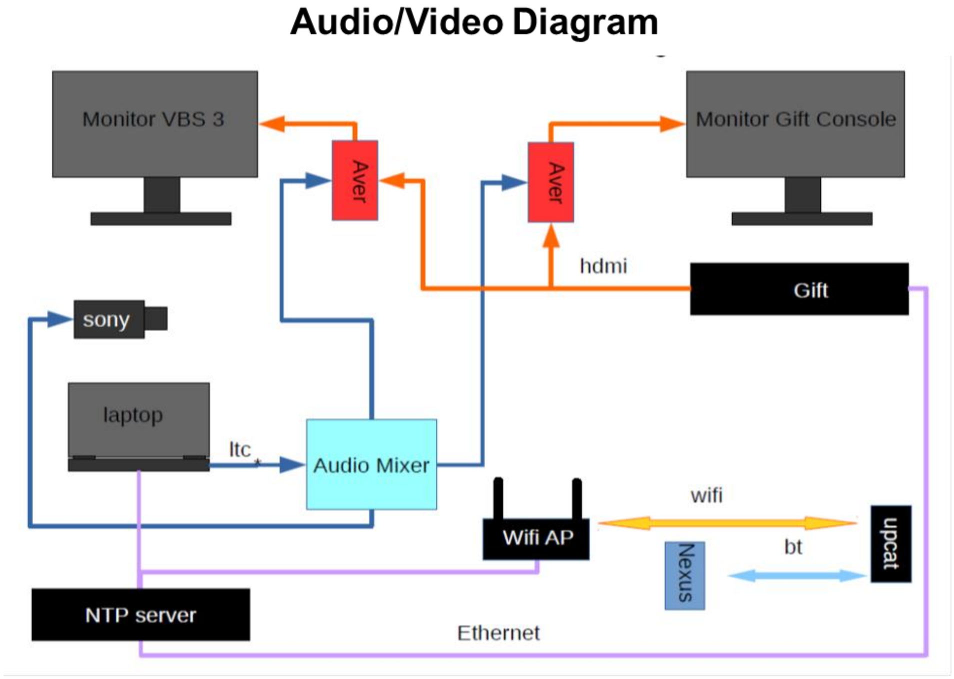


Figure 29. System Architecture for Video and Workload Data Capture

Each participant will complete a GIFT based training course in accordance with the group assigned adaption scheme (P+WL), (P). Within this course, each participant will complete a number of tasks. With the exception of the non-adaptive introduction (warmup), each task has three levels of difficulty (i.e. Easy, Medium and Hard). Participants will return approximately 14 days after their initial course to complete the capstone check-ride. This general GIFT course flow can be seen in Figure 5.

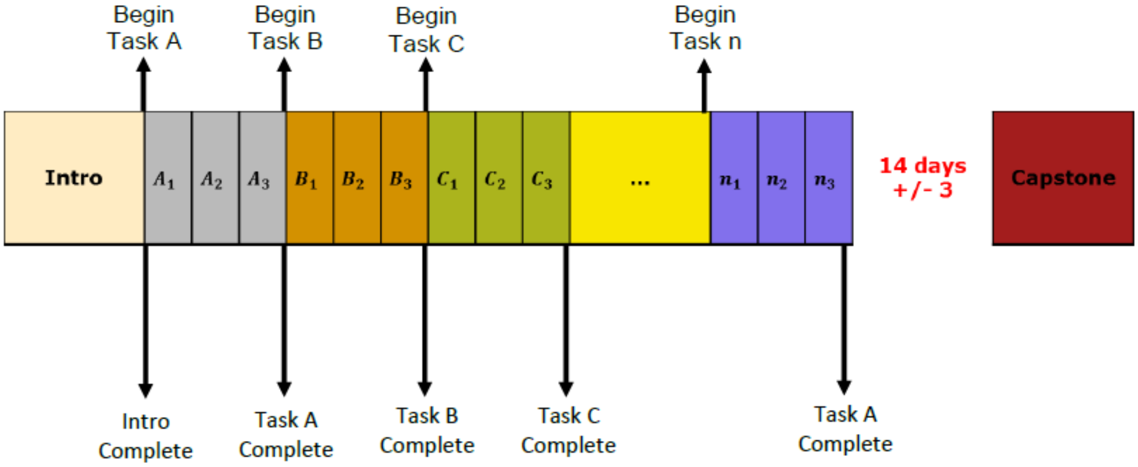
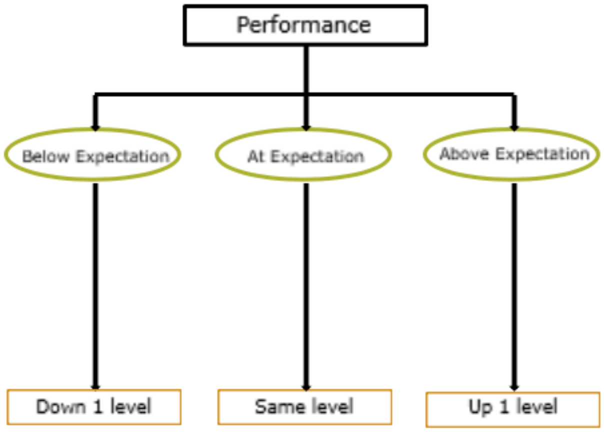


Figure 5. Training Course Flow

Each participant will first attempt each task at the medium difficulty level. Participant performance and workload are assessed throughout the training task and summarized for the adaptation decision at the end of each attempt. For each separate level of difficulty, participant outcomes are classified into one of three groups based on their performance score (green bubbles) as being below expectation, at expectation, or above expectation. The transition to the ensuing task level follows the decision tree shown in Figure 30 and Figure 31 for (P) and (P_+WL) groups, respectively. These adaptation decision trees were adapted from (Mark et al., 2018).



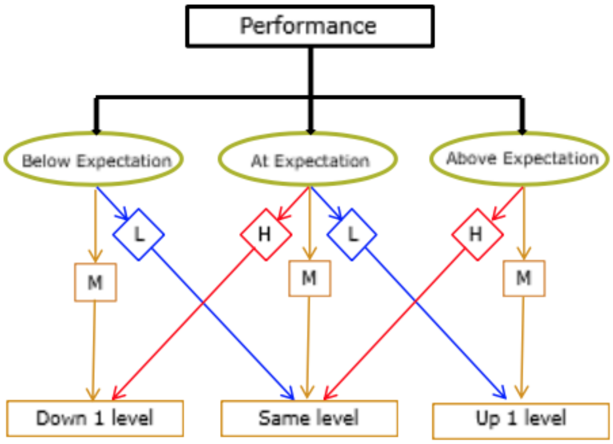


Figure 30. Adaptation Flow for Performance Only Adaptation (P)

Figure 31. Adaptation Flow for Performance with Workload Adaptation (P+WL)

The capstone check-ride consists of one ever increasingly difficult task that encompasses all task elements from all previous part tasks. Participants continue through this increasingly difficult capstone check-ride until they fall below performance thresholds. The point in the check-ride where they fail is the dependent measure of training effectiveness with a later failure being better than an early one. We chose this method of testing to avoid ceiling or floor effects where many or all participants pass or fail a check-ride of a selected level of difficulty.

Throughout the experimental GIFT driving course we evaluate four conditions. They include the GIFT Corridor Boundary, OPL Workload, Maintain Speed, and Collision Avoidance. GIFT evaluates both Corridor Boundary and Workload Classifier conditions while VBS 3 evaluates Maintain Speed and Collision Avoidance conditions. VBS 3 maintains a state variable for each Corridor Boundary, Workload Classifier and Maintain Speed conditions. Each GIFT condition has three state transition strategies: one for each of the below, at or above expectation evaluations (increasing, decreasing and maintaining for workload), in accordance with the flow graphs shown in Figure 30 and Figure 31.

We added six new Environmental Control Enums; one for each condition at each evaluation which are used in GIFT state transition strategies. Using the *sendCommand()* function from GIFT’s VBS 3 Plugin Interface we are able to send any valid VBS 3 script command. For example, assume that the subject has trouble with tracking the vehicle in the middle of the driving lane. Therefore, the Corridor Boundary condition will evaluate to a value of *below expectation*. GIFT executes its corridor boundary *from anything to below expectation* state transition strategy which sends the VBS 3 command *["BELOW”] call setCorridorState*, and the Corridor boundary state variable maintained by VBS 3 is updated to BELOW. The same happens for all evaluations and accompanying state transition strategies for both the Corridor Boundary and Workload Classifier conditions.

Currently we have hard-coded the commands through the use of the Environmental Control Enum. This is restrictive as VBS 3 allows for thousands of commands. We experimented with the sendCommand() function, and were able to send multiple commands separated by a semi-colon with a single call to sendCommand(). We believe the ability to create custom commands within the state transition strategies instead of the restrictive hard coded example we are using to be an appropriate addition to GIFT. We could add a single CUSTOM_COMMAND enum to the list of GIFT Environmental Control Enums. The command(s) could then be written into, and read from, the course .dkf file when GIFT calls the state transition strategy implementing that command.

Both Maintain Speed and Corridor Boundary Conditions are called inside of an event handler attached to the subject object which fires every time the subject object moves. The event handler includes a timer that only calls the evaluation functions for both conditions for every evaluation interval (currently every 1 second while the vehicle is moving). For both the Corridor Boundary and Maintain Speed conditions, VBS 3 maintains a timer for each of the below, at or above expectation evaluations. At every evaluation interval, VBS 3 checks the current state of the two conditions and adds the elapsed time from the previous evaluation to its corresponding timer. The final evaluation for each of the Corridor Boundary and Maintain Speed conditions is assigned based on what percentage of the total time was spent in each state based on Table 1 (note that actual logic accounts for ranges and not set values).

Table 1. Evaluation Assignment for Corridor Boundary and Maintain Speed Conditions

Below/Total	At/Total	Above/Total	Evaluation
0%	0%	100%	ABOVE
0%	25%	75%	ABOVE
0%	50%	50%	ABOVE
0%	75%	25%	AT
0%	100%	0%	AT
25%	0%	75%	ABOVE
25%	25%	50%	AT
25%	50%	25%	AT
25%	75%	0%	AT
50%	0%	50%	AT
50%	25%	25%	BELOW
50%	50%	0%	BELOW
75%	0%	25%	BELOW
75%	25%	0%	BELOW
100%	0%	0%	BELOW

Maintain Speed condition is graded through the use of a target speed and a speed window. If the subject is outside the speed window, they are evaluated to below expectation. If the subject is inside the center one-third of the speed window, then they are evaluated to above expectation. If the subject is between inner one-third and outside of the speed window, then they are evaluated to at expectation. Let the target speed be 35 km/h, and the speed window be 6km/h. If the subject's speed is more than 41 km/h or less than 29 km/h, then they are outside the speed window and are evaluated to below expectation. If the subject's speed is between 37 and 41 km/h or between 29 and 33 km/h, then they are evaluated to above expectation. If the subject's speed is between 33 and 37 km/h, then they are evaluated to at expectation.

Collision Avoidance is graded through the use of upper and lower bounds. If, at the end of an attempt, the subject has had fewer collisions than the lower bound they are evaluated to above expectation. If the subject has had more collisions than the upper bound, then they are evaluated to below expectation. Anything in between receives an evaluation of at expectation.

The performance evaluation used for adaptation is an aggregate of these three condition evaluations. Each task weights the evaluation of the three conditions differently, and some are not even used at all for some tasks. Let Task one be driving in reduced visibility, where the subject is evaluated on maintaining speed and corridor boundary while driving through a sandstorm. It is important to maintain their speed, but it is more important to stay on the road. So a fair weighting of the singular evaluations to determine the aggregate performance evaluation could be set to Equation 1. The aggregate performance and workload evaluations are then used to decide on the scenario adaptation based on the adaptation trees from Figure 6 and Figure 7.

$$aggPer = 0.40 \times speedEvaluation + 0.60 \times corridorEvaluation$$

Equation 1. Aggregate Performance Grade for Task 1

The aforementioned evaluation and adaptation logic is controlled by various scripts and event handlers. VBS 3 init.sqf script (called at the start of the scenario) compiles multiple scripts that set-up the global variables; the VBS 3 waypoints, create the files used for data collection; task, time, grading and GIFT message related functions; event handlers; and scripts that set-up the evaluation of conditions. The scenario adaptations needed for each level of difficulty for each task are also contained within their own scripts.

The current GIFT Corridor Boundary condition did not allow an evaluation of above expectation, and we were concerned about fairness in the evaluations of the two groups (A & B). For example: the ability of subjects from group A to reach an evaluation of above expectation and an adaptation of up 1 level compared to subjects from group B's ability to reach the same adaptation through an evaluation of at expectation with a decreasing workload as shown in the adaptation trees in Figure 6 and Figure 7. We saw a potential for a biased evaluation and made changes to allow GIFT's Corridor Boundary condition to evaluate to above expectation.

It works in much the same way as the Maintain Speed condition. If the subject is outside the corridor, they are evaluated to below expectation just as before. The change we made affected the way the subject is graded while inside the corridor. If the subject is inside the center one-half of the corridor, then they are evaluated to above expectation. If the subject is between inner one-half and outside the corridor, then they are evaluated to at expectation. Let the corridor be 10 meters wide. If the subject is more than 5 feet away from the center of the corridor, then they are outside the corridor and are evaluated to below expectation. If the subject is less than 2.5 meters from the center of the corridor, then they are evaluated to above expectation. If the subject is less than 5 meters but more than 2.5 meters away from the center of the corridor, then they are evaluated to at expectation.

For purposes of our study, we write all data related to decision making with respect to the evaluation of the different conditions, aggregate scoring and adaptations throughout the course to .csv files. Each data point is timestamped with the system time (the exact time and date according to the computer the subject is using).

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

We invested considerable effort in the establishment of an architecture that tightly integrates the capabilities of the GIFT framework with VBS 3 as a representative of an Army Synthetic Training Environment (STE). This architecture provides a robust control interaction capability between the two systems. Additionally, this architecture includes tight integration of a continuous workload assessment system (CATS) using a deterministic nonlinear workload classifier that analyses the ECG waveform in embedded phase space. This apparatus is capable of assessing learner state in real-time, in this case using a driving task, and applying performance and workload assessments to automatically configure scenario transitions for adaptive training.

Additionally, we invested significant effort integrating an existing GIFT learner affect classifier library (Ocumpaugh et al., 2017) into this framework. This classifier uses a Kinect sensor to track features on the learner's face to classify states of emotion. Even though we spent a tremendous amount of effort in an attempt to integrate this library, we were, to date, not yet able to gain a reliable classification from it. Therefore, in the upcoming validation study using this apparatus, we decided not to use learner affect as a state variable to invoke scenario transitions. If we manage to get the affect state library to work, we will collect data from its affect state classifier for separate and off-line analysis.

REFERENCES

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

Engler, J., & Schnell, T. (2013). *Deterministically Nonlinear Dynamical Classification of Cognitive Workload.* Paper presented at the I/ITSEC 2013, Orlando, FL.

FAA. (2008). *Pilot's Handbook of Aeronautical Knowledge* (FAA-H-8083-25A). Retrieved from

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience, 8*(385). doi:10.3389/fnins.2014.00385

Lintern G., & Gopher D. (1978). Adaptive training of perceptual-motor skills: issues, results, and future directions. *International Journal of Man-Machine Studies, 10*(5), 521-551.

Mane A., & Wickens C. (1986). *The Effects of task Difficulty and Workload on Training.* Paper presented at the 30th Annual Meeting of the Human Factors and Ergonomics Society.

Mané, A. M., Adams, J. A., & Donchin, E. (1989). Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, 71*(1), 179-196. doi:https://doi.org/10.1016/0001-6918(89)90008-5

Mark, J., Thomas, N., Kraft, A., Casebeer, W. D., Ziegler, M., & Ayaz, H. (2018). *Neurofeedback for Personalized Adaptive Training*, Cham.

Matthews, R., Turner, P., McDonald, N., Ermolaev, K., McManus, T., Shelby, R., & Steindorf, M. (2008, Aug. 20-24, 208). *Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes.* Paper presented at the The 30th Annual Intl Conf of the IEEE on Engineering in Medicine and Biology, Vancouver, BC.

Ocumpaugh, J., Andres, J. M., Baker, R., DeFalco, J., Paquette, L., Rowe, J., . . . Sottilare, R. (2017). *Affect Dynamics in Military Trainees Using vMedic: From Engaged Concentration to Boredom to Confusion*, Cham.

OPL. (2014). Cognitive Assessment Tool Set (CATS) User Manual Retrieved from Iowa City, Iowa:

Richter, M., & Schreiber, T. (1998). Phase Space Embedding of Electrocardiograms. *Physical Review E, 58*(5), 6392-6398. doi:10.1103/PhysRevE.58.6392

Schnell T, Reichlen C., Geiselman E, Knox J., Williams H., & Ercoline W. (2017, May 8 - May 11, 2017). *A Comparison of Helmet-Mounted Display Symbologies During Live Flight Operational Tasks.* Paper presented at the 19th International Symposium on Aviation Psychology, Dayton, Ohio, USA, May 8 - May 11, 2017.

Schnell T., Hoke J., & Romeas T. (2017, Mar 7-9 2017). *Achieving the Third Offset: Maximizing Human-Machine Symbiosis.* Paper presented at the 2017 NDIA Human Systems Conference,, Waterford at Springfield.

Schnell T., Reichlen C., & C., R. (2017). Spatial Disorientation Threat Characterization for F-35 Representative Helmet-Mounted Display Use in the Flight Environment Retrieved from Dayton, OH:

Wickens C. D. (1992). *Engineering Psychology and Human Performance*. New York: Harper Collins.

Wickens C. D. (2008). Multiple resources and mental workload. *The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449-455.

Wilson, G., & Russell, C. (2003). Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors*, *45*(4), 635-644.

ABOUT THE AUTHORS

Mr. Nathan D. Smith *Nathan joined the Operator Performance Laboratory's (OPL's) team in September of 2017. He currently works as a Research Assistant and is also a student at the University of Iowa College of Engineering. Nathan will graduate in the fall of 2018 BSE in Electrical and Computer Engineering, with a focus in Software Design and a minors in both Business and Computer Science.*

Mr. Ezekiel D. Gunnink *is a full-time Research Assistant at the Operator Performance Lab (OPL). He is currently completing his computer & electrical engineering degree, with a specialization in software engineering. He is an expert in computer vision, graphics, and advancement in sensor technology. He is one of the lead developers and software architects of the CATS software.*

Mr. Chris Reuter *joined the Operator Performance Laboratory in October 2012. He has a bachelor's degree in mathematics from the University of Iowa and is currently pursuing a PhD in Industrial Engineering. Prior to joining OPL, Chris worked in the financial industry and brings to the team a wide range of management and operational experience. Chris's primary focus at OPL is working with human factors studies and data analysis.*

Dr. Tom Schnell *is a Professor in the Department of Industrial and Mechanical Engineering at the University of Iowa. He is also the Director of the Operator Performance Laboratory (OPL) where he has been the principal investigator on around 230 research projects. Tom has an undergraduate degree in Electrical Engineering and a MS and PhD in Industrial Engineering. He is a commercial pilot, test pilot, and flight instructor for fixed wing airplanes and rotorcraft.*

Dr. Jason D. Moss *is a Research Psychologist for the U.S. Army Research Laboratory (ARL)– Human Research and Engineering Directorate (HRED), Advanced Training and Simulation Division (ATSD) in Orlando, Fl. He has over 14 years of experimental research experience in the areas of military psychology, training, perception, human factors psychology, simulator sickness, and virtual environments. He has a Ph.D. in Human Factors Psychology from Clemson University.*

Predicting Students' Unproductive Failure on Intelligent Tutors in Adaptive Online Courseware

Seoyeon Park and Noboru Matsuda
Department of Teaching, Learning and Culture Texas A&M University

INTRODUCTION

The wheel-spinning phenomenon in the current paper refers to students' unproductive failure within a computer-based learning environment using Intelligent Tutoring Systems (ITSs). Beck and Gong (2013) found that students often spend a considerable amount of time practicing a skill in ITSs without making progress. This phenomenon is coined *wheel spinning* because students' learning pattern is like a car stuck in the mud. The wheel-spinning phenomenon has been observed universally on many ITSs (Beck & Gong, 2015). When wheel spinning, students often become frustrated and demotivated to learn (Cen, Koedinger, & Junker, 2007; Baker, Gowda, & Corbett, 2011). Therefore, several studies explored building an effective and reliable wheel-spinning detector to detect the moment of wheel spinning. Beck and Gong (2015) suggested a generic model using logistic regression to predict wheel spinning with three aspects: student's performance on the skill, the seriousness of the learner, and general factors of the learning material such as skill difficulty. Matsuda, Chandrasekaran, and Stamper (2016) built a more simplified wheel-spinning predictor as a combination of the probability of mastery based on Bayesian knowledge tracing, and a neural-network model.

In the current paper, we investigate the wheel-spinning phenomena in the context of adaptive online courseware where many ITSs are embedded into the online courseware. Students are provided with multimedia instruction, including paragraph text instruction, images, videos, and traditional formative assessments such as multiple choice and fill-in-the-blank questions. ITSs are embedded in the courseware as a type of formative assessment as well. In this rich learning environment, we aim to predict the moment of wheel spinning so that the system can provide proactive scaffolding to maintain students' motivation and engagement.

The goal of the current paper is to contribute to the Generalized Intelligent Framework for Tutoring (GIFT) framework by investigating the wheel-spinning phenomena on the adaptive online course platform with many ITSs on which wheel spinning will happen. We discuss the unique nature of the wheel-spinning in this environment and our current progress. The current work is part of our on-going project where we develop evidence-based learning-engineering methods to build adaptive online courseware, called PASTEL (**P**ragmatic methods to develop **A**daptive and **S**calable **T**echnologies for next generation **E-L**earning).

The existing models for wheel-spinning detection have some limitations. First, existing models have low recall rates around 0.25-0.50, suggesting that these models are weak and can only detect less than half of actual wheel-spinning cases. Since not catching a moment of wheel spinning would impact students' motivation, we need to develop a model that has a high sensitivity to wheel spinning.

Second, most of the existing models are aimed to *detect* a moment of wheel spinning, instead of *predicting* students who are likely to get stuck. Matsuda et al. (2016) applied a neural-network model to predict wheel spinning at an early stage of learning. However, its prediction power is approximately 0.25, which is still insufficient for practical use. The primary purpose of catching wheel spinning is to maintain students' motivation for learning, it is crucial to *predict* the moment of wheel spinning in advance. With the early prediction, we can provide students with proactive scaffolding that keep those students from experiencing wheel spinning.

Third, existing wheel-spinning detectors/predictors explain wheel spinning on individual skills (the skill-level model), indicating the likelihood of a student to fail to obtain mastery on a particular skill. Historically speaking, this trend has been held because problems on ITSs are broken down into a fine-grained skill set, often called a knowledge component (KC) model (Koedinger, Corbett, & Perfetti, 2012). Taking skills as a unit of analysis works well for a "standalone" ITS (including ITS with "units"). As mentioned above, we target the adaptive online courseware as the platform for wheel-spinning prediction. During our initial trial for creating an instance of adaptive online courseware (called *CyberBook*) with in-service teachers as curriculum consultants, we asked in-service teachers to tag each ITS with the most essential skill that students will learn by solving problems on a corresponding ITS. We observed that in-service teachers often tagged an ITS with a skill that does not appear in any steps on the ITS (as opposed to selecting one of the steps on an ITS as the most essential step hence the most essential skill). For example, an ITS that teaches how to compute the slope of a given linear equation involves steps such as subtracting and dividing terms, but no single step is about "computing the slope." On CyberBook, when a student gets stuck (i.e., wheel spins) on a particular ITS, the system provides the student with proactive scaffolding by showing a link to the related instruction paragraph. A naïve research question therefore is: *Should wheel spinning be predicted on steps within an ITS (hence triggers the proactive scaffolding) or on the ITS as a whole?* Given our observations from in-service teachers tagging ITSs with a skill, we hypothesized that the ITS as a whole should be the unit of analysis for wheel-spinning prediction.

The goal of the current study is to develop a wheel-spinning predictor, which can distinguish students who have a high possibility to wheel-spin as quickly as possible, at the problem level. The specific research questions are as follows:

1. How accurately can we predict wheel-spinning at the problem level?

2. How early can we detect wheel-spinning at the problem level?

To build a wheel-spinning prediction model that can find wheel-spinning cases with high accuracy and speed, we propose to use four general factors, students' performance, hint usage, the sum of response time, and difficulty of each problem type. These factors are generally available on most ITSs and are known to be effective in predicting students' academic performance. We have previously built a wheel-spinning predictor at the step-level (Park & Matsuda, under review). In the current paper, to understand whether the problem-level prediction is any better than step-level prediction, we apply logistic regression and an ensemble modeling to predict wheel-spinning cases at the problem-level.

DATA PREPROCESSING

We used an existing dataset from DataShop, entitled 'Cog Model Discovery Experiment Spring 2010' in the 'Geometry Cognitive Model Discovery Closing-the-Loop study' project. There were 49 skills forming 45,597 observations done by 123 students in the 'KTracedSkills' model in this data. This dataset contained 5,279 student-skill pairs. The DataShop data uses fine-grained skills that are decomposed by Learning Factor Analysis. In order to predict wheel-spinning at the problem level, we needed to create 'problem type' as a different dimension of measuring wheel-spinning. We used a text-mining technology named SMART to create 'problem type'. SMART is an AI technology that can compute the similarity among words within the text and extract a keyword. We input hint message of each intelligent tutor and set an arbitrary k number; k=25, 50, 75, 100. After SMART generates problem types, those problem type models were validated with the DataShop knowledge component model. Table 1 shows the result of comparing SMART generated problem type models.

Table 1. Comparison of SMART generated problem type models

Model name	Problem types	Observations with Problem types	AIC	BIC	RMSE (student stratified)	RMSE (item stratified)
SMART k=25	17	85,115	46,986.00	48,454.30	0.273130	0.271673
SMART k=50	28	85,115	46,787.87	48,461.83	0.272680	0.271298
SMART k=75	40	85,115	47,114.50	49,012.91	0.274457	0.272230
SMART k=100	39	85,115	47,145.30	49,025.00	0.273595	0.272066
KTracedSkills	49	41,756	29,096.28	31,005.13	0.333781	0.324864

KTracedSkills row is the baseline when comparing other SMART generated problem type models. We chose to use the problem type model named ‘SMART k=50’ because this model shows the lowest root mean squared error (RMSE). Comparing to KTracedSkills model, ‘SMART k=50’ has bigger AIC and BIC, but these figures are affected by the number of observations. Considering that the number of observations of our SMART generated problem type models is more than twofold, the AIC and BIC figures make sense.

We employed the ‘SMART k=50’ problem type model and did data preprocessing. There were 28 problem types and we created 1,889 student-problem type pairs. Mastery in this study is defined as three consecutive correct responses on one’s first attempt within 10 practice opportunities (Beck and Gong, 2013) on a problem level. We filtered out “indeterminate” students, who did not practice on enough opportunities, which was 10 opportunities in this study, for us to define their mastery (Beck & Gong, 2015). After removing indeterminate student-problem type pairs, this dataset came to contain 1,794 student-problem type pairs and 31,801 observations with 123 students. The dependent variable is whether a student shows mastery (M) or wheel-spinning (W) on a problem type within 10 opportunities, based on the response sequences of each student-problem type pair. In order to see how early we can predict wheel-spinning on a problem type, we made subset at each practice opportunity from the third opportunity to the ninth opportunity.

FEATURES

We used four features that are all general factors in any dataset of ITSs. This is because first, we want to show that predicting wheel-spinning at the problem level can be generalized among any ITS construct, and second, we want to build a more simple and scalable wheel-spinning model.

Student’s performance on each problem type

The first feature we used is how well a student did on a problem type. This represents a student’s ability to solve a certain type of problem. This is calculated as the average probabilities of correct first attempts per each student-problem type pair.

Problem type difficulty

The second feature is the difficulty of each problem type. We calculated this variable by getting the average correct response rate of each problem type across all students who practiced the problem type.

Max_hint

Hint usage is regarded as one of the important factors in explaining students' learning (Feng, 2009; Rivers, 2017). Thus, we used the maximum number of hint usage of students on each problem type.

Sum_duration

Response time is one of the key features in a wheel-spinning model (Beck and Gong, 2015). Each problem type has several steps, so we added step duration of constituent steps to get the response time of a student on each problem type.

PREDICTION MODELS AND RESULTS

A basic model for wheel-spinning prediction at the problem level

With the combination of features above, we trained a logistic regression to build a basic model for wheel-spinning prediction at the problem type level with ten-fold cross validation. The coefficients would not be suggested due to the limit of space. This basic model for wheel-spinning prediction shows high accuracy throughout practice opportunities in Table 2. The overall accuracy in percent correct is 92.75% and overall AUC is 0.916. Considering the accuracy of the generic wheel-spinning model in a skill level (Beck and Gong, 2015), which was less than 90% in percent correct and 0.9 in AUC, this basic model shows a sufficient performance with even using the smaller number of features.

Table 2. Accuracy of a basic model per practice opportunity

	opp3	opp4	opp5	opp6	opp7	opp8	opp9
Percent correct	0.908	0.91	0.917	0.923	0.932	0.949	0.950
AUC	0.856	0.863	0.894	0.939	0.938	0.949	0.975

We not only need to see the accuracy of this model but also precision and recall rates in order to have an insight into its classification. Table 3 shows the precision and recall rate of this model at each opportunity. Both rates are increasing by each opportunity. However, the precision rate is 60% and recall rate is 33.65% on average across the third through ninth opportunity. These figures are relatively low comparing to those of existing wheel-spinning models (around 70% in precision rate and 25~50% in recall rate). Moreover, using this basic model, we cannot predict wheel-spinning on a problem type level as early as possible due to its weak recall rate in every practice opportunity.

Table 3. Precision and Recall rates of a basic model per practice opportunity

	opp3	opp4	opp5	opp6	opp7	opp8	opp9
Precision	0.358	0.440	0.551	0.619	0.666	0.755	0.814
Recall	0.0798	0.176	0.230	0.285	0.417	0.612	0.554

The upgraded model for wheel-spinning prediction at the problem level using gradient boosted decision tree model. We found that the basic model has some limitations in terms of its precision and recall rates. Thus, other data mining techniques were explored to find a better prediction model. Especially, we focused on getting a higher recall rate in the early phases so that we can predict wheel-spinning on a problem level as quickly as possible. We discovered that the gradient boosted decision tree model using the same combination of features shows much better performance in accuracy, precision, and recall rate. Gradient boosted trees is an ensemble of multiple tree models to create a powerful prediction model for classification. This algorithm generates a series of trees where trees are made by correcting poor predicted examples of the previous trees in the series. We trained this model with a ten-fold cross validation by each practice opportunity. The overall accuracy of the upgraded model is 96.90% and 0.97 in AUC. Table 4 shows that this model shows higher accuracy throughout opportunities.

Table 4. Accuracy of the upgraded model per practice opportunity

	opp3	opp4	opp5	opp6	opp7	opp8	opp9
Percent correct	0.953	0.955	0.961	0.972	0.974	0.985	0.981
AUC	0.942	0.96	0.974	0.985	0.987	0.991	0.997

This model also has a much higher performance on both precision and recall rates than those of our basic model. Overall, the precision rate is 87% and recall rate is 75% across the third through ninth opportunity. These figures are showing that this upgraded model has greater wheel-spinning prediction power than other existing models. Applying this model, we can predict wheel-spinning on a problem type on students' fifth opportunity with 65% accuracy and over 80% accuracy on the sixth opportunity.

Table 5. Precision and Recall rates of the upgraded model per practice opportunity

	opp3	opp4	opp5	opp6	opp7	opp8	opp9
Precision	0.792	0.834	0.867	0.843	0.840	0.958	0.963
Recall	0.616	0.606	0.651	0.829	0.865	0.864	0.813

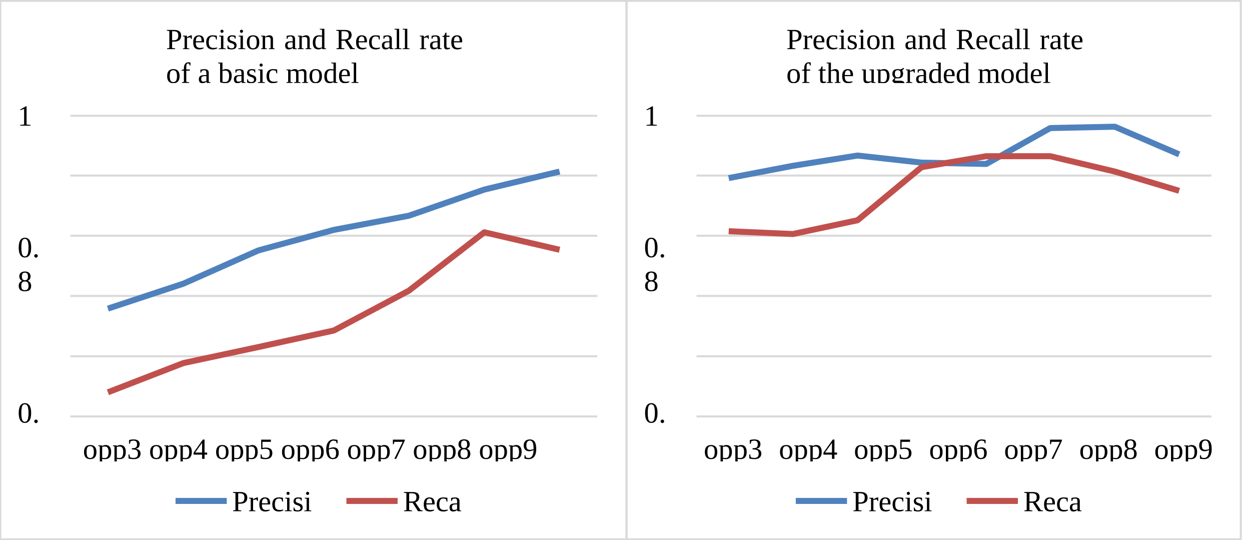


Figure 1. Precision and Recall rate of two models

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The goal of the current study is to seek the way in which we can predict wheel-spinning at the problem level (i.e., an individual ITS as opposed to a step in an ITS) with high accuracy, prediction power, and speed. We have some important findings in this work. First, we found that the four general variables (i.e., students' performance, hint usage, the sum of response time, and difficulty of each problem type) that are available for most ITSs can sufficiently build a prediction model for wheel spinning at the problem level. Our basic model with four general variables shows similar performance with existing models in its accuracy (average percent correct is 0.93 and overall AUC is 0.92). Its recall rate (0.34) is higher than that of the other wheel-spinning prediction model (Matsuda, Chandrasekaran, and Stamper, 2016).

Second, we explored other machine learning techniques to improve the accuracy of wheel-spinning prediction. Our upgraded model with gradient boosted decision tree algorithm shows enhanced precision and recall rate with an average recall rate of 0.75. A pragmatic merit of this upgraded model is its speed—the recall rate on the sixth practice opportunity is around 0.83. This would expand our chance to promote students' efficient learning in ITSs by keeping them from wheel spinning in advance.

As for the contribution to the Generalized Intelligent Framework for Tutoring (GIFT), the current study demonstrated a generic technique to predict students' unproductive failure (wheel spinning) on an ITS embedded into adaptive online courseware. The adaptive online courseware with embedded intelligent tutors has a tremendous potential for future online learning hence investigating fundamental techniques such as the wheel-spinning prediction plays an important role. We also demonstrated an importance of building the wheel-spinning predictor at the different level of granularity of the skill model.

For future study, one intriguing topic would be to find what we should do once we predict wheel-spinning cases. What would be an effective intervention for those who are predicted to wheel spin on a problem? Another suggestion is to explore other machine learning techniques to improve the current wheel-spinning prediction model. This study used logistic regression and gradient boosted decision tree. Our upgraded model using gradient boosted decision tree shows significant improvement in predicting wheel-spinning, however, a drawback of using this technique is that it is hard to interpret the model itself. Finally, it would also be an interesting idea to extend the research regarding why students show unproductive failure in learning by using ITSs.

ACKNOWLEDGEMENT

This study is supported by NSF grants No. 1623702 and 1644430.

REFERENCES

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. American Psychologist, 51(4), 355.

Baker, R. S., Gowda, S. M., & Corbett, A. T. (2011). Towards predicting future transfer of learning. In International Conference on Artificial Intelligence in Education. Springer, Berlin, Heidelberg, 23-30.

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. In International Conference on Artificial Intelligence in Education. Springer, Berlin, Heidelberg, 431-440.

Beck, J., Ostrow, K. & Wang, Y. (2016). Students vs. Skills: Partitioning Variance Explained in Learner Models. In The 9th International Conference on Educational Data Mining. ACM

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In International Conference on Intelligent Tutoring Systems. Springer, Berlin, Heidelberg, 164-175.

Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In Icml. 96, 148-156. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Gong, Y., & Beck, J. E. (2015). Towards detecting wheel-spinning: Future failure in mastery learning. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale. ACM, 67-74.

Gong, Y., Wang, Y., & Beck, J. (2016). How long must we spin our wheels? Analysis of student time and classifier inaccuracy. Student modeling from different aspects, 32-38.

Heffernan, N., Heffernan, C., & Ostrow, K. (2018). The Assessments underlying ASSISTments. In Proceedings of the AERA 2018. NY.

Koedinger, K.R., Corbett, A.C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive Science, 36 (5), 757- 798.

Maclin, R., & Opitz, D. (1997). An empirical evaluation of bagging and boosting. AAAI/IAAI, 546-551.

Matsuda, N., Chandrasekaran, S., & Stamper, J. C. (2016). How quickly can wheel spinning be detected?. In EDM.607-608.

Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. Instructional-design theories and models: A new paradigm of instructional theory, 2, 161-181.

Schapire, R. E. (2003). The boosting approach to machine learning: An overview. In Nonlinear estimation and classification. Springer, New York, NY, 149-171.

Seymour, E., & Hewitt, N. M. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview.

Stamper, J., & Ritter, S. (2010). Cog Model Discovery Experiment Spring 2010. Dataset 392 in DataShop. Retrieved from https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=392.

Park, S., & Matsuda, N. (under review). Early wheel-spinning detection in student performance on cognitive tutors using an ensemble model. Journal of Educational Data Mining.

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007). The Cognitive Tutor: Applied research in mathematics education. Psychonomics Bulletin & Review, 14(2), 249-255.

Watkins, J., & Mazur, E. (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. Journal of College Science Teaching, 42(5), 36-41.

ABOUT THE AUTHORS

***Seoyeon Park** is a Ph.D. student at Texas A&M University and a research associate in the Innovative Educational Computing Laboratory, whose research interest is educational data mining and stem education with adaptive scaffolding in computer-based learning environments.*

***Dr. Noboru Matsuda** is an Associate Professor of Cyber STEM Education at the Department of Teaching, Learning, and Culture; and the director of the Innovative Educational Computing Laboratory. He leads the NSF funded project on the data-driven learning engineering methods to build adaptive online courseware where the research team investigates scalable AI techniques to evidence-basely build adaptive online courseware.*

Modeling Training Efficiency and Return on Investment for Adaptive Training: GIFT Integration

Gregory A. Goodwin[1], James Niehaus[3]
ARL HRED[1], Charles River Analytics[2]

ABSTRACT

Adaptive training promises more effective training by tailoring content to each individual. Where non-adaptive training may be just right for one segment of the student population, there will be some students that find it too easy while others find it too difficult. Another, often ignored benefit of adaptive training, is improved training efficiency by minimizing the presentation of unnecessary material to learners. One implication of this is that intelligent, adaptive training should require less time to train a population of learners to a given level of proficiency than non-adaptive training. The gains in efficiency should be a function of several factors including 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). This paper describes work in the second year of a three year effort showing the results of a predictive model for training efficiency based on those factors and how it could be integrated into the Generalized Intelligent Framework for Tutoring (GIFT) architecture. How this model supports return on investment decisions for authors is also discussed.

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT) is an open-source, modular architecture developed to reduce the cost and skill required for authoring adaptive training and educational systems, to automate instructional delivery and management, and to develop and standardize tools for the evaluation of adaptive training and educational technologies (Sottilare, Brawner, Goldberg, & Holden, 2012a; Sottilare, Goldberg, Brawner, & Holden, 2012b). By separating the components of ITSs, GIFT seeks to reduce development costs by facilitating component reuse.

Meta-analyses and reviews support the claim that intelligent tutoring systems (ITS's) improve learning over typical classroom teaching, reading texts, and/or other traditional learning methods. (Dynarsky et al. 2007; Dodds and Fletcher 2004; Fletcher 2003; Graesser et al. 2012; Steenbergen-Hu and Cooper 2013, 2014; VanLehn 2011). In fact, ITSs have been shown to improve learning to levels comparable to Human tutors (VanLehn et al. 2007; VanLehn 2011; Olney et al. 2012).

As shown in Figure 1, while improved training effectiveness is certainly a benefit of ITS technology, an- other important benefit is improved training efficiency over one-size-fits-all training. The goal of an ITS is to identify the gaps in knowledge specific to each learner so that training can focus on filling just those gaps. One of the problems of one-size-fits-all training is that to insure all trainees can comprehend the instruction, it must be developed for trainees with the least experience, knowledge, and aptitude. Though less costly to develop, the material is presented a pace that is slow and that includes content not needed for more experienced, higher aptitude trainees. An ITS would be expected to reduce the time needed to deliver training to such trainees.

The reduction in time to train (i.e., improved acquisition rate) is an important metric because reductions in training time represent cost savings. This is especially true for military trainees who are paid a salary.

Reductions in the time needed to train those trainees save salary costs for both trainees and instructors. For large-volume courses, those savings can be substantial.

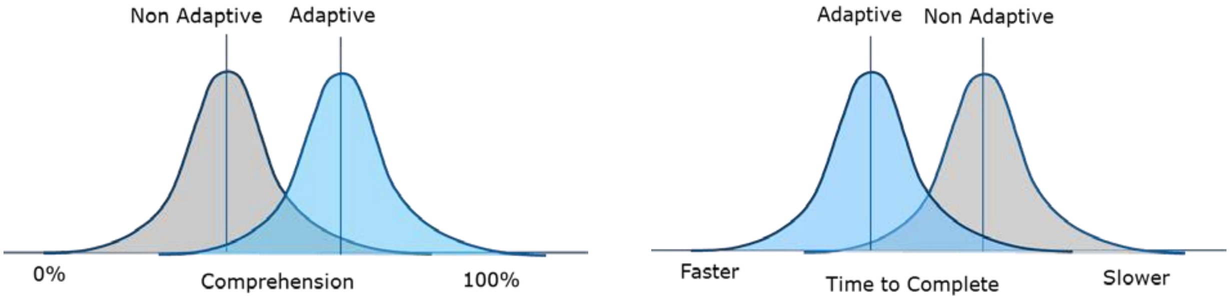


Figure 1: Benefits of adaptive training. On the left, adaptive training can increase the subject comprehension from a fixed time to complete. On the right, adaptive training can decrease the time to complete training content with a fixed level of comprehension.

All of this highlights the need for a means to model and predict training efficiency gains (i.e., time saved) by ITSs generally and GIFT specifically. Having the ability to model time saved by the use of adaptive, intelligent training, as compared to existing or non-adaptive training would have benefits throughout the lifecycle of a course. During the design of new training, the training developer could more easily make decisions about the relative costs and benefits of adding adaptive features. For example, adding extensive remedial training for easy-to-understand concepts may benefit such a small percent of the population of learners, that the net reduction in training time would be too small to make those features worth the cost of development.

During training delivery, actual trainee data could be used to verify and/or improve the model. For example, suppose the model assumed that learners with an aptitude above criteria A would have a 95% probability of understanding concept B without needing any remediation. Learner data could then be used to validate or adjust that probability. This improved model could then be used to better determine the true time-savings of the course when delivered by GIFT.

During training evaluation and refinement, the disparity between predicted and observed training outcomes could be used to refine the training. For example, if a segment of training proves to be more difficult than anticipated for a group of learners, it is possible that the training segment should be refined or redeveloped.

An example of such a model was developed by McDonnell Douglas (1977). This model incorporated predictor variables in four broad categories: course content (e.g., difficulty, length of content), instructional design (e.g., instructional strategies/techniques), test characteristics (e.g., difficulty, number of items), and trainee characteristics (e.g., aptitude, motivation). The model predicted about 39% of the variability in trainee’s first-attempt lesson time for self-paced computer-based instruction.

To understand how GIFT might begin to model and predict training time for learners, it is necessary to understand how training is adapted by this system. GIFT is a framework that modularizes the common components of intelligent tutoring systems. These components include a learner module, an instructional or tutor module, a domain module, and a user interface. One of the main motivations for creating this framework was to lower the cost and labor needed to create intelligent tutoring systems by facilitating re- use of components and by simplifying the authoring process (Sottilare et al., 2012a).

GIFT adapts training using the learning effects model. At the first point of this model, learner data informs the learner state in the learner module. The learner module receives assessments from both sensors and the domain module. The learner state is used to determine the appropriate instructional strategy by the tutor module. The instructional strategy is then interpreted by the domain module and used to determine the domain specific learning activities needed to instruct the learner in that domain. The responses of the learner to that activity then update the learner module which starts the cycle over again.

Developing a predictive model in GIFT is not a straightforward process given the ways that training is adapted to each individual. We should note that our goal is not to predict the single path that a trainee would be expected to take through a specific course, but rather the probability associated with all possible paths through the training for a given learner. From that we can determine the range and distribution of times that would be expected for that learner to complete the training. Taking this one step further, we could apply this to a population of learners and predict the range and distribution of the time for that population to complete that training.

The development and integration of a probabilistic model for predicting time to train into the GIFT architecture is currently in the second phase of a three phase plan. Goodwin, Kim, and Niehaus (2017) reviews the approach and results of the first phase of this effort, which focused on the design and feasibility of these predictive models of tutor time to complete. In this paper, we describe work being done in the second phase. In the second phase, we are enhancing a predictive model for training efficiency and integrating this model with GIFT architecture, so that GIFT course creators can use these models directly with their GIFT tutors. In the third phase of the work, we will empirically validate the predictive model in GIFT and enhance the models with experimental and collected data.

METHODS

This section (1) reviews our method for modeling adaptive training content and predicting distributions of completion times for both individuals and groups using the GIFT excavator trainer as an example and (2) describes our approach for integrating these models with the GIFT architecture.

Modeling the Content of Adaptive Training

Predicting completion time for a tutor requires a model of the content and how the student can transition between the content. In GIFT, this transition logic is maintained in the Adaptive Course Flow object (formerly known as the Engine for Management of Adaptive Pedagogy – EMAP, e.g., Sottilare, 2014; Goldberg, 2015). It supports adaptive capabilities for training based on instructional strategies such as the Component Display Theory (CDT, Merrill, 1983). The CDT supports a general framework of skill training that progresses through two types of learning activities, each with two categories: expository (rules and examples) and inquisitory (recall and practice). According to Merrill, learners should progress through these four quadrants in order starting with rules (presentation of general principles), then to examples (presentation of a specific instance), then to recall (declarative knowledge test of the trainee's comprehension), and finally to practice (opportunity for the trainee to perform the skill). By sorting learning activities into these four quadrants, adaptive training systems like GIFT can apply the CDT to any domain as long as content for that domain is so labeled.

To model the content of adaptive training, we use the *Methodology for Annotated Skill Trees* (MAST) (Bauchwitz et al. 2018). In MAST, the "skeleton" of the skill tree breaks down entire procedures into constituent steps, tasks, and subtasks. Annotations are added to the procedure model. Figure 2 shows a portion of a MAST skill tree for an example training GIFT course, the excavator tutor. This skill tree focuses on the information elements that most heavily influence the completion time. On the left, the overall course on Excavator is the root of the tree structure. Its children are the different topics covered by the course, including the Boom Movement topic. This topic features a number of slides with Pictures, Audio, and Text Components. Individual trainees may vary in the amount of time they spend examining the Pictures, whether or not they listen completely to the Audio, and the amount of time taken to read the text. Trainees may also choose to view optional Slides explaining concepts that they may not be familiar with, adding more time. If trainees fail to demonstrate sufficient knowledge in the quiz or fail to complete the simulation tasks appropriately, they are sent back to the beginning of the Boom Movement topic on Slide 1, adding significant time to completion of the course. This model may be expanded to represent a maximum number of failures before the trainee either moves to a different topic or ends the course.

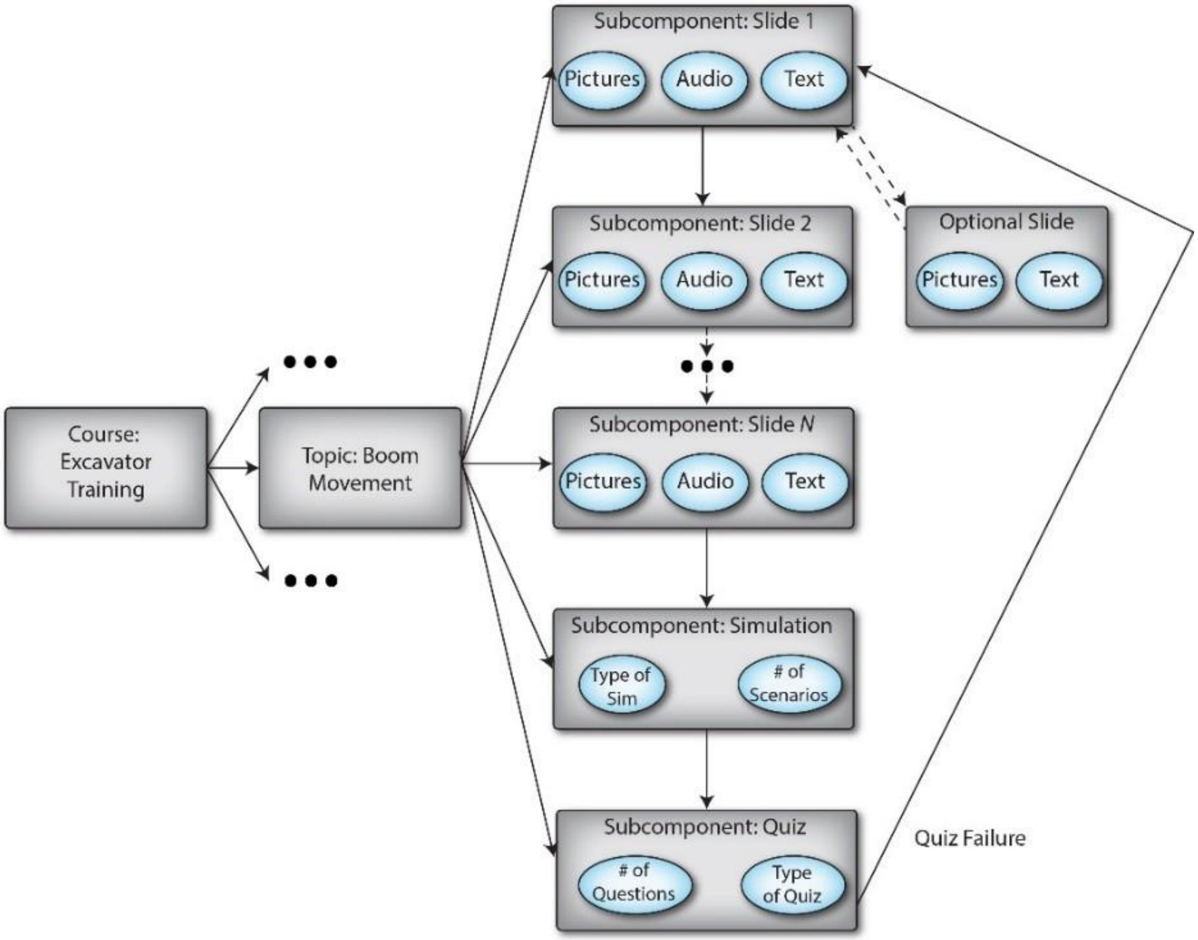


Figure 2: High-level design of a MAST skill tree of a GIFT module with representations of individual instructional elements, branching content, and variables that influence completion times.

After reviewing the Slides, the trainees are asked to practice their skills in Simulation. The MAST model of the simulation can be either a complex procedure describing the steps needed to complete the scenario and optional steps that may or may not contribute to the overall goal. The MAST simulation model may also be simple, representing just the type of simulation and the number of scenarios. To save modeling time and effort, these MAST models are constructed with only the level of detail needed to sufficiently and accurately predict the completion time.

Integrating with GIFT

To effectively predict completion time, we must combine models of students with models of the adaptive training content. We construct probabilistic models of students in the Figaro probabilistic programming language with key variables that influence their completion time of generic content. Figure 3 shows an example of one such set of variables. This student has a fatigue value to represent how tired they are. They have a read speed variable to represent how many words per minute they read under normal conditions. They have an expertise variable that represents how familiar they are with the concepts in the tutor. They have an effort level variable that represents how much effort they are putting into the training. They have an innate comprehension level that represents their general learning aptitude. They also have some status variables that record how many repetitions of different drills and quiz failures they have had during the course of training, for reporting purposes. These and similar parameters can be used to characterize the main student features that influence their completion time of training content. To be used in actual courses, these parameters must be learned and validated with real world data, such as records of previous students attempting a course.

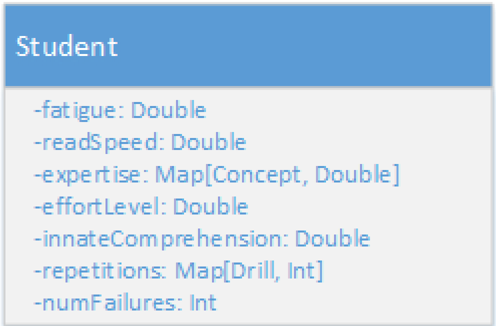

Student
-fatigue: Double
-readSpeed: Double
-expertise: Map[Concept, Double]
-effortLevel: Double
-innateComprehension: Double
-repetitions: Map[Drill, Int]
-numFailures: Int

Figure 3: UML for example Student model

Figure 4 shows a model of how a GIFT course can be represented as a set of learning material that the student must read or experience. At the top, the course is composed of multiple concepts. According to Merrill's CDT theory, each concept is taught by presenting a number of rules (on slides), examples (on slides), and quizzes to test rules, and exercises to test understanding of the examples. Each slide has a selection of media, which can include text, audio, and video, and is also rated for comprehensibility (e.g., more difficult slides take more time to comprehend). Quizzes are composed of a set of questions, which rules for how many must be answered correctly before the quiz is passed. Exercises are similarly composed of a set of drills with individual difficulties. Representing in the course in this way, along with the control logic that determines which piece of content the student is provided with next, enables the probabilistic modeling of the interaction between anticipated student populations and course content. It also enables the analysis of which content or sections of the course are contributing most to the completion time.

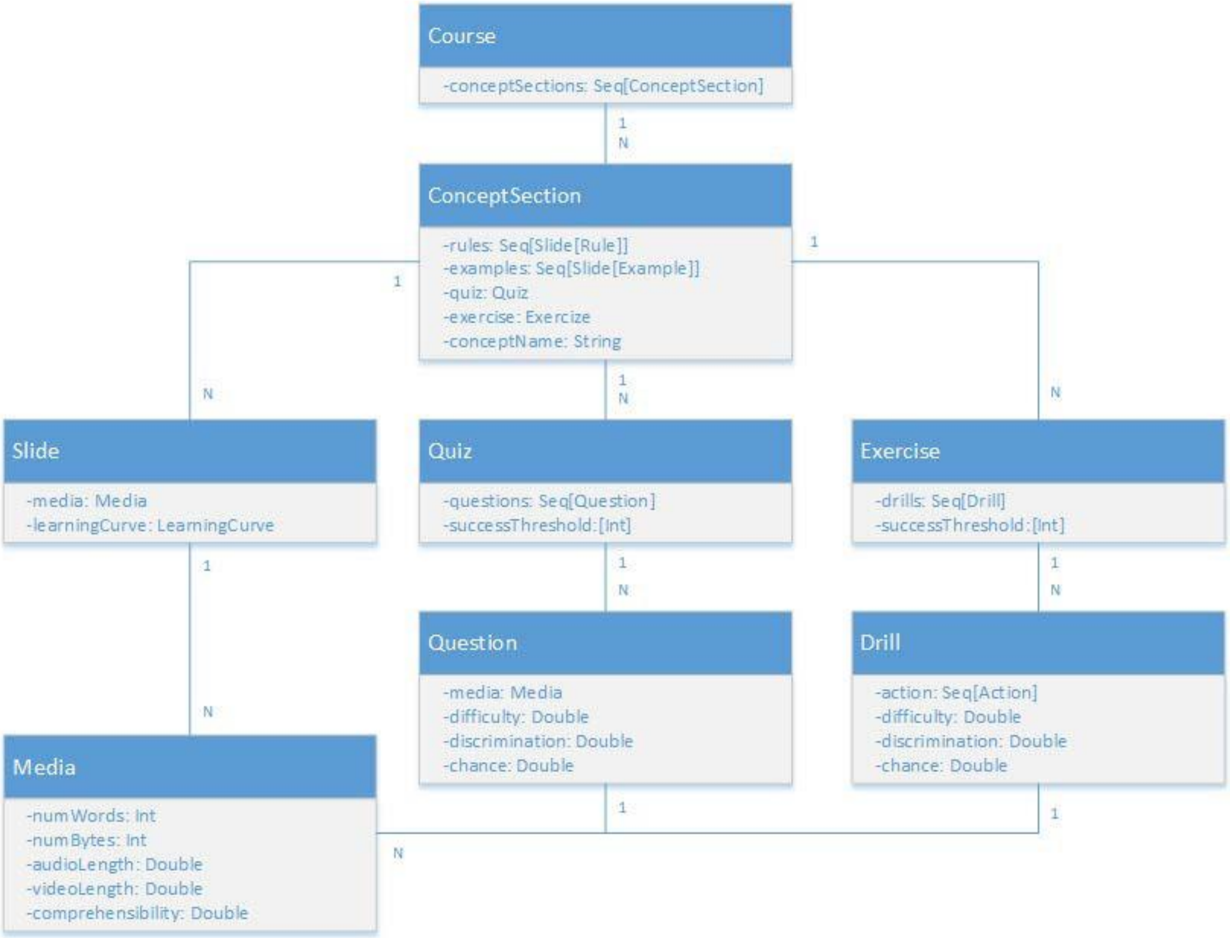
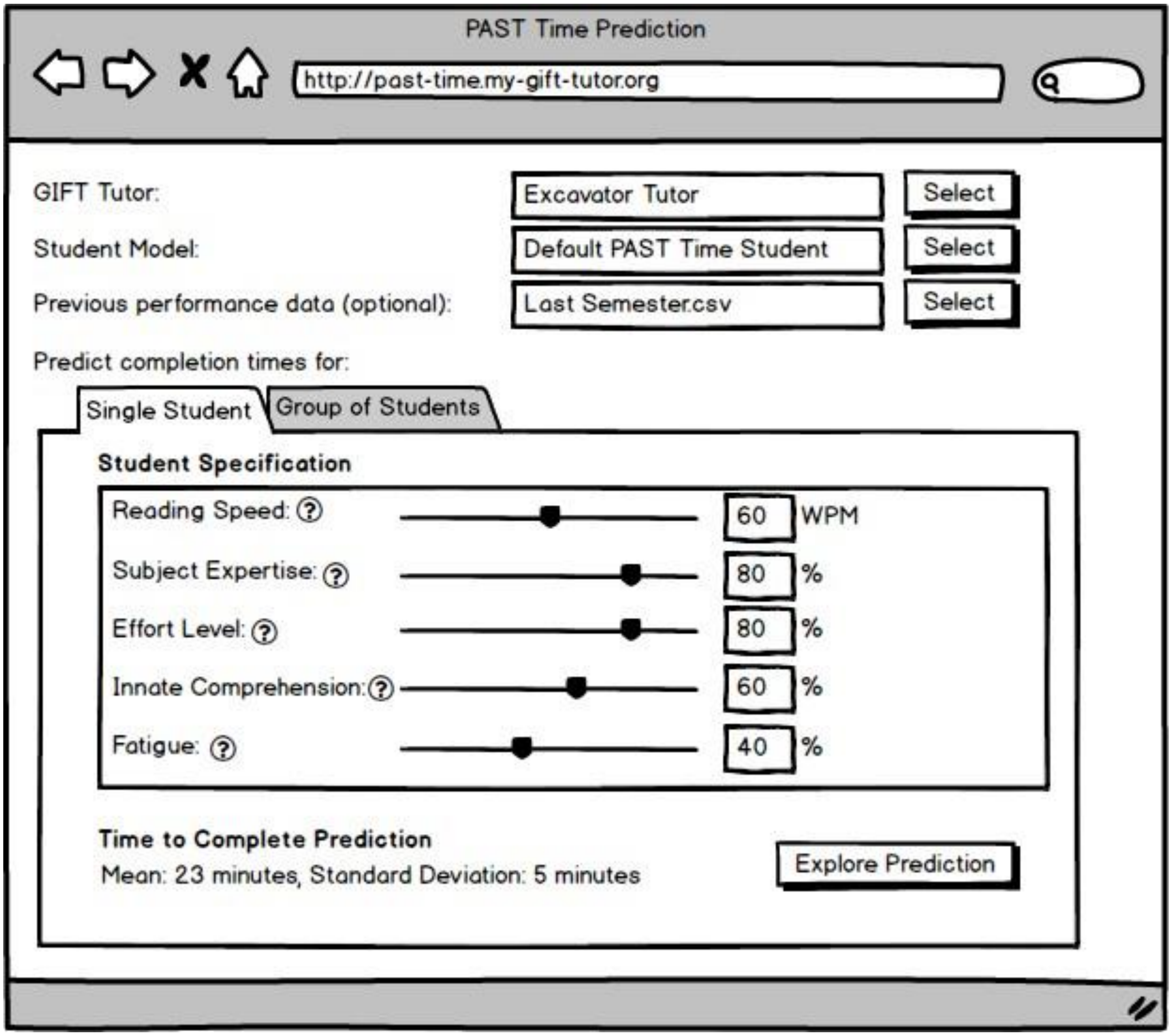


Figure 4: Partial UML diagram for PAST Time model of a Merrill's instructional theory GIFT Tutor

To effectively use these models, they must be integrated into the authoring cycle of adaptive training. Figure 5 shows a mockup of an interface to enable GIFT adaptive course authors to use these models for predicting completion time and understanding the impact of course design decisions on the ROI of adaptive training. At the top, the user specifies the GIFT tutor of interest, and which student model to use. The student model determines which parameters will be used to represent the student, such as those in Figure 3. The user is also presented with the option of including previous performance data to better tune the models to the population of interest.

To request a prediction, the user specifies a single student or a group of students according to the student model. In the single student case, this can be done by selecting exactly which values are set for each student parameter. In the group of students case, this can be done by specifying joint distributions of these values for the group of interest. Once these parameter values are specified, the models are executed and summary statistics of the prediction are presented, with the option for the user to explore the various components of the prediction.



GIFT Tutor: Excavator Tutor

Student Model: Default PAST Time Student

Previous performance data (optional): Last Semester.csv

Predict completion times for: Single Student

Student Specification

Reading Speed:	60	WPM
Subject Expertise:	80	%
Effort Level:	80	%
Innate Comprehension:	60	%
Fatigue:	40	%

Time to Complete Prediction

Mean: 23 minutes, Standard Deviation: 5 minutes

Figure 5: Mockup of PAST Time GIFT integration interface.

RESULTS

Implementing the Adaptive Training Models

This section presents a sample of the implementation of completion time models, and analysis of results of running these models with mock data. The probabilistic model is being implemented using Charles River Analytics' open source probabilistic programming language, Figaro™ (Pfeffer 2012), to construct and learn probabilistic models of the relationships between these factors. The use of Figaro greatly simplifies the authoring of these models which can be complex and require a high degree of experience by users who may not be experts in probabilistic reasoning.

Figure 6 shows an example Figaro function that predicts the completion time for a student reading a slide of information and updates the effect of reading the slide on the individual student. The reading time is computed by summing the media ingestion time of all the media on the slide (e.g., reading some text, looking at pictures, listening to audio, and watching video). The student's internal variables are then updated to reflect the effects of reading this slide; their fatigue is increased by a small amount and their expertise in the current concept is increased according to a specified function. The student is updated, the reading time is recorded, and the simulated student is then given the next piece of course content to complete.

```
def readSlide[T <: TrainingElement](slide: Slide[T], concept: Concept, student: Student): (Student,
Element[Double]) =

{

val readingTime: Element[Double] = mediaIngestionTime(slide.media, student)

val newFatigue = Math.min(1, Math.max(1.005*student.fatigue, 0.00000001))

val newExpertise = student.expertise |+| Map(concept -> expertiseIncrease[T](slide, con- cept, student))
```

Figure 6: Figaro function that models a student's reading time for a slide

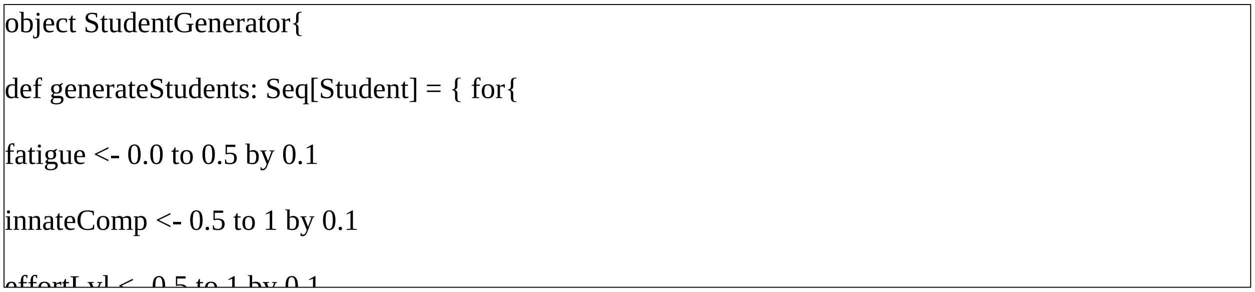
Figure 7 shows the model for a student taking a quiz as part of the adaptive training content. A quiz has multiple questions, which take time to complete. Based on the student's performance and the quiz passing threshold, the student may be sent to remediation for the current concept. In this function, the probabilities of success for each question are determined by the questions difficulty (currently, in classic item response theory) and the student's current aptitude at the concept. The reading time is given by the sum of the reading and thinking times for all the questions. The fatigue is updated by a marginal amount, and the new student is created. The probability of success on this quiz is also returned, so that Figaro can sample across the space of probabilities.

```
def takeQuiz(quiz: Quiz, concept: Concept, student: Student): (Student, Element[Double], El-
ement[Boolean]) = {

val probs: Seq[Element[Boolean]] = quiz.questions.map(q => probOfSuccess(q, concept, stu- dent))

val questions = Container(probs: _*)

val readingTime = Reduce((x: Double, y: Double) => x+y)(quiz.questions.map{q => mediaIn-
gestionTime(q.media, student)}: _*) // thinking time
```

Figure 7: Figaro function that models a student's completion time of a Quiz

Figaro probabilistic programming is useful in this context for a number of reasons: We can automatically build a model given a specification of the MAST skill tree, the trainee model, and a set of known relation- ships. Prediction based on the model is already coded in Figaro's inference algorithm, so additional effort is not required to use the model. Figaro supports the creation of dynamic Bayesian networks that model the temporal processes of variables, simulating fatigue and practice effects. We can continuously learn using these models; the probabilistic programs are flexible enough to update relationships between variables based on historical or dynamic data. Figaro's encapsulation mechanism enables easy creation of reusable components. Trainee models and MAST skill trees can be reused for future prediction models. It is embedded in a general purpose language, Scala, which allows the creation of front end graphical interfaces that can edit and invoke the models created in Figaro.

To test the current models, we created a set of mock simulated students to run on a mock tutor. Figure 8 shows the student sampling process, with samples across the space of initial fatigue, innate comprehension, effort level, and reading speed.



```
object StudentGenerator{

def generateStudents: Seq[Student] = { for{

fatigue <- 0.0 to 0.5 by 0.1

innateComp <- 0.5 to 1 by 0.1

effortLvl <- 0.5 to 1 by 0.1
```

Figure 8: Mock student generation for system testing

Figure 9 shows the results of this sample set on completion times. Because this is a sample to test model dynamics, the completion time units on the y-axis are arbitrary. In this plot, there is a combined effect of the main variables. In the left, when fatigue is low, low effort and comprehension only make a moderate difference in completion time. On the right, when initial fatigue is high, low effort and comprehension are compounded, as the mock students begin failing quizzes and drills, which causes them to repeat content, which causes them to become more fatigued. At the far right, several samples included one or more students that exceeded our modeling time limit, causing them to be marked as 0 for the purposes of this graph. This effect can be seen in Figure 10 where the relationship between failures and completion time is exponential due to the compounding fatigue factor.

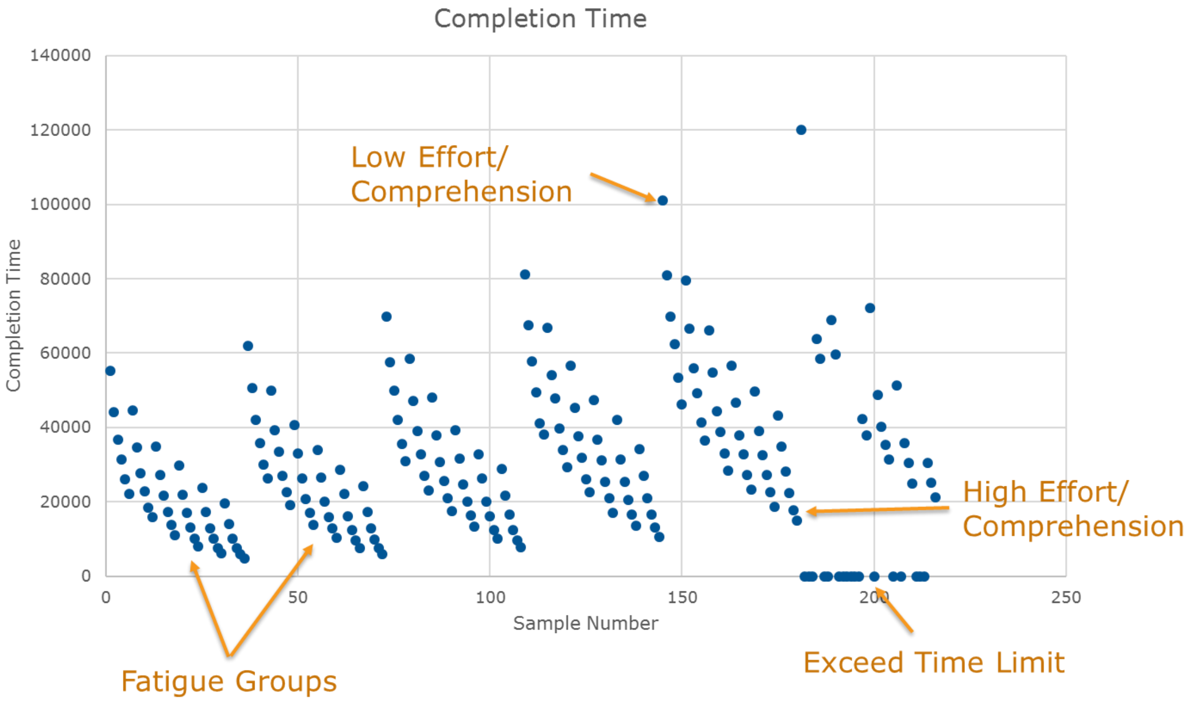


Figure 9: Completion time of mock students, sampled across the space

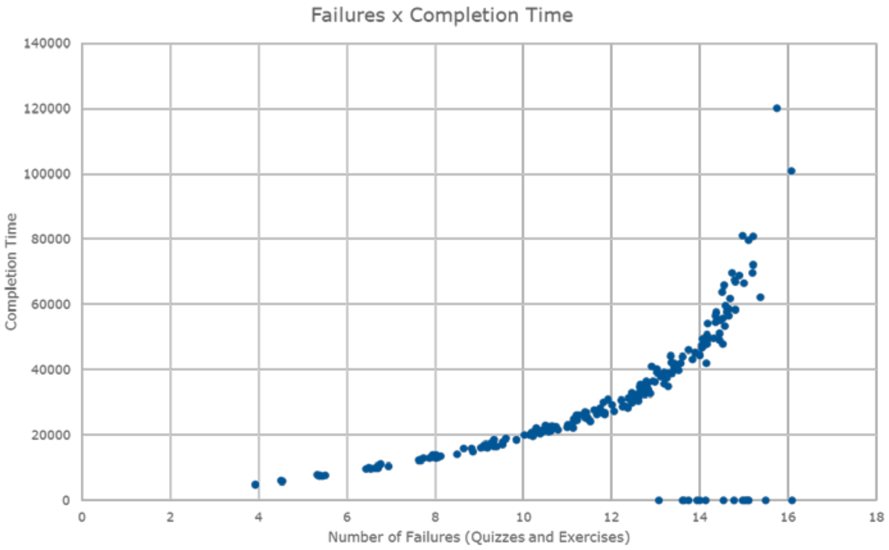


Figure 10: Relationship between failures on quizzes and drills and completion time of mock students

Models like these enable adaptive training course authors to quickly explore what-if scenarios with ranges of students and different configurations of adaptive content. The third phase of our effort will focus on learning actual student models from empirical data, enabling us to calibrate these models to actual performance, realistic completion times, and identify which observable and latent variables are most valuable in this prediction.

DISCUSSION

We believe that including a capability to predict training time for trainees in GIFT has several significant advantages for accelerated learning. First, it facilitates return on investment (ROI) calculations by enabling the author to determine training time reductions resulting from the addition of adaptive features. Second, it provides a means for GIFT to monitor student progress against an expected timeline. Students who take much longer to complete training than expected may not be fully engaged in the training or may be having difficulty with the material. These are conditions that might prompt a response by GIFT. Finally, it can play a role in quality control of GIFT courses. For example, if segments of a course take much longer than expected across multiple trainees, GIFT could flag those sections for review by the course author to insure that the material is presented clearly.

Determining the ROI for training is not always easy. As Fletcher and Chatham (2010) put it, how does one determine the benefit of a pound of training? In some cases it may be fairly straight forward. For example, one might measure the increase in revenue produced by the introduction of new training for a sales staff. While this may work for commercial businesses, the military is not a profit making organization, therefore one must look at other factors like cost avoidance to get a measure of ROI.

Determining this can be quite difficult as one rarely has before and after data on the operational impact of training. In rare cases it can be found. For example, Fletcher and Chatham (2010) examined the benefits of Top Gun training given to pilots during the Vietnam war. Because of this training, kill ratios of Navy pilots improved from 2.4 enemy kills per loss up to 12.5 enemy kills per loss. The authors determined that the training had reduced U.S. losses by about 10-12 aircraft during the war When they looked at the cost of procuring and employing that many aircraft during the war, they calculated that the training had saved the Navy about $132 million dollars for an ROI of about 2.5.

Determining the ROI for adaptive vs. non-adaptive training in terms of cost avoidance measures in an operational context would be very difficult. Adaptive training is still relatively new and opportunities to do side-by-side comparisons with traditional non-adaptive training are virtually non-existent. Rather than try- ing to quantify an impact in the operational environment however, we can look at the impact in a training environment. Specifically, one of the key advantages of adaptive training would be to reduce the overall time needed to deliver the training to a population of trainees.

A challenge for authors of adaptive training is determining how *adaptive* the training should be. While adding adaptive features can potentially save training time, it also increases the cost of development. How does one determine, when the training is adaptive enough? Using an ROI metric can help to answer this question. On one hand is the cost of adding the adaptive feature. On the other hand is the value of the time saved by that adaptive feature. The value of that time could be calculated by looking at the total salary paid to the trainees over that time (e.g., 1,000 trainees/year x .5h/trainee x $35/h = $17,500/year). So, as long as the cost of adding the adaptive feature was less than value of the time saved, there would be a positive ROI and therefore justification for adding that particular adaptive feature.

As can be seen, our model supports this strategy for the design and development of adaptive training in GIFT by helping to predict the effect of adaptive features on the training time for a known population of learners.

There are several challenges we may face as we develop this model. First, the initial MAST skill tree may not contain sufficient variables to predict adaptive training completion times. Our initial literature review and analysis have identified a potential set of most influential variables, but these variables may not be reflective of the completion time upon closer inspection. We will mitigate the identified risk by widening the scope of task models to incorporate more predictive variables if necessary.

Second, while the model predictions may be highly accurate, there is a risk that the system will be too difficult or time consuming to use for some or all of the target populations of instructional designers, course managers, and instructional staff. We mitigate this risk by conducting a requirements analysis early in the effort to closely examine the needs of these user groups and design our system and interfaces to best meet those needs. We will apply human factors and user-centered design and understand the challenges of and methods for developing highly useful and usable decision-aiding tools for practitioners.

Third, while this approach combines state of the art probabilistic approaches and identifies key variables from the literature and past experience, there is a potential that the initial predictions will not sufficiently account for the variability of trainee completion times. We plan to mitigate this risk by incorporating historical data early and adjusting the analysis techniques to capture the maximum amount of variability from data that can be reasonably collected in the field.

When complete, this will be the first system to predict the completion times of GIFT and to enable effective assessments of the ROI that is useful for key design and implementation decisions of an adaptive training system. It includes an innovative application of the procedure skill modeling the MAST skill tree to flexibly represent the adaptive training content for analysis. It is the first application using a probabilistic programming language (i.e., Figaro) to predict completion times for adaptive training technologies, including both unobserved latent variables and temporal factors, such as trainee fatigue, boredom, or flow.

REFERENCES

Army Research Laboratory. (2015). GIFT 2015-1, Generalized Intelligent Framework for Tutoring Release Page.

Retrieved from http://www.gifttutoring.org

Bauchwitz B, Niehaus J, Weyhrauch P. Modeling and Comparing Nurse and Physician Trauma Assessment Skills.

Military Medicine, Volume 183, Issue 1, 2018, Pages 47–54,

Dodds P, Fletcher JD. Opportunities for new “smart” learning environments enabled by next generation web capabilities. Journal of Educational Multimedia and Hypermedia. 2004;13:391–404.

Dynarsky M, Agodini R, Heaviside S, Novak T, Carey N, Camuzano L, Sussex W. Effectiveness of reading and mathematics software products: findings from the first student cohort; March Report to Congress 2007. [accessed 2015 May 15]. http://ies.ed.gov/ncee/pdf/20074005.pdf.

Fletcher JD. Evidence for learning from technology-assisted instruction. In: O’Neil HF, Perez R, editors. Technology applications in education: a learning view. Mahwah (NJ): Erlbaum; 2003. p. 79–99.

Fletcher, J. D., and R. E. Chatham. 2010. “Measuring Return on Investment in Military Training and Human Performance.” In Human Performance Enhancements in High-Risk Environments, edited by J. Cohn and P. O’Connor, 106–28. Santa Barbara, CA: Praeger/ABC-CLIO.

Goodwin GA, Kim JW, Niehaus J. Modeling Training Efficiency and Return on Investment for Adaptive Training. In Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5) 2017 Jul 17 (p. 229). Robert Sottilare.

Goldberg, B., & Hoffman, M. (2015). Adaptive course flow and sequencing through the engine for management of adaptive pedagogy (EMAP). In *Proceedings of the AIED Workshop on Developing a Generalized Intelligent Framework for Tutoring (GIFT): Informing Design through a Community of Practice* (pp. 46- 53). Madrid, Spain.

Graesser AC, Conley M, Olney A. Intelligent tutoring systems. In: Harris KR, Graham S, Urdan T, editors. APA Educational Psychology Handbook: vol. 3. Applications to Learning and Teaching. Washington (DC): American Psychological Association; 2012. p. 451–473.

McDonnell Douglas Corporation. A survey and analysis of military computer-based training systems: (A two part study). Vol II: A descriptive and predictive model for evaluating instructional systems; 1977. Defense Advanced Research Projects Agency. [accessed 2015 October] http://www.dtic.mil/dtic/tr/fulltext/u2/a043358.pdf.

Merrill, M. D. (1983). Component display theory. In C. M. Reigeluth (Ed.), *Instructional-design theories and models: An overview of their current status* (pp. 282-333). Hillsdale, NJ: Lawrence Erlbaum Associates.

Olney AM, Person NK, Graesser AC. Guru: designing a conversational expert intelligent tutoring system. In: Boon-thum-Denecke C, McCarthy P, Lamkin T, editors. Cross-disciplinary advances in applied natural language processing: issues and approaches. Hershey (PA): Information Science Publishing; 2012. p. 156–171.

Pfeffer, A. (2012). Creating and manipulating probabilistic programs with Figaro. Workshop on Statistical Relational Artificial Intelligence (StarAI).

Sottilare R, Brawner KW, Goldberg BS, Holden HK. The generalized intelligent framework for tutoring (GIFT). Orlando (FL): Army Research Laboratory (US); Human Research and Engineering Directorate (HRED); 2012a [accessed 2015 May]. https://gifttutoring.org/attachments/152 /GIFTdescription_0.pdf.

Sottilare R, Goldberg BS, Brawner KW, Holden HK. A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In: Proceedings of the Interservice/Industry Training Simulation and Education Conference; 2012 Dec 3–6; Orlando, FL. Arlington (VA): National Defense Industrial Association; 2012b.

Sottilare, R. A. (2014). Using learner data to influence performance during adaptive tutoring experiences. In *Proceedings of International Conference on Augmented Cognition--HCII2014* (pp. 265-275). Crete, Greece.

Springer.Steenbergen-Hu S, Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on K-12 students’ mathematical learning. Journal of Educational Psychology. 2013;105(4):970–987.

Steenbergen-Hu S, Cooper H. A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning. Journal of Educational Psychology. 2014;106:331–347.

VanLehn K, Graesser AC, Jackson GT, Jordan P, Olney A, Rosé CP. When are tutorial dialogues more effective than reading? Cognitive Science. 2007;31(1):3–62.

VanLehn K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems.

Educational Psychologist. 2011;46 (4):197–221.

ABOUT THE AUTHORS

***Dr. Gregory Goodwin** is the acting Branch Chief of the Creative and Effective Technologies Branch at the Army Research Laboratory – Human Research and Engineering Directorate in Orlando, FL. For the last decade, he has worked for the Army researching ways to improve training methods and technologies. He holds a Ph.D. in Psychology from Binghamton University and an M.A. in Psychology from Wake Forest University.*

***Dr. James Niehaus** is a senior scientist at Charles River Analytics. Dr. Niehaus’s areas of expertise include artificial intelligence, training systems, and health technology. Dr. Niehaus has a B.S. in computer science from the College of Charleston and a Ph.D. in computer science from North Carolina State University.*

Personality: A Key to Motivating our Learners

Elizabeth Biddle[1], Elizabeth Lameier[2], Lauren Reinerman-Jones[2], Gerald Matthews[2], Michael Boyce[3]
The Boeing Company[1], University of Central Florida[2], Army Research Laboratory[3]

INTRODUCTION

The Motivation Assessment Tool (MAT), currently in development (Lameier et al, pending) will assess a learner's motivation profile and provide instructional guidance via the Generalized Intelligent Framework for Tutoring's (GIFT's) authoring tool to enable an instructor to design a lesson that will personalize the learner's experience to support and/or improve their motivation. Specifically, the MAT will provide a methodology for personalizing learning in GIFT. Learner motivation is influenced by a variety of traits and factors, which include student personality, learning performance history, intrinsic vs. extrinsic motivation tendencies, and self-regulatory skills (Reinerman-Jones et al., 2017). Understanding a learner's composition of these traits is key to tailoring the instructional environment to support and encourage learner motivation. Intelligent tutoring systems provide a learning environment in which it is possible to seamlessly assess and tailor instruction to support the learner's motivation. The goal for the MAT is to develop a method for identifying the motivational dispositions of GIFT learners. In turn, assessments using the MAT may guide methods for personalizing training to capitalize on the learner's motivational profile with the outcome being improved mastery and retention. This paper will describe an effort in support of the MAT's development and validation to determine how strongly a learner's personality profile describes their motivation. After discussing the analysis of the personality relation to the MAT, the paper will then focus on how the MAT will be implemented in GIFT and the associated benefits and the barriers.

Motivation in Learning

Motivation has been defined as being "moved to do something" (Ryan & Deci, 2000) and is essential to learning (Keller, 1987). When in a motivated state, a learner is inclined to initiate a task and persevere throughout its completion. As a result, motivation increases an individual's level of engagement (Magill, 1980). When learners are not motivated, they are more likely to disengage from the task. Motivation can be classified into two types (delSoldato & duBoulay, 1999; Kember, Wong, & Leung, 1999; Noels, Clement, & Pelletier, 1999): (1) intrinsic motivation, which refers to an individual's internal desire to achieve, and (2) extrinsic motivation, which refers to external rewards that encourage an individual to achieve. Both intrinsic and extrinsic motivation are approached by the MAT as traits. However, intrinsically motivated individuals rely on self-regulatory processes and internally driven incentives, whereas extrinsically motivated individuals need an instructor or automated learning environment to influence their motivation throughout learning.

Personality in Learning

An individual's personality traits influence their cognitive, affective, and motivational processes (Matthews & Zeidner, 2004; Blickensderfer et al, 2003). Consequently, a learner's personality profile will affect their reaction and experience with different learning environments and strategies (Komarrajuq et al, 2011; Costa & McCrae, 1992). The Big Five model (Goldberg, 1981) is one of the most commonly used personality theories. The five traits are: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Extraversion is related to an interest in social events, talking with others and interaction with environments. Agreeableness refers to a cooperative disposition with little interest in competition, a willingness to put others first, being compliant, and trusting others. Conscientiousness refers to behaviors that focus on attention to details, organization, and being goal-directed. Neuroticism describes a tendency to exhibit negative emotions such as stress, anxiety, irritability, or sadness due to a predisposition to perceiving the environment as negative or threatening. Openness describes an inclination to explore, try and learn new things, and enjoy intellectual and creative activities.

The "Big Five" personality traits have been connected to behaviors, academic achievement, and job performance (Judge et al., 2007; Larson et al., 1990). Further, 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). Komarraju and Karau (2009) found that Conscientiousness was the most influential trait and had positive correlations with intrinsic motivation and high GPA, while correlating negatively with extrinsic motivation and amotivation. They also obtained evidence that individuals with: 1) high intrinsic motivation also had higher tendencies towards Openness, 2) high Neuroticism was seen to have a higher amotivation, and 3) high Extraversion was more closely associated with extrinsic motivation. The authors obtained similar results in a later, related study (Komarraju et al., 2011).

MAT WAVE 3 ANALYSIS

The MAT was developed to evaluate the multiple variables that influence a learner's motivation to increase the precision in providing learning in GIFT tailored to the learner's needs. The MAT has been constructed with two sections: 1) General Motivation, to assess the learner's motivation traits, and 2) Motivator Inventory, to determine the optimal reinforcers that motivate the individual learner (see Table 1). The MAT development has undergone three waves of data collection and analysis. The first wave evaluated the original iteration of the MAT, which was created by combining and clustering items from prior motivation assessments, which each addressed a limited set of motivation variables (Reinerman-Jones et al, 2017). Additional items were created and included in this first iteration to evaluate the types of reinforcers that support an individual's motivation. In the second iteration, two scales for items important to motivation, attitudes and autonomy, were added to ensure these constructs were addressed by the MAT (Lameier et al, pending publication). This paper focuses on the third wave analysis, which was used to: 1) check reliability and factor structure, and provide the final refinement to the MAT, and 2) evaluate relationships between the MAT scales and the Big Five personality traits.

Table 1. MAT Scales

General Motivation		Motivator Inventory Scales	
1. Attitudes	10. Workload	1. Feedback	10. High-value
2. Learning Driven	11. Organize and Structure	2. Intrinsic feedback	11. Self-reward
3. Autonomy	12. Social	3. Extrinsic feedback	12. Activity
4. Goal Orientation	13. Breaks	4. Recognition	13. Time

5. Loss of Effort	14. Extinction	5. IMI	14. Sensors
6. Worry	15. Relatedness	6. Digital	15. Hobbies
7. Freeze, Fear, Flight	16. Effort Based on Punishment	7. Energizer	16. Time During learning
8. Competition	17. Positive outlook	8. Logical Consequences	17. Negative Time
9. Challenge	18. Self-regulation	9. Low-value	18. Activity

Participants

For the wave 3 analysis, 249 participants (112 females, 137 males) were recruited through Amazon Mechanical Turk, with ages ranging from 19 to 71 years.

Materials and Procedures

The participants completed the MAT, along with the Big Five Aspect Scales (BFAS; DeYoung et al, 2007) to assess the Big Five personality traits, the Reinforcement Sensitivity Theory of Personality Questionnaire (RST-PQ; Corr & Cooper, 2016), and three assessments that evaluate aspects of motivation, which were the Portrait Value (Schwartz & Butenko, 2014), Grit and Ambition scale (Duckworth, 2009) and the 3x2 Achievement Goal scale (Elliot, Murayama, & Pekrun, 2011). The MAT contained 293 items across both sections of the MAT (general and motivator inventory). All of the questionnaires, including the MAT, were loaded into GIFT as evaluations. When the participants accessed Mechanical Turk, they were taken to GIFT via a weblink to complete the questionnaires. This paper is only addressing the evaluation of the relationships between the MAT and personality, while the evaluation of the MAT against the constructs evaluated by the other tools (e.g., grit, value, goal orientation) will be reported in subsequent publications.

Results

First, Cronbach α coefficients were calculated to evaluate the internal consistency of the scales identified in Table 1. The coefficients ranged from .794 (relatedness) to .955 (Freeze, Fear, Flight) in the General Motivation section and .791 (Intrinsic Feedback) to .935 (Recognition) in the Motivator Inventory indicating that each of the scales generally had satisfactory internal consistency.

Table 2. BFAS Scales

Big Five Personality Trait	Trait Dimensions
Neuroticism	• *Volatility* – tendency for extreme variability in response to external environment

	• *Withdrawal* – tendency to focus inward
Agreeableness	• *Compassion* – interest in feelings of others • *Politeness* – tendency to treat others with respect
Conscientiousness	• *Industriousness* – tendency to work hard to complete tasks • *Orderliness* - detailed and organized nature
Extraversion	• *Assertiveness* – tendency to dominate social interactions • *Enthusiasm* – tendency to exhibit energy and positive attitude
Openness	• *Openness* – willingness to explore new ideas and activities • *Intellect* – tendency to synthesize information to guide, objective decision making

To simplify the analysis, an exploratory factor analysis was conducted to determine the higher-order factor structure of the MAT General Motivation scales. A principal factor method was used for factor extraction, followed by oblique rotation using the direct oblimin criterion. The three primary factors identified were Social (need for social interaction and competition), Self-Directed Learning (ability to keep on task and determine appropriate use of time to complete learning tasks), and Threat Vulnerability (tendency to become anxious or stressed during learning). On the basis of the scree test and parallel analysis three factors were extracted, explaining 65 % of the variance. The range of factor correlations was maximum of .891 and minimum of .508. The three factors were then scored by their mean. The Social factor included scales from challenge, extinction, competition, relatedness, social link, and punishment. The Self-Directed Learning is composed from the autonomy, positive outlook, self-regulation, organized structure, and break scales. Threat Vulnerability has loss of effort, workload, worry, and fear-freeze-fight scales. A similar process was performed for the MAT Motivator Inventory and two primary factors were identified. The factors identified for the Motivator Inventory were Motivator (preference for specific type of reinforcer) and High Value Motivator (preference for reinforcer of high value). Based on the scree test and parallel analysis two factors were extracted, explaining 59% of the variance. The motivator scale was created by intrinsic feedback, extrinsic feedback, acknowledgement, digital, energizer, logical consequence, low value, self, activity, sensor, hobby, level of interactivity, and time during learning scales. The high value factor was the only factor that loaded on the high value scale. The range of factor correlations was a maximum of .814 and a minimum of .478.

Table 2. Higher Order Factors

	Primary Factors	Correlations (*r)* with Personality Traits and Facets
General Motivation	Social	• Neuroticism (.204) • Agreeableness (-.276)

		○ Politeness (-.407)
	Self-Directed Learning	• Openness (.554) ○ Intellect (.529) ○ Openness (.478) • Agreeableness (.465) ○ Compassion (.456) ○ Politeness (.373) • Neuroticism (-.335) ○ Withdrawal (-.311) ○ Volatility (-.322) • Conscientiousness (.457) ○ Industriousness (.403) ○ Orderliness (.384) • Extraversion (.373) ○ Enthusiasm (.396) ○ Assertiveness (.253)
	Threat Vulnerability	• Neuroticism (.730) ○ Withdrawal (.714) ○ Volatility (.663) • Openness (-.485) ○ Intellect (-.575) ○ Openness (-.226) • Agreeableness (-.437) ○ Politeness (-.386) ○ Compassion (-.398) • Conscientiousness (-.467) ○ Industriousness (-.650) • Extraversion (-.438) ○ Enthusiasm (-.418) ○ Assertiveness (-.340)
Reinforcer Inventory	Motivator	• Extraversion (.321) ○ Enthusiasm (.300) ○ Assertiveness (.257) • Openness (.234) ○ Openness (.237) ○ Intellect (.201)
	High Value Motivator	• Openness (.371) ○ Openness (.371) ○ Intellect (.322) • Conscientiousness (.329) ○ Orderliness (.357) ○ Industriousness (.214)

		• Agreeableness (.347) ○ Compassion (.311) ○ Politeness (.310) • Extraversion (.164) ○ Enthusiasm (.201)

A bivariate correlation analysis was conducted to examine the relationships of the other motivation assessments' scales and the IPIP scales. Table 2 provides the Pearson correlation coefficients (*r*) for IPIP scales for each of the MAT higher order factors. There were notable correlations for the primary factors for the both the General Motivation and Motivator Inventory sections of the MAT.

Threat Vulnerability demonstrated the strongest correlations with the Big Five personality traits and the 10 facets. The strongest, and only positive correlation, was with Neuroticism (*r* = .730) and its facets Withdrawal (.714) and Votility (.663). This type of learner would view the learning environment as intimidating making it difficult for the learner to maintain motivation due to feelings of hopelessness and likely have random reactions based on their successes and failures during the learning process. Threat Vulnerability was negatively correlated with the other 4 Big Five personality traits, indicated that this type of learner is most influenced by their predisposition to interpret their learning environment negatively. The strongest negative correlations were with the facet of Industriousness (Conscientiousness, *r* = -.650) and Intellect (Openness, *r* = -.575), reflecting a learner that is not productive due to their tendency to worry and their reluctance to experience new activities and experiences.

Self-Directed Learning was positively correlated with Openness (*r* = .516) and its two facets, Openness (*r* = .592) and Intellect (*r* = .470). Given that Self-Directed Learning refers to an individual with an intrinsic motivation tendency and ability to complete learning tasks on their own, it makes sense that this type of individual would be open to new ideas and experiences. Self-Directed Learning was also correlated with Conscientiousness (*r* = .457) and its facet Industriousness (*r* = .403), which is indicative of the focus and follow-through a self-directed learner would need. Finally, Self-Directed Learning was also positively correlated with Agreeableness (.465) and its facet, Compassion (.456). While a student who can work autonomously does not require social skills, the relationship may be explained that this type of student is not threatened or competing with other students.

While neither the trait of extraversion nor its dimensions were correlated with Social, the Agreeableness facet of Politeness was negatively correlated with Social. This may indicate that while some learners need interaction with others to be motivated to learn, they are not necessarily interested in the other students' well-being, but having interaction with other learner.

The Motivator Inventory demonstrated weaker correlations with personality. The Motivator scale had weaker correlations – primarily with Extraversion, which can be explained by their need and higher threshold for, stimulation from the external environment. The High Value scale had week correlations with all traits except Neuroticism. The lack of any correlation between the Motivator Inventory scales and Neuroticism is interesting and may suggest that it is the learning environment, interactions and feedback style, which is most important to motivating this learner type, rather than an externally provided reward.

The results of this study indicate a learner's personality trait composition is related to their motivation trait composition. Identifying a learner's personality composition can provide insights that will support the provision of instruction that is tailored to optimize the learner's motivation. Specifically, personality trait identification can help determine whether the learner is intrinsic and able to learn independently or whether the learner is going to need positive support and encouragement.

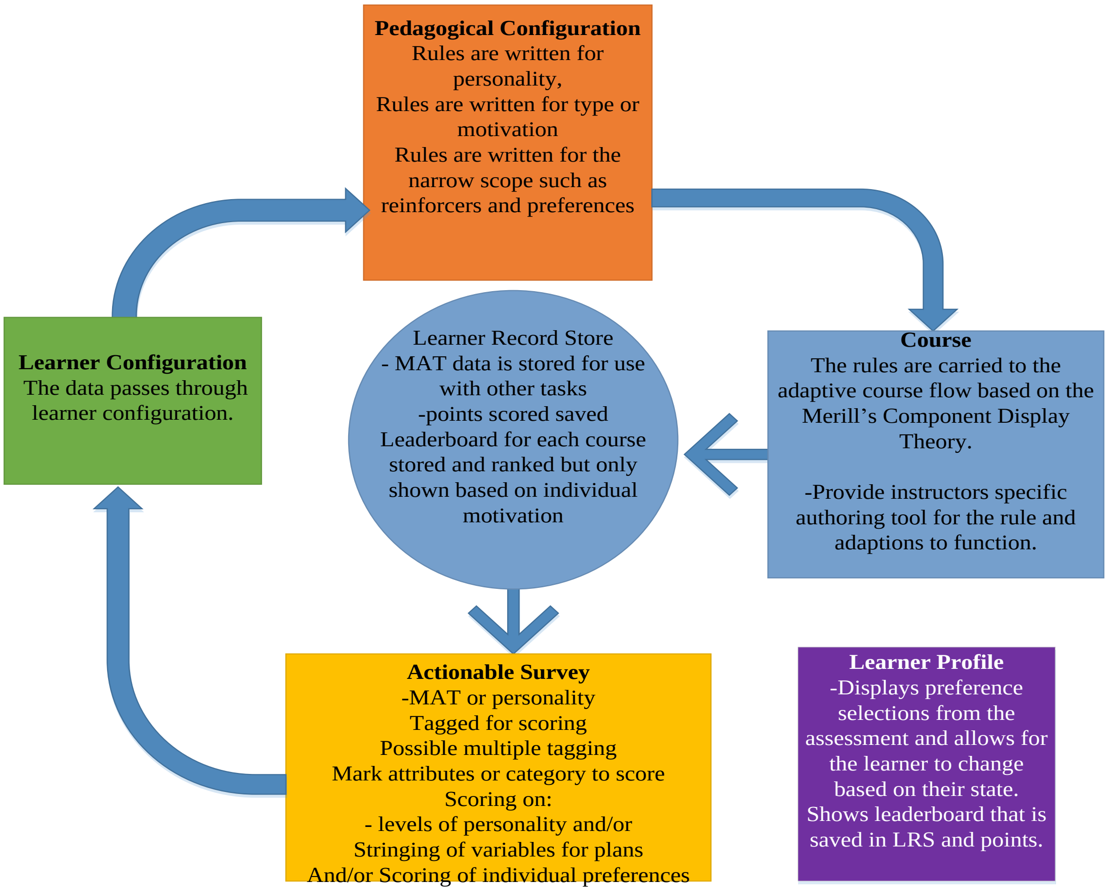
MAT IMPLEMENTATION IN GIFT

The final version of the MAT will be implemented as an actionable survey within GIFT with its implementation functionally aligned with the pedagogical module and long-term learner module (LTLM). Currently, actionable surveys in GIFT use the results of the survey to immediately update the learner model and the pedagogical model, which results in a course adaptation. An actionable survey is scored based on the tags authored and attached to the concepts addressed by the individual survey questions to create the logic for scoring the survey. The information collected from the survey is sent to the learner model (found in advanced settings) and the scores for the concepts are updated.

Implementation of the MAT will follow the process described above. However, rather than designating a learner as a novice, journeyman or expert, or high or low motivated, the resulting adaptations will be designed to implement a Learner Plan, which will be further described, that will support the learner's motivation. Furthermore, the results of the MAT will be stored in the LRS and use to select the optimal Learner Plan when the student enters GIFT and launches a lesson. Figure 1 depicts how the MAT actionable survey will be implemented within GIFT.

MAT Actionable Survey Implementation

The final state of the MAT will be shaped by the results from the present study, as well as the planned verification experiment, which will evaluate the effectiveness of the Learner Plans based on MAT assessment to improve or maintain motivation and learning effectiveness. For implementation within GIFT, the MAT will be created as an actionable questionnaire. Currently, the MAT is divided into sections based on groupings the ITS would need to know such as intrinsic motivation, level of effort, affective tendencies, comparing/competitiveness, task (preference and strategies), reward orientation, and motivator inventory. Extrinsic tendencies will be scored from the reverse of the intrinsic tendencies.



Pedagogical Configuration
Rules are written for personality,
Rules are written for type or motivation
Rules are written for the narrow scope such as reinforcers and preferences

Course
The rules are carried to the adaptive course flow based on the Merill's Component Display Theory.

-Provide instructors specific authoring tool for the rule and adaptions to function.

Learner Record Store
- MAT data is stored for use with other tasks
-points scored saved
Leaderboard for each course stored and ranked but only shown based on individual motivation

Learner Configuration
The data passes through learner configuration.

Actionable Survey
-MAT or personality
Tagged for scoring
Possible multiple tagging
Mark attributes or category to score
Scoring on:
- levels of personality and/or
Stringing of variables for plans
And/or Scoring of individual preferences

Learner Profile
-Displays preference selections from the assessment and allows for the learner to change based on their state.
Shows leaderboard that is saved in LRS and points.

Figure 1. MAT Implementation in GIFT

Based on the results of the analysis presented earlier, as well as the wave 1 and 2 analyses, the final version of the MAT will likely be reduced, focusing on the higher level scales such that only a few Learner Plans may be required, such as Intrinsic/Self-Learner, Threat Vulnerable and Social. For Social, there may be two different plans – one focused on challenge and one focused on reward. For example, the cumulative scores for these higher-level scales will be made actionable by having specific delivery and pedagogy preferences associated with each scale. A tag will be set to score the various sections from the actionable survey (experiment dependent). For example, Tag 1 would be scored with the Intrinsic (Self-Learner) Learner Plan such that an intrinsic learner's correlates are with a set of variables that need to be scored throughout the assessment and not just based on a few questions measuring one attribute. Tag 2 might be tied to the Extrinsic Learner Plan such that the extrinsic learner will need to provide the personality type (Scenario Developed below for further explanation) to help determine whether they are Threat Vulnerable or Social for instance. Tag 3 might be with additional MAT (e.g., challenge, breaks) or Motivator Inventory sections that will further guide the Social Learner Plan to accurately provide the type of schedule, level of support needed, and so forth (yet to be determined based upon the verification results and synthesis of the wave 1-3 analyses).

The results of the MAT Actionable Survey will need to be stored into the LTLM rather than feed real-time into the pedagogical configuration of the lesson. The next section discusses the LTLM implementation.

LTLM Implementation

In the current version of GIFT, if a lesson has been implemented using an actionable survey, the data is then immediately used to configure the student's lesson in run-time. Therefore, the learner must complete the survey each time he or she takes a lesson Trait information, such as the type of data being obtained by the MAT, is generally fixed for long durations. Repeating the survey each time the learner completes a lesson results in collecting the same data and will serve to demotivate the student. Therefore, we are recommending that the student be asked to take the MAT Actionable Survey the first time they log into GIFT and have the results saved to the long-term learner model (LTLM). Rather than the results directly feeding the pedagogical module during the run-time configuration of the lesson, it can pull the data from the LTLM based on the student's login.

The authors are anticipating that the LTLM will be implemented with a learner record store (LRS). Given the goal of GIFT is to include a LTLM that provides a historical learner model that contains previous learning experience data, as well as data pertaining to individual differences in learning, this project is planning to leverage this future capability. In this way, the LTLM will be used to tailor the pedagogy and delivery mechanisms without requiring the learner to complete surveys each time they enter GIFT to complete a lesson. However, the learner will need to retake the assessment after a period, such as a year, or for major life events that could jeopardize the stability of the trait. Additionally, we recommend that the learner have the option to retake the MAT or other relevant survey at any time if they feel the plan is not right from not answering honestly or a major life event. Some of the information should be shown to the learner on the profile where course history is kept. Students should be able to view the specific outcomes from the MAT scales including motivator preferences

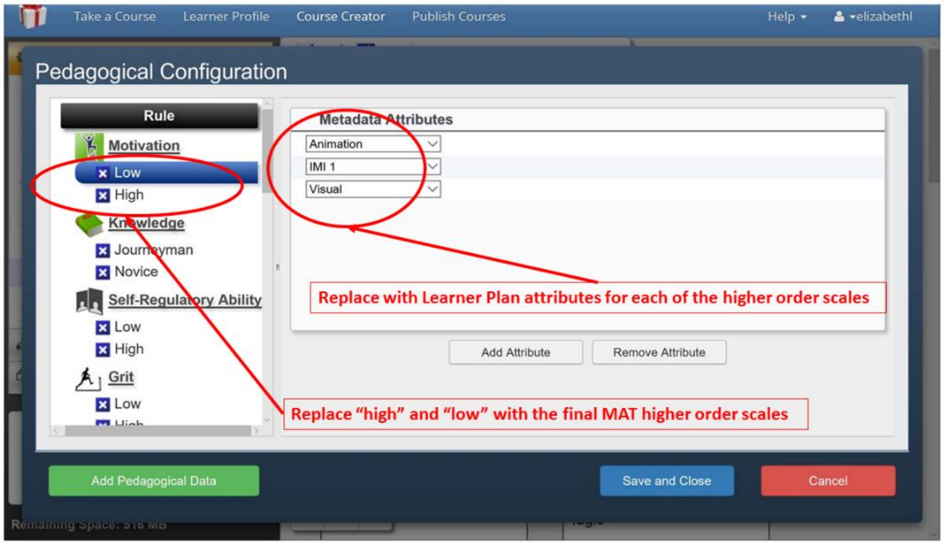
Learner Plan Overview

At the end of Phase I of this project, a set of 4 Learner Quadrants (Intrinsic, High Neuroticism, High Neuroticism with Low Conscientiousness and Low Openness, and Low Conscientiousness and/or Openness) was proposed (Reinerman-Jones et al, 2017), as a means of identifying learner strategies that could be authored in GIFT to support learning motivation based on an assessment of the learner's motivation and personality traits as assessed with the MAT. This present analysis supports Quadrant 1, which resembles the factor of Self-Directed Learning and high levels of Conscientiousness and Openness. These results from this study suggest that Quadrants 2 and 3 can be combined because an individual high in Neuroticism and the MAT factor of Threat Vulnerability is likely to be low in Conscientiousness and Openness. The results in general support Quadrant 4. Further analysis of the MAT scales relevant to the Social factor is warranted and may provide a way of decomposing into more specific learner plans. For example, competition and challenge are two scales of the MAT associated with the Social factor, so there could be a learner plan focused on including a challenge aspect to learner, such as providing a leaderboard with points or badges. A different learner plan may focus more on providing breaks to the learner. In addition, given the slight correlation to with Neuroticism, the type of social interaction may need to provide supportive interaction.

Pedagogical Module Implementation

In order to realize the Learner Plans in GIFT, the pedagogical module requires changes so that it can receive input from the LTLM at lesson run-time. The pedagogical module-authoring tool needs to be expanded to support options for the final MAT higher order scales and attributes for the associated learning plans. Figure 4 identifies the parts of the current pedagogical module authoring tool to be modified.

Attributes being considered for the learning plans are intended to promote and improve student motivation. For example, the Intrinsically Motivated (Self-Learner) student’s Learner Plans would include: options to write in the learning goal, complete a pre-test, and potentially demonstrate competency and earn credit for the lesson or portion of the lesson and the ability to select their preferred method of task completion (e.g., text, video, game). Whereas, the Learner Plan for a Threat Vulnerable student may include: sub-goals for dividing the lesson into smaller segments, incorporation of positive feedback throughout the lesson and incorporation of relaxation techniques throughout the learning process. Finally, a Social Learner Plan for the Extrinsically Motivated learner who is low in Neuroticism may incorporate: leaderboard for competition with other students, and feedback to help the student maintain focus.



Replace with Learner Plan attributes for each of the higher order scales

Replace “high” and “low” with the final MAT higher order scales

Figure 4. Recommended Modifications to Pedagogical Module Configuration Tool

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The results of the study described in this paper provide support that personality is a contributing factor to how a student interacts and performs within a given learning environment. Further, the study provided support that the use of a learner’s personality composition can be useful in developing a learning plan to support their motivation. The next step is to further analyze the MAT wave 3 results to better understand the variables that underlie the Social factor.

Benefits of GIFT Implementation

A verification experiment is planned to assess the contribution of Leaner Plans tailored to the MAT and personality assessments on mastery level (performance score) and retention of learning. It is anticipated that participants who receive instruction with the Learner Plan associated with their motivation and personality traits will improve their performance and retention, due to an increased level of motivation. With the MAT implemented within GIFT, it will help enable the provision of instruction to the learner in a manner that optimizes their learning outcomes.

Barriers to GIFT Implementation

Implementation of the MAT into GIFT will require some changes to GIFT authoring (configuration) tools and run-time engine. First, the results of the MAT Actionable Survey will need to be stored in a LTLM. Secondly, the pedagogical module configuration tool will need to be modified to support the MAT final scales and Learner Plan attributes. In order to implement some of the attributes being recommended, such as the ability to write in goals or select method of task completion (e.g., game or videos), extensions will need to be made to GIFT to support learning environments beyond those currently supported by GIFT. Finally, the pedagogical module will need to be able to receive data from the LTLM after the student logs into GIFT, rather than pulling the results in run-time from an actionable survey.

Summary

The authors have designed a study to evaluate tailored learning plans that are providing support for Quadrant 1, 2, and 4 in the above model. The results of this study will be used to better inform modifications to the GIFT authoring environment.

REFERENCE

Blickensderfer, E., Johnston, J., Paris, C., & Wilson, J. (2003). E-Learning: Implications of training theory & research. In, Proceedings to the 2003 Interservice/Industry, Training, Simulation, and Education Conference.

Costa, P. T., & McCrae, R. R. (1992). NEO-PI-R professional manual. Odessa, FL: Psychological Assessment Resources.

del Soldato, T., & du Boulay, B. (1995). Implementations of motivational tactics in tutoring systems. Journal of Artificial Intelligence in Education, 6(4), 337-378.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. Journal of personality and social psychology, 93(5), 880.

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.

Elliot, A. J., Murayama, A. J., & Pekrun, R. (2011). A 3x2 achievement goal model. Journal of Educational Psychology, 103(3), 632-648.

French, E. (1955). Some characterizes of achievement motivation. Journal of experimental Psychology, 50, 232-236.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. Review of personality and social psychology, 141-165.

Judge, T. A., Jackson, C. L., Shaw, J. C., Scott, B. A., & Rich, B. (2007). Self-efficacy and work-related performance the integral role of individual differences. Journal of Applied Psychlogy, 92, 107-127.

Keller, J. (1987). Strategies for simulating the motivation to learn. Performance and Instruction, 1-7.

Kember, D., Wong, A., & Leung, D. (1999). Reconsidering the dimensions of approaches to learning. British Journal of Educational Psychology, 60, 323 - 343.

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits n predicting college students' academic motivation and achievement. Learning and Individual Differences, 19(1), 47-52.

Komarrajuq, M., Karau, S. J., Schmeck, R. R., & Avdic, A. (2011). The Big Five personality traits, learning styles, and academic achievement. Personality and Individual Differences, 472-477.

Lameier, E., Reinerman-Jones, L., Matthews, G., Biddle, E. & Boyce, M. (pending). Motivational Assessment Tool (MAT): Enabling Personalized Learning to Enhance Motivation.

Sottilare, R., Goldberg, B., Brawner, K., & Holden, H. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.

Larsen, R. J., & Kasimatis, M. (1990). Individual differences in entrainment of mood to the weekly calendar. Journal of Personality and Social Psychology (58) , 164-171.

Magill, R. A. (1980). Motor learning: Concepts and application. In Hays, R. (Eds.), The Science of Learning: A Systems Theory Approach. Dubuque, IA: Wm. C. Brown Company Publishers.

Matthews, G., & Zeidner. (2004). Traits, states, and the trilogy of mind: An adaptive perspective on intellectual functioning. In, (David, Y.D. & Sternberg, R.J, Eds.). Motivation, Emotion & Cognition: Integrative Perspectives on Intellectual Functioning and Development, 143-174.

Noels, K., Clement, R., & Pelletier, L. (1999). Perception of teachers' communicative style and students' intrinsic and extrinsic motivation. The Modern Language Journal, 83, 23 - 34.

Pavlov, L. (1927). Conditioned Reflexes. Oxford: Oxford University Press.

Reinerman-Jones, L., Lameier, E., Biddle, E. & Boyce, M. (2017). Informing the Long-Term Learner Model: Motivating the Adult Learner (Phase 1). (Technical Report).Sottilare, R., Brawner, K., Goldberg, B. & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.

Ryan, R. M., & Deci , E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology, 54-67.

Schwartz, S. H., & Butenko, T. (2014). Values and behavior: Validating the refined value theory in Russia. European Journal of Social Psychology, 44, 799-813.

ACKNOWLEDGMENTS

This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911 NF-16-2-0062. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

ABOUT THE AUTHORS

Dr. Elizabeth Biddle *is a Boeing Technical Fellow and currently provides technical leadership in the Boeing Research & Technology (BR&T) Advanced Learning organization to support the development of technologies.*

Elizabeth Lameier *is a Research Associate in the University of Central Florida’s Institute for Simulation and Training Prodigy Lab. Previously, she was an Intervention Specialist and Early Childhood Educator at public school.*

Dr. Lauren Reinerman-Jones *is the Director of Prodigy Lab at the University of Central Florida’s Institute for Simulation and Training, focused on explaining, predicting, and improving human performance and systems.*

Dr. Gerald Matthews *is a Research Professor at the University of Central Florida’s Institute for Simulation and Training. His research interests include psychometrics, human performance, and workload assessment.*

***Dr. Michael Boyce** is a research psychologist with ARL's adaptive training research program. For the past 3 years his emphasis has been in using technologies like GIFT to accurately assess learner knowledge and performance.*

Theme IV: Team Modeling

Team Models in the Generalized Intelligent Framework for Tutoring: 2018 Update

Anne M. Sinatra[1]
U.S. Army Research Laboratory[1]

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT) is a domain-independent intelligent tutoring system (ITS) framework that has many features and applications that can be used by ITS authors (Sottilare, Brawner, Sinatra & Johnston, 2017). The research and development associated with GIFT is divided into a number of different research vectors including, architecture, individual learner modeling, team modeling, instructional management, domain modeling, and training effectiveness. Many of the research projects that have been associated with GIFT have focused on developing new features and tools that ITS authors can use to create their courses. An ultimate goal of GIFT is being able to provide tutoring to teams. Examples of teams that GIFT plans to support are at the Squad level (9 people) and higher. An initial summary of the work in the team modeling vector through May 2017 was provided in the GIFT Symposium 5 proceedings (Sinatra, 2017). The current paper provides an update in the progress and work that has been done in team models in GIFT.

TEAM MODELING IN GIFT

Theoretical Background

Work in the area of team modeling in GIFT has been separated into two different divisions: theoretical and applied. The theoretical basis for the work in team modeling in GIFT was done as part of a large scale meta-analysis that covered the relevant team literature from 2003 to 2013. The results of this meta-analysis were recently published in the Journal of Artificial Intelligence in Education (Sottilare, Burke, Salas, Sinatra, Johnston & Gilbert, 2017). As part of this project, behavioral markers were also identified. These behavioral markers provided ways of assessing team performance during a session. Many of the markers are heavily communication focused, and are traditionally assessed in person by a human observer. Future work is planned in which these behavioral markers will be specifically selected and operationalized in the context of real-time team intelligent tutoring.

Applications of Team Modeling and Team Tutoring in GIFT

Surveillance Tasks

Team tutoring has been successfully demonstrated within the GIFT architecture in the form of a Surveillance Tutor. This tutor was developed using GIFT 4.0, and involved tracking both individual performance and team performance in real-time. This was initially implemented in the form of an experiment, which had two players who were working together to surveil an area in the Virtual Battlespace 2 (VBS2) software. The initial version had two team members who each surveilled a 180 degree area and communicated to their teammate when they saw a threat (OPFOR) crossing into the other person’s sector. The lessons learned and information about the creation of this tutor was documented in a recently published article (Gilbert, et al., 2017). Depending on the condition that the team was in, feedback was either provided at the individual level, the team level, or not at all. The GIFT system was adapted to support individual assessment files (Domain Knowledge Files; DKFs) and a team assessment file. Each of the team members had his or her own DKF which tracked their individual actions and performances, and there was an overall team DKF which assessed team performance. The team DKF assessed the team tasks as a whole, and was unable to discern which team member engaged in which action.

As a follow up to the initial implementation, the task was extended such that there were three individuals who worked together as a team to perform the surveillance task. While in the first implementation of the task there were two individuals who each performed the same role (spotters), in the expanded version there were the two spotters and an additional role, a sniper. The spotters surveilled their 180 degrees and instead of telling their partner when they saw a threat passing to the other sector, they communicated this information to the sniper. The sniper then was tasked with acknowledging that the information was received, locating the possible threat and identifying the threat level associated with the spotted individual. The sniper would determine if it was a civilian, a potential threat, or an imminent threat. This implementation used a similar DKF structure in GIFT, where each of the spotters had their own task specific DKF, and another DKF was generated for the sniper and provided assessment that was associated with that role. Finally, there was an overall team DKF that examined the team actions and could provide feedback based on them. In this version of the experiment, feedback was provided in one of two ways: at the individual level or at the team level. Lessons learned from this approach included that this particular approach to assessing team tutoring in GIFT would result in an increasing number of DKFs as the number of team members and roles were increased. Further, the overlap and reassessment in the team DKF required additional authoring and duplication of efforts. While not the ideal scalable approach to team tutoring in GIFT, this implementation was an important step forward, as it demonstrated the simultaneous assessment/tutoring of three individuals, and the ability to assess individuals in different roles.

Search and Rescue Task

The next implementation of team tutoring in GIFT will be in the Search and Rescue domain using the Virtual Battlespace 3 (VBS3) software. This work is still in initial development and the implementation is in progress (McCormack et al., in press). As part of this project, there will be effort made to operationalize previously identified behavioral markers (e.g., cohesion and cooperation), and create a task that will elicit appropriate team behaviors that are associated with them. The search and rescue task is being developed in such a way that it is military relevant, and subject matter experts are being consulted in order to ensure that the tasks within the scenario are as realistic as possible. The focus for this implementation is on the team performance overall, with less focus on the individual. The actions that the team members perform will ultimately be assessed in one team DKF, and feedback will be provided at the team level. This reduces some of the complications of using multiple DKFs as was done in previous work using GIFT. Additionally, as this scenario is expected to be made up of 9 people, and include subteams within the structure, providing assessment/feedback in this manner will reduce complications and allow for focus to be on developing a rich relevant scenario. While this project is ongoing, the GIFT team will be developing a scalable solution to the GIFT team architecture, which can both be merged with this scenario and used for future implementations in GIFT.

TEAM MODELING WORKSHOPS AND OUTPUT

Team Workshop, March 2016

As part of the meta-analysis project, a workshop titled "Building Intelligent Tutoring Systems for Teams: What Matters" was conducted in March 2016. This workshop focused on teamwork as it applied to ITSs. Brainstorming and discussion happened during the workshop about best practices in ITSs and what is relevant in order to conduct team tutoring. The discussion lead to the output of an edited book volume. Individuals who attended the workshop, in addition to others who were experts in the field were invited to contribute chapters. The book was recently completed and is in the editing process. The final book titled *Building Intelligent Tutoring Systems for Teams: What Matters, Volume 19*, with editors Joan Johnston, Robert Sottilare, Anne M. Sinatra, and C. Shawn Burke is scheduled to be released in September 2018.

Team Taskwork Expert Workshop, June 2017 and Design Recommendations Book Volume

In June 2017 a Team Taskwork Expert Workshop was held at Iowa State University in Ames, Iowa. This workshop was held as part of the ARL-University of Memphis cooperative agreement, and one of the goals was bringing together a group of experts in different areas of team research (including collaborative learning, team performance, and team tutoring) to discuss their work and how it is applicable to team taskwork in ITSs. The focus of the workshop was specifically on taskwork, or ways that intelligent tutors could be applied for specific tasks or domain areas. There were a wide range of presentations that included discussions about applications in the medical field, in military domains, in analyzing the content of team messages, and more. In addition to the discussions, the formal output of the expert workshop is in the form of an edited volume. The book, tentatively titled: *Design Recommendations for Intelligent Tutoring Systems: Team Taskwork*, includes four focuses areas about team taskwork: modeling, socio-cultural applications, system design and assessment. The editors of the book are Robert Sottilare, Art Graesser, Xiangen Hu, and Anne M. Sinatra. The book is expected to be released in summer 2018.

Assessment and Intervention during Team Tutoring Workshop, Artificial Intelligence and Education Conference, June 2018

The GIFT team has an accepted workshop at the Artificial Intelligence and Education (AIED) Conference in June 2018 in London. Papers have been accepted to the workshop that highlight different areas and applications of team tutoring in ITSs. Areas of focus include collaborative problem solving, demonstration of team tutoring in action, and communication during team tutoring. During the workshop there will be a discussion of the commonalities in the different approaches to team tutoring and a discussion of gaps and steps forward overall in the problem area. The output of this workshop will be proceedings papers that will be available online and the information gathered from the workshop will impact the way forward for team tutoring in GIFT.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The lessons learned from the surveillance tutor and the recommendations for both intelligent team tutoring systems and GIFT in particular that have come out of the team workshops will be taken into consideration while developing the team architecture in GIFT. Providing real-time feedback and assessment for not only individual team members, but a team as a whole is a difficult challenge. As was demonstrated in the initial surveillance tutor, it can be difficult to interrupt individuals in the middle of a scenario to provide real-time feedback, and it may result in them either reducing performance, or not being able to attend to the feedback. As a result of this, initial implementations in the future may use the approach of focusing on after action reviews that occur at the team level after the completion of subtasks within a scenario. By engaging in this manner it will not interrupt the event that is occurring and will ensure that the feedback is viewed by the team members. Additionally, as authoring multiple DKFs would be cumbersome, work should continue to be done in order to implement a scalable team architecture that lessons the authoring burden but still provides relevant assessment and feedback during tutoring sessions.

Research into team modeling in GIFT should continue to be actively developed, and careful thought should be given into the implementation of the team architecture in GIFT. Additionally, work should continue to be done to use the theoretical foundation that was identified in order to implement successful team tutoring in GIFT. By operationalizing the behavioral markers and determining which can be generalized it will provide a powerful theory driven approach to team tutoring that tutor authors will be able to implement. Through the development of scenarios and the architecture, the initial plans for a team tutor authoring tool can be put into place. Ultimately, as GIFT is adapted for use with teams it will lead to it becoming more powerful and incorporating many additional relevant features that can be authored. Similar to the other authoring tools in GIFT, it is expected that the team tutor author will not need to be heavily versed in the team literature, but can use the tools, prompts, and recommended pedagogy within GIFT to construct a highly relevant team tutor that is pedagogically sound. GIFT continues to be developed in order to support team tutoring, with future work including demonstration of a Squad level team tutor, and approaches to assessing teams in real-time.

REFERENCES

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., ... & Winer, E. (2017). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Education*, 1-28.

McCormack, R., Kilcullen, T., Sinatra, A.M., Brown, T., & Beaubian, J. (in press). Scenarios for Training Teamwork Skills in Virtual Environments with GIFT. In Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6).

Sinatra, A. M. (2017, July). Team Tutoring in the Generalized Intelligent Framework for Tutoring: Current and Future Directions. In Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5), p.123.

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, 1-40.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring.org*.

ACKNOWLEDGEMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. Additionally, the author would like to acknowledge and thank the entire U.S. ARL Adaptive Training team, who have all been instrumental in the design and development of GIFT and the GIFT authoring tools.

ABOUT THE AUTHORS

***Dr. Anne Sinatra** is part of the adaptive training research team within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab, she is lead on the Team Modeling vector and works on the Generalized Intelligent Framework for Tutoring (GIFT) project. Her background is in Human Factors and Cognitive Psychology.*

Team Performance and Assessment in GIFT – Research recommendations based on Lessons Learned from the Squad Overmatch Research Program

Joan H. Johnston, Ph.D.
Army Research Laboratory

INTRODUCTION

In 2015 a set of research objectives were developed for the Army Research Laboratory's (ARL) adaptive tutoring program focusing on designing and developing effective team tutoring environments in GIFT (Goodwin, Johnston, Sottilare, Brawner, Sinatra, & Graesser (2015). An initial objectives is determining the important variables that drive small unit team performance and developing ways to measure and model those factors in an adaptive training system. At the time the objectives were formulated the ARL research team had already begun a joint, collaborative research program called Squad Overmatch (SOvM) that conducted a series of team-based research studies that in part focused on addressing this question (Milham, Phillips, Ross, Townsend, Riddle, Smith, et al., 2017). The purpose of this paper is to describe how the SOvM program approached the problem of team performance measurement and describes lessons learned for measuring and modeling those factors in an adaptive training system.

SQUAD OVERMATCH

The SOvM research objective is to improve dismounted squad decision making under stress, with a focus on the following five skill domains: Tactical Combat Casualty Care (TC3); Advanced Situation Awareness (ASA); Resilience and Performance Enhancement (RPE); Team Development (TD); and conducting an Integrated After Action Review (IAAR). In 2016 an experiment with eight squads was conducted to determine the effect of training these skills using classroom, simulation, and live training compared to traditional live training exercises (Townsend, Johnston, Ross, Milham, Riddle, Phillips, & Woodhouse, 2017). The four-day SOvM curriculum involved Subject Matter Experts (SMEs) conducting classroom instruction on days 1 and 2 that was immediately followed by skills development in a virtual team training simulation, and then conducting practical skills application in an outdoor training facility on days 3 and 4. Following each 45 minute scenario, the Platoon leader and learning domain SMEs led the squads in guided team self-correction IAARs. The IAAR was aligned with the U.S. Army AAR doctrine for discussing the movement and engagement actions the squad performed during significant tactical events during the scenario. The IAAR focused on developing squad member skills in how to take personal responsibility for identifying behaviors that need correction, develop team cohesion, and set goals for improvement in the next scenario. For the first 20 minutes, the Platoon leader led the squad members in a critique of their tactical performance using video snippets of the critical events collected during the exercise. Then each domain SME spent about 5 to 7 minutes leading squad members in identifying tactical triggers, behaviors, solutions, and outcomes as they reflected on each of the areas, sometimes reviewing video snippets. Finally, the Platoon leader led the squad members in setting goals for improvement in the next scenario. In this manner, the teaching points were reinforced based on practical application, and provided a way to "adapt" how they used the next scenario to focus on performance objectives they had set themselves.

Team Performance Measurement Approach

A major goal of the study was to test the hypothesis that squads receiving the SOvM training would demonstrate better performance on TC3, ASA, and TD over the control condition squads during and after the live training exercises. To construct measures of these skill domains, researchers leveraged two types of team competency models and measurement methods that had been previously validated in earlier research. The Team Tactical Decision Making competency model and the Teamwork competency model were used to develop performance objectives and measures for ASA and TD.

Team Tactical Decision Making Competency Model

The Team Tactical Decision Making model was developed by Paris, Johnston, and Reeves (2000) and is comprised of the four related dimensions of Identification, Elaboration, Planning, and Execution. Johnston, Fiore, Paris, and Smith (2013) validated the model by mapping Navy combat team behaviors to the four categories based on their performance objectives (i.e., the detect-to-engage sequence) for managing their air warfare tasks and assessing performance with the measure (Air Warfare Team Performance Index or ATPI) in simulation-based training exercises. Spiker, Johnston, Williams, and Lethin (2010) then used the identification and elaboration categories of the TDM model to characterize dismounted rifle infantry squad member behaviors during training exercises designed to improve their collective decision making skills. The SOvM study used these identification and elaboration behaviors to guide development of ASA performance objectives in the simulation and live scenarios.

Identification processes involve strategies for employing and manipulating one’s own cognitive resources and available assets to orient, observe, recognize, and identify potentially important hostile, friendly, and neutral players based on a particular configuration of features. Such configurations tap an individual’s knowledge of cues in the environment, thereby enabling identification of hostile intent, projecting future actions of the players, and ultimately assigning threat potential (e.g., friendly, hostile, neutral, unknown) to them. Identification is an inherently team task as it requires the exchange of timely and accurate reporting of the ongoing state of those features to team members within the team and up the chain of command to feed the common operational picture. Table 1 lists the identification skill definitions and performance objectives developed by Spiker et al. (2010).

Table 1. Identification Skills and Example Performance Objectives. (Adapted from Spiker et al., 2010)

#	Identification Skills	Example Performance Objective
1	Establish a geometry of fires to create an interlocking network of optics, intelligence, and communications	Team members triangulate their communication, optics, and intelligence data to ensure comprehensive coverage of an event, individual, vehicle, anchor point, or habitual area.
2	Utilize organic assets and natural light to make positive identification	Team members use optics (e.g., binoculars and thermals) as effective substitutes in determining, for example, what part of a body was shot and how bad the wound is based on the color of the blood on the ground.
3	Make innovative use of optics (and other organic assets) to help construct a baseline or	Team members use range estimation capability in optics to determine opposing forces social status indicators (e.g., to

	profile	determine if a person of interest is a leader).
4	Shift field of view – from wide to narrow and back – and thereby avoiding focus lock	Team members watch a distant target for awhile with binoculars and then switch to naked eye in order to better interpret the context surrounding the specific action they are watching.
5	Efficiently refocus observation scan to include both near and far objects in the scene	Team members keep all parts of their viewing sector, both near and far, within their visual field scan and in their focal attention so that no important cues are missed.
6	Orient observation or tracking toward potentially hostile players or good guys and ignore neutrals	Team members economize their profiling by concentrating observations on potential hostiles (insurgents, informants) and potential friendlies (police, security), while reducing attention to the neutrals (regular population).
7	Make effective and efficient identification of anchor points and indications of anti-tracking	Team members economize their observations by localizing their viewing on areas–anchor points–where hostiles tend to concentrate their illicit activities, such as specific parts of town or a building.
8	Make effective and efficient identification of habitual areas and action indicators	Team members economize their observations by localizing their viewing on areas–habitual areas–where townspeople congregate and which might represent a “soft target” for hostile activity, such as a market or mosque.
9	Make effective and efficient identification of opposing force leaders	Team members determine who the leader is in a village by using the four key indicators (entourage, direction, mimicry, adoration) of leadership.
10	Adopt appropriate criteria based on objective cues observed to make timely, accurate decisions	Team members use clue clusters to collect three pieces of evidence, such as three indicators of a leader or a terrorist planning cycle, before taking action.
11	Induce or generalizes a pattern from a few individual cues	Team members infer the presence of a larger event–such as a Vehicle Born Improvised Explosive Device (VBIED) or a complex ambush–by generalizing from the presence of a few cues (e.g., how a car is parked, or how a sniper team has been deployed).
12	Look for prototypes (vs. template matching)	Team members look for signature behaviors (e.g., insurgent, HVT, vehicle, or a track) and signature locations (e.g., habitual area, anchor point, or aerial spoor) through a cluster of cues.
13	Establish an observation baseline to extract normalcy	Team members make a systematic, sustained observation on a person, event, location, or vehicle to determine what behavioral profile constitutes “normal,” where this normal is used as the baseline against which deviations are noted. A baseline, for example, might be established for market behavior when insurgents

		are not present.
14	Look for anomalies – above and below baseline (including the absence of something)	Team members look at the elements to note anything out of place or anomalous, either something there that should not be or something missing. As an example, team members should observe a group of people to see if someone seems out of place based on biometrics (e.g., they are sweating from running) or if a vehicle is parked in an unusual location (possible VBIED).

Elaboration involves tapping into one's background store of information that summarizes what has been learned previously about similar situations; it enables the team members to create a shared mental model of the situation. Effective elaboration involves applying and discussing with other team members previous knowledge (e.g., of hostile profiles) about the current situation, such that the most reliable and acceptable hypothesis may be found with regard to the intent of a potentially hostile actor. Team members map their current experiences onto a cognitive template they had developed from previous experiences, and then attempt to match each part of this template with some aspect of the current situation. Table 2 is a list of the elaboration skills that Spiker et al. (2010) produced from their study.

Table 2. Elaboration PCR Skills and Example Performance Objectives. (Adapted from Spiker et al., 2010)

#	Elaboration Skills	Example Performance Objective
15	Take evidence-based approach to identifying hostiles using hard data to confirm or disconfirm a hypothesis	Team members take the time to list three reasons why an individual is a body bomber or an HVT, rather than going with a hunch to save time.
16	Generate explanatory storylines that tie individual items of information together	Team members construct alternative explanations for how individual events or pieces of evidence might be related and part of a larger whole.
17	Imagine alternative courses of action or alternative event outcomes by what-if mental simulations	Team members attempt to "think through" what might be happening in an unfolding event (e.g., a possible complex ambush) by rapidly reviewing different, but plausible, alternative outcomes.
18	Detect an unfolding event or activity by identifying a piece of it and inferring the rest	Team members view a sequence of events as being tied together by some underlying process-unfolding like a movie- such as the steps to create and plant a bomb or the cycle of planning a terrorist attack.

Teamwork Competency Model

The Teamwork competency model is comprised of the four dimensions of information exchange, communication delivery, supporting behavior, and initiative/leadership. Information exchange involves team members passing relevant information to the right team member at the right time, seeking information from all relevant sources, and providing periodic situation updates that summarize the big picture. Communication delivery involves using proper terminology, avoiding excess chatter, speaking clearly and audibly, and delivering complete standard reports containing data in the appropriate order. Supporting behaviors consists of offering, requesting, and accepting backup when needed, and noting and correcting errors, as well as accepting correction. Initiative and leadership consists of explicitly stating priorities and providing guidance, suggestions, or direction to other team members. Smith-Jentsch, Johnston, & Payne (1998) developed and validated the teamwork competency model in a series of studies with Navy combat teams. Then Smith-Jentsch, Cannon-Bowers, Tannenbaum, and Salas (2008) demonstrated in an empirical, field experiment that Navy combat teams that participated in facilitator-led guided team self-correction structured around the expert model of teamwork developed more accurate mental models of teamwork, demonstrated more teamwork processes, and achieved more effective performance outcomes after two training cycles than did those briefed and debriefed using the traditional Navy AAR method. The SOvM program adapted the Teamwork competency model and guided team self-correction method to establish the Team Development and IAAR performance objectives.

Translating Competency Models into Event-Based Training Scenarios

The event-based approach to training method was applied to the SOvM training scenario design to ensure the skills identified in the TDM and Teamwork competency models would be learned (Rosen, Salas, Tannenbaum, Pronovost, & King, 2011). Critical tasks, task stressors, learning objectives, exercise design and execution, performance measurement, and feedback were clearly linked and documented prior to completing the scenarios. An important feature in designing scenarios was including as much of the knowledge-rich environment in the virtual and live scenario events as possible so that pre-specified cue-strategy relationships could be observable and would result in producing team responses that were observable and measureable.

Five event-based scenarios of approximately 45 minutes in length were developed with a single overarching narrative that had the scenarios taking place over a fictional four week period of time. Two scenarios were designed for the team training simulation (B1 and B2) and three scenarios (M1, M2, and M3) were developed for the live training environment. Following the graduated exposure to stress guidelines each scenario was designed to provide an increasing number of task stressors (Driskell, Salas, & Johnston, 2006). Key events and associated ASA, TD, and TC3 performance objectives were developed for each scenario. For example, Scenario M2 had the squad mission objective of conducting a zone reconnaissance in order to conduct a key leader engagement; exploit intelligence; confirm location of a suspected arms cache; and, exploit the site, if able.

The chronological list of nine key events for M2 were:

1. Establish listening post (LP)/observation post (OP).

2. Depart LP/OP.

3. Observe civilian interactions in village.

4. Conduct key leader engagement and tactical questioning with high value target.

5. Observe proxemics push as village civilians move away from the central square.

6. Squad moves north to tea shop to interview civilian woman.
7. Sniper fire results in soldier receiving gunshot wound (GSW) to arm and a civilian woman receiving GSW to chest.
8. Squad conducts movement toward sniper locations.
9. Soldier receives GSW to chest at sniper location.

Performance objectives were then mapped to scenario events. Table 3 presents 51 objectives developed for M2 and shows many objectives are repeated across events. The ASA behaviors represent the identification and elaboration behaviors described in Table 1 and the TD behaviors representing the four dimensions of teamwork. Multiple performance domains are represented in specific events to ensure the scenarios had sufficient levels of stressors. For example, the last row in Table 3 shows events 1, 7, and 9 had many more performance objectives (15, 22, and 19, respectively) compared to the other six events.

Many of the ASA identification and elaboration performance objectives were planned in Events 1 through 6. For example, in Table 3, the ASA performance objective #1 expected squad members to "use tools or otherwise visually identify objects that are hidden in windows or shadows through the town." From Table 1, this behavior represents Identification skill #2 – "utilize organic assets and natural light to make positive identification." The TD performance objectives were inserted throughout all the events. For example, the TD performance objective #17 expected squad members to "pass information among teams about their observations of the town." This behavior is representative of the TD behavioral dimension of "information exchange." During Event 7 a sniper fire results in a Soldier receiving a GSW to his arm and a civilian woman receiving a GSW to her chest. This was expected to elicit multiple TD behaviors, such as objective #26 – "provide complete and accurate medical reports" (Communication Delivery), and objective #31 – "squad leader and team leaders provide guidance and state priorities regarding roles for continuing mission" (Initiative/Leadership). Event 7 also involved the TC3 behaviors, such as objective #38 - waits for suppressive fire or other cover before retrieving casualty (Care Under Fire), and Objective #49 - provides medical updates to Squad Leader; completes MIST report, and 9-Line (Casualty Evacuation Activities).

Table 3. Event-Based Performance Objectives for ASA, TD, and TC3 in Scenario M2.

	PERFORMANCE OBJECTIVES	M2 EVENTS								
		1	2	3	4	5	6	7	8	9
	Advanced Situation Awareness									
1	Squad divides into two separate forces for two LP/OPs to establish geometry of observation	X								

	PERFORMANCE OBJECTIVES	M2 EVENTS								
		1	2	3	4	5	6	7	8	9
2	Use tools OR OTHERWISE visually identifies objects that are hidden in windows or shadows through the town	X								
3	Establishes geographic points of interest (avoidance or common use of an area)	X								
4	Establishes atmospheric details (information that is or is not in line with baseline from intelligence)	X								
5	Establishes that groups of civilians are engaging in mimicry, adoration, directing attention, or are part of an entourage	X								
6	Positively identifies Key Leader	X								
7	Establishes key leader identification to include how key leader was identified and why it is believed it is the key leader	X								
8	Establishes baseline behaviors of target	X								
9	Employs guardian angel / geometries of observation		X		X		X			
10	Verbalizes nature of target nonverbal behaviors		X		X					
11	Communicates an assessment to include why s/he believes the validity, quantity of the information received		X							
12	Communicates deviations in baseline of behavior of target		X							
13	Offers some medical care to local national (good shepherd)			X						
14	Identifies that townspeople exhibit slight proxemics push away from the squad					X				
15	Identifies nonverbal and paralanguage cues that					X				

	PERFORMANCE OBJECTIVES	M2 EVENTS								
		1	2	3	4	5	6	7	8	9
	townspeople are uneasy about squad's presence									
	Team Development									
16	Squad leader gives direction to separate into two LP/OPs	X								
17	Squad members pass information among teams about their observations of the town	X								
18	Use available resources to determine identifying characteristics (e.g., OPORD)	X								
19	Communicate to team members when a townsperson fits description of key leader	X								
20	Communicate to team members when groups of people are engaging in mimicry, adoration, directing attention, or are part of an entourage	X								
21	Communicate to chain of command when key leader is identified	X								
22	Correct errors in information repeated on radio	X								
23	Backup is provided to the squad member engaging in the interview		X		X					
24	Communicates a situation update up the chain of command		X		X					
25	Communicates changes in priority from chain of command to other team members		X		X					
26	Provides complete and accurate medical reports							XXX		XXX
27	Support Squad Leader & establish medical SA exchanges casualty information with Squad Leader							X		X

	PERFORMANCE OBJECTIVES	M2 EVENTS								
		1	2	3	4	5	6	7	8	9
	and Village Leader / casualty.									
28	Squad asks higher for guidance in further care of civilian casualty							X		
29	Directs TMs to provide care							X		X
30	Squad leader and team leaders exchange information about status of the squad								X	
31	Squad leader and team leaders provide guidance and state priorities regarding roles for continuing mission								X	
32	Squad members call out enemy position and status to squad, giving a complete report									X
	Tactical Combat Casualty Care									
33	Delivers some medical care to local national (good shepherd)			X						
34	Returns fire/provide security; weapons up; scans for enemy; fires weapon							X		X
35	Provides MAN Down Report to Squad Leader							X		X
36	Provides casualty status info to medic							X		X
37	Establish security / provide cover after injury occurs, TMs face outward from casualty (360); guns up, looking for enemy. TMs lay suppressive fire to provide cover							X		X
38	Waits for suppressive fire or other cover before retrieving casualty							X		X
39	Retrieves casualty							X		X
40	Treats casualty							X		X
41	Squad Leader directs TLs to suppress enemy to maintain tactical focus							X		X
42	Squad Leader collects medical and tactical info							X		X

	PERFORMANCE OBJECTIVES	M2 EVENTS								
		1	2	3	4	5	6	7	8	9
43	Squad asks higher for guidance in further care of civilian casualty							X		
44	Squad leader determines capability to continue mission							X		X
45	Assigns medical & tactical resources to establish CCP							X		X
46	Send up first 5 lines of 9-line report; Complete, accurate, brief, and clear reporting							X		X
47	Medic provides advanced care							X		
48	Directs TMs to provide care							X		
49	Provides medical updates to Squad Leader; completes MIST report, and 9-Line							X		
50	Squad leader decides that squad remains combat effective and decides to move forward with the mission								X	
51	Consolidates CCP									X
	Total Objectives Per Event	15	7	1	5	2	1	22	3	19

Measures Development and Application

The performance objectives for ASA, TD, and TC3 in each scenario were transformed into individual behavioral observation checklists in a spreadsheet format and on an android tablet so that SME raters could assess the squads during the scenarios. Observations of behaviors in virtual scenarios B1 and B2 were attempted, but proved difficult as it was challenging to hear and see squad member behaviors within the virtual world (Townsend et al., 2017). In addition, squad members were sitting next to each other using VBS3 and they often communicated face-to-face instead of using their radios, which added to the challenge to effectively observe. It was also difficult to observe multiple team members in the virtual environment from one control station. These challenges made it difficult to determine whether behaviors occurred or not, or were simply missed.

During the live scenarios, assessors observed squad members moving through the urban village buildings and outdoor spaces on multiple video screens in the control room, and listened to squad communications via an audio system that was specifically developed for the experiment to enable isolation of communications among any needed subset of squad members in real time. The ASA and TC3 instructors followed and observed squads in the outdoor training site. The ASA and TC3 raters used spreadsheet based checklists. Following the exercises, they met with the respective SME instructors to establish ground truth for squad performance on ASA and TC3 behaviors. This approach enabled the ASA and TC3 raters to obtain almost 100% certainty about squad performance.

The two TD observers used the android tablet – based Mobile Performance Assessment Tool to make their event-based ratings during each live scenario run. Townsend et al. (2017) found the average percent agreement score for scenarios M2 and M3 was 80%. The M2 scenario agreement score was higher (89%) than the agreement score for scenario M3 (70%), and the raters suggested that because the M2 scenario had fewer complex events it may have been easier to see squads and hear their communications, whereas, scenario M3 was more complex and the raters may have had more trouble seeing or hearing the squad members. In addition, raters determined that more practice was needed to make the right assessments of squad behaviors. All raters also used the recorded videos and squad member communications following the experiment to correct missing ratings and for the TD raters to develop 100% consensus on the performance assessments.

Team scores for ASA, TD, and TC3 performance were calculated as the percent of tasks accomplished in a scenario. It was calculated by summing the number of behaviors performed by the squad on each of these skill domains and dividing it by the total possible number of behaviors that were expected to be performed in that skill domain.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Implications for the GIFT architecture

The measurement strategy defined in this study has implications for the Sottilare, Brawner, Sinatra, and Johnston (2017) GIFT functional concept for a “learning effect model for teams” that we briefly discuss here. The GIFT learning effect model presents an iterative data collection and learning methods delivery approach that presents specific functional features for team assessment (as noted in italics below). In this conceptual architecture, *team members* produce behavioral data during a training exercise that are detected and tagged by pre-defined *behavioral markers.* The *behavioral markers* populate the *initialization data for teams (e.g., competencies)* that in turn populates the *long-term team model*, and also the *team data* function. The *team data* function informs the *team states*. The TDM and Teamwork models could provide the competency framework for *the initialization data for teams* function and the *long –term team model*. The TDM and Teamwork competency behaviors could serve as the *behavioral markers* that GIFT needs to seek from the behavioral data generated by the *team members* during the exercise. As the *behavioral markers* of TDM and Teamwork are collected GIFT would generate *team states* for each of the four TDM and four Teamwork dimensions. *Team states* for the TDM and Teamwork dimensions would then be able to inform the GIFT *team instructional strategy selection*. For example, if the team is doing well on information exchange, but they are not catching and correcting errors (supporting behaviors), then GIFT would provide feedback in the AAR that the team needed to improve on supporting behaviors such as error correction, and sustain their good information exchange.

Future Research

Below are several research recommendations to continue to address the initially stated objective in this paper to develop ways to measure and model team behaviors in an adaptive training system.

Recommendation 1

The SOvM study demonstrated that team competency models for TDM and Teamwork are generalizable for assessing dismounted squads conducting tactical and TC3 tasks and can be used to assess team performance progress during training. It is recommended that these competency models be used as a tool to diagnose team performance and that further analysis of the SOvM data needs to be conducted to categorize observed behaviors into the TDM dimensions for planning and execution to further validate the model and increase the diagnosticity of the measures.

Recommendation 2

The majority of TDM and TD behaviors assessed were obtained from a team's verbal and non-verbal communications that trained human raters could hear, see, and categorize. A fairly high level of rater agreement can be achieved on TD behaviors using a tablet-based device, but increased rater error likely occurs as scenario events become more complex. It is recommended that adaptive tutoring needs to develop natural language recognition and processing to automatically categorize verbal behaviors into the TDM and TD competency models.

Recommendation 3

It was easier to observe and evaluate squads in the live exercises because the audio and video technologies were available and configured to the raters' needs. Team assessment in the virtual training environment was impossible due to the noisy communications and inability to effectively observe the squad actions in the scenario on the small PC monitors. Research needs to focus on developing technologies that can diagnose squad performance information in a rapid and organized method in both simulations and live training exercises. Tools need to be developed for capturing event-based team simulation interactions representative of the TDM and TD models and organized for the event-based IAAR. For example, the virtual team simulation currently records squad actions in a scenario for human-controlled replay in the AAR, but it is labor intensive and complicated to manipulate, and does not support the event-based approach to conducting the IAAR. With simulation recordings and speech to text recordings a more accurate representation of TDM and TD could be obtained with few to no humans in the loop collecting this information. In the live environment, sensor worn technologies that record audio and visual information, and location would enable more accurate and efficient assessments.

REFERENCES

Driskell, J. E., Salas, E., & Johnston, J. H. (2006). Decision Making and Performance under Stress. In T. W. Britt, C. A. Castro, & A. B. Adler (Eds.), Military life: The psychology of serving in peace and combat: Military performance (Vol. 1) (pp. 128-154). Westport, CT: Praeger.

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 U.S. Army learning model: Research outline (No. ARL-SR-0336). Aberdeen Proving Ground, MD: Army Research Laboratory.

Johnston, J.H., Fiore, S.M., Paris, C. & Smith, C.A.P. (2013). Application of cognitive load theory to develop a measure of team cognitive efficiency. Military Psychology, 25(3), 252-265.

Milham, L. M., Phillips, H. L., Ross, W. A., Townsend, L. N., Riddle, D. L., Smith, K. M., ... & Johnston, J. H. (2017). Squad-level training for Tactical Combat Casualty Care: instructional approach and technology assessment. The Journal of Defense Modeling and Simulation, 14(4), 345-360.

Paris, C. R., Johnston, J. H., & Reeves, D. (2000). A schema-based approach to measuring team decision-making in a Navy combat information center. In C. McCann & R. Pigeau (Eds.), The human in command: Exploring the Modern Military Experience (pp. 263-278). NY: Kluwer Academic/Plenum Publishers.

Rosen, M. A., Salas, E., Tannenbaum, S. I., Pronovost, P. J., & King, H. B. (2011). Simulation-based training for teams in health care: Designing scenarios, measuring performance, and providing feedback. Human factors and ergonomics in health care and patient safety. CRC Press, London, 573-594.

Smith-Jentsch, K.A., Cannon-Bowers, J.A., Tannenbaum, S.I., & Salas, E. (2008). Guided team self-correction impacts on team mental models, processes, and effectiveness. Small Group Research, 39(3), 303–327.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. Making decisions under stress: Implications for individual and team training, 1, 61-87.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Retrieved from https://gifttutoring.org/documents/31

Spiker, V. A., Johnston, J. H., Williams, G., & Lethin, C. (2010, December). Training tactical behavior profiling skills for irregular warfare. In Proceedings of the 2010 Interservice/Industry Training, Simulation, and Education Conference.

Townsend, L., Johnston, J., Ross, B., Milham, L., Riddle, D., Phillips, H., & Woodhouse, B. (2017, July). Development of a Mobile Tool for Dismounted Squad Team Performance Observations. In International Conference on Virtual, Augmented and Mixed Reality (pp. 312-321). Springer.

ABOUT THE AUTHOR

Dr. Joan Johnston is a Senior Scientist with the U.S. Army Research Laboratory where she leads research on training effectiveness for simulations and team training. Prior to ARL, she was a senior research psychologist and NAVAIR Fellow at the Naval Air Warfare Center Training Systems Division. Dr. Johnston received her M.A. and Ph.D. in Industrial and Organizational Psychology from the University of South Florida.

Scenarios for Training Teamwork Skills in Virtual Environments with GIFT

Robert K. McCormack[1], Tara Kilcullen[1], Anne M. Sinatra[2], Tara Brown[1], Jeffrey M. Beaubien[1]
Aptima, Inc.[1], US Army Research Laboratory, Human Research & Engineering Directorate[2]

INTRODUCTION

Breakdowns in teamwork are often cited as a cause for poor, and at times, devastating outcomes that lead to loss of life, limb, and material resources (Wilson, Salas, Priest, & Andrews, 2007). Such failures are often attributed to breakdowns in essential teamwork skills, such as coordination and communication, and emergent team states, such as cohesion and shared situational awareness. The relevance of these team constructs is evidenced in the academic literature across a variety of domains, including both medicine (Hughes et al., 2016) and the military (Sottilare, Burke, Salas, Sinatra, Johnston & Gilbert, 2017; Wilson et al., 2007). Additionally, the Army has recognized the importance of Soldiers demonstrating these teamwork concepts. Specifically, several of the principles of Mission Command outlined in ADRP 6-0 align with these concepts, including "Build cohesive teams through mutual trust" and "Create shared understanding" (U.S. Department of the Army, 2012).

While the importance of these teamwork concepts is recognized, there remain challenges to training them efficiently. To maximize effects while simultaneously minimizing costs, there has been a push toward the use of intelligent tutoring systems (ITSs) and ITS frameworks in training. However, these systems have been almost universally designed to train individuals, not teams. In particular, the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012; Sottilare, Brawner, Sinatra & Johnston, 2017) enables the ITS community to efficiently achieve learning effects for individuals. To date, all of the GIFT-based team tutoring applications have been limited to dyads or triads (Bonner, Walton, Dorneich, Gilbert, Winer, & Sottilare, 2015; Gilbert, et al., 2017). However, in order for GIFT to fully support Army training needs, it must scale to larger team structures, such as squads, platoons, and above. In this paper, the authors outline development of such a system within the GIFT framework and development of supporting training scenarios within the Virtual Battlespace (VBS3) simulation environment. There are two overarching objectives to this effort:

1. The first objective is to demonstrate the utility of GIFT for adaptive team training in rich Virtual Environments (VEs), and specifically VBS3. Previous team tutoring implementations using GIFT focused on dyads and/or triads. The current effort aims to assess the utility of using GIFT for larger organizational structures (e.g., squads).

2. Previous efforts have relied heavily upon expert observer rating scales and self-report surveys of team processes and performance. The current effort seeks to assess the utility of using unobtrusive measurement methods (Orvis, Duchon, & DeCostanza, 2013) instead.

To accomplish these two objectives, the authors are developing a prototype training system in GIFT that can capture meaningful team processes and emergent states in a virtual training environment. In addition, the authors are developing realistic training scenarios that provide sufficient complexity and team interaction opportunities to enable effective team training. Specifically, the authors are developing scenario frameworks that enable GIFT to read data collected unobtrusively from teams training using the VBS3 platform, and computing measures of key teamwork constructs that will be used to assess and debrief team performance. This paper will summarize developments to date towards achieving the above-mentioned goals.

TRAINING ENVIRONMENT

Intelligent Tutoring System

GIFT is a domain-independent intelligent tutoring system framework (Sottilare, Brawner, Sinatra & Johnston, 2017). Much of the research and the efforts to date in GIFT have focused on individual tutoring. However, the ultimate goal of GIFT is for tutoring to be conducted with teams. Both theoretical and practical work has been done with GIFT that will prepare it for scaling up for use with teams. A large-scale literature search and meta-analysis has served as the theoretical foundation for team tutoring in GIFT (Sottilare, Burke, Salas, Sinatra, Johnston, & Gilbert 2017). As part of this effort, relevant behavioral markers were identified for several team constructs. Additionally, initial work has also been done to adapt GIFT for use by multiple users engaging in the same scenario simultaneously.

The first work to implement team tutoring in GIFT created a two-person surveillance task using the Virtual Battlespace 2 (VBS2) software. The task consisted of two individuals (spotters) each monitoring their own sector and reporting to their teammate if a threat was passing to the other's sector (Gilbert et al., 2017). This task demonstrated that GIFT could have two individuals simultaneously engage in a simulation-based environment, and was able to provide feedback based on the actions of both individuals separately, as well as the team as a whole. The next step was adjusting the surveillance task such that it had three individuals working together as a team to achieve their goals. Two spotters continued to monitor their respective areas or responsibility, and a third role – a "sniper" who received information from both spotters – was added. The role of the sniper included receiving information from the two spotters, acknowledging receipt of that information, and making decisions based on it. Through the development of this scenario, it was shown that GIFT was capable of providing tutoring and real-time feedback to a three-person team in a simulation-based environment.

Looking toward the future, it is important to demonstrate that GIFT is capable of tutoring large numbers of individuals simultaneously, such as a squad of, which is typically composed to two four-person fire teams plus a squad leader. Scaling GIFT's team tutoring capabilities will require consideration of not only how to deal with the data of nine separate team members, but also how to measure teamwork within a VE and how to handle different team member roles. Therefore, new approaches should focus on defining roles within a GIFT tutoring scenario, simultaneously assessing the behavior of multiple Soldiers, and efficiently determining the team's overall performance in real-time. The teams' performance will then need to result in the proper feedback being given to the team either during or after engagement with a game scenario. Additionally, future work should find efficient ways to implement team behavioral markers in the GIFT software so that the team's performance can be assessed in real-time.

Virtual Training Environment

To train and assess teamwork skills, the authors utilize the VBS3 software. The decision to use VBS3 was two-fold. First, GIFT has integrated with VBS3 (and previous versions of the Virtual Battlespace software) throughout its development. Therefore, it already interoperates with the VBS3 architecture and data structures. Secondly, VBS3 is widely used throughout the Army. While we are not assuming that all research participants that will come through the training will have had exposure to VBS3, it is a readily available training asset at many Army installations. This will ensure adequate locations and candidates for validation of the training and teamwork measures.

When it comes to infantry, virtual training proves to be an overall challenge. Many virtual training platforms have proven to be ineffective for numerous reasons. These include overly cumbersome or counter-intuitive software interfaces, the system being too time-consuming to set up and tear down, and the lack of validated human performance measures. With the instantiation of VBS3 into their virtual training toolbox, infantry Soldiers and Marines are able to gain valuable, training experiences prior to completing live training. The flexibility of VBS3 – in terms of actions, assets, and customization – means that it can support the development of scenarios that are rich enough to enable measurement of teamwork skills.

Inherent in training and assessing teamwork skills is the ability for individual Soldiers to interact and communicate with one another. VBS3 includes a built-in text chat feature will serve as a primary means of team communication and information coordination, as well as providing a rich set of data for teamwork measurement. Team members will need to communicate about a number of issues throughout the scenario, such as detecting a threat, reporting a threat, and handing off a threat. Interaction and communication complexity can be manipulated by putting constraints on the communication structure. For example, the communication structure can be set up such that certain members of a team cannot communicate directly with members of another team, which mimics communication breakdowns during a mission.

The specifications for an initial VBS3 scenario, as just described, should provide enough complexity to require sufficient teamwork interaction. However, the goal is to make the scenario readily scalable to accommodate different team sizes, as well as to support the training of Soldiers at different expertise levels. The VBS3 simulation engine itself has been shown to support over 100 simultaneous learners, and the structure and number of the teams, assets, threats and communication constraints can be scaled to support more or less complex conditions, as desired.

SCENARIO DEVELOPMENT FOR VBS3

To support teamwork training within VBS3, realistic scenarios are needed that provide ample opportunities for assessment and feedback. The authors have identified a number of constraints for training scenarios:

1. Must be implemented within the constraints of the simulation environment (VBS3);

2. Must represent realistic tasks, interactions, and outcomes to ensure Soldier engagement and buy-in;

3. Must support the training and assessment of teamwork-related constructs (e.g., coordination, communication, cohesion) that emerge as a function of the team members' interactions;

4. Must allow team members to communicate both naturally and in a manner that enables assessment of communications for measurement purposes;

5. Must initially focus on the squad level, but also enable larger team structures to train within the simulation environment;

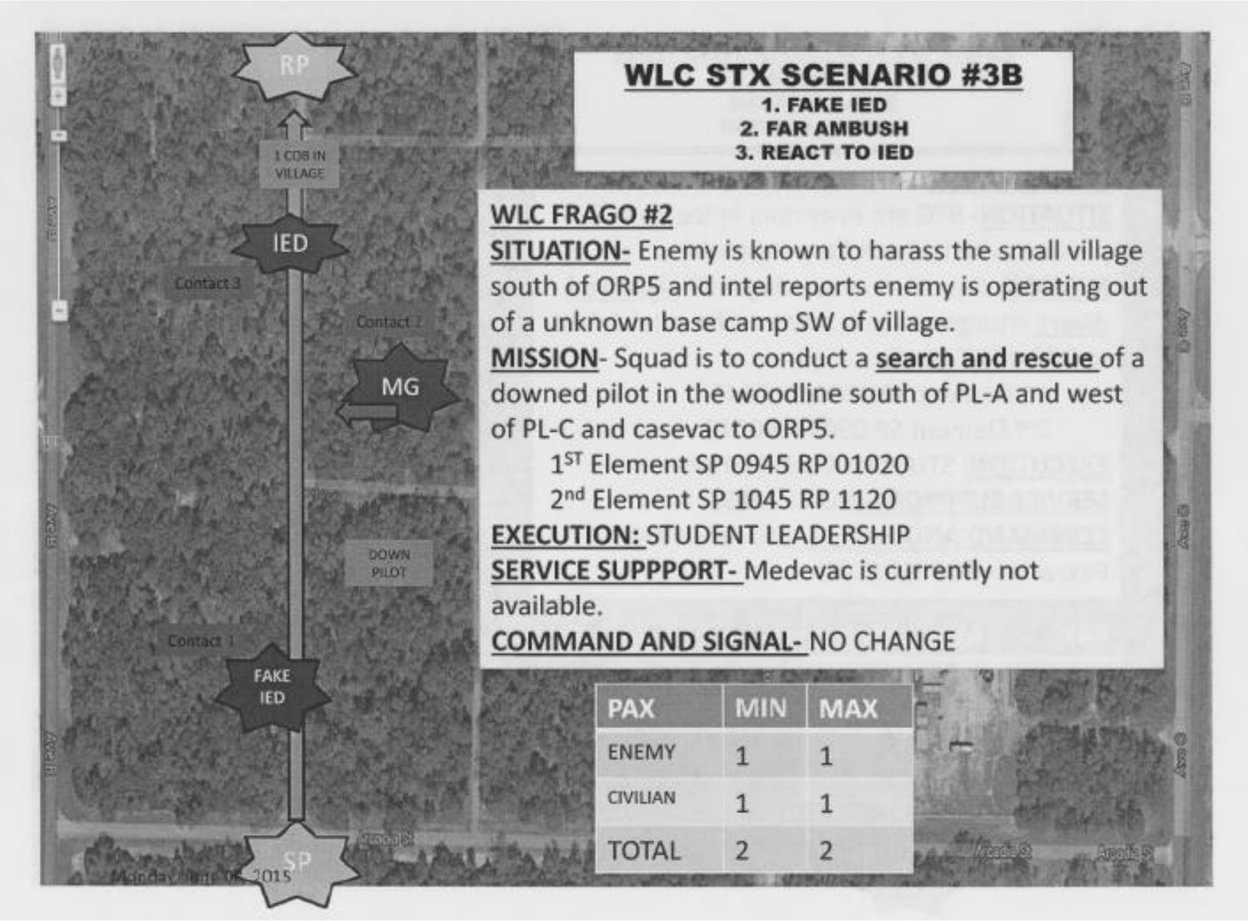
6. Be scalable to support higher echelon training objectives with more complex scenarios.

Scenario Overview

Working with an active duty Army infantry Subject Matter Expert (SME), the research team modified an existing Combat Search and Rescue (CSAR) training scenario that is currently being used at the Army's Basic Leader Course (BLC) to train and assess small unit leadership skills (See Figure 1 for an overview of the scenario). The scenario focuses on search and rescue of a downed pilot within the Area of Operations (AO). The team is a squad-sized element that is comprised of two four-person fire teams. The squad is led by a squad leader (a Sergeant); each fire team includes a fire team leader (a Corporal), as well as a Rifleman, a Squad Automatic Weapon (SAW) operator, and a Grenadier.

The scenario unfolds over a roughly 1-mile linear pathway through a forest which includes a mixture of tall trees and scrub brush that are common to northern Florida (Camp Blanding Joint Training Center). While Soldiers were able to venture from the path into the forest, it both slowed their movement and impaired their visual scan. Because of this, the forest also served as an ideal place for small groups of enemy fighters to launch ambushes against the squad.

Prior to starting the scenario, the Squad Leader is provided with a tactical map of the AO, along with a Fragmentary Order (FRAGO) that describes their mission. The squad leader is also provided with available intelligence (INTEL) about the location of the downed pilot as well as the number and disposition of enemy forces in the AO.



WLC STX SCENARIO #3B
1. FAKE IED
2. FAR AMBUSH
3. REACT TO IED

WLC FRAGO #2

SITUATION- Enemy is known to harass the small village south of ORP5 and intel reports enemy is operating out of a unknown base camp SW of village.

MISSION- Squad is to conduct a **search and rescue** of a downed pilot in the woodline south of PL-A and west of PL-C and casevac to ORP5.

1ST Element SP 0945 RP 01020
2nd Element SP 1045 RP 1120

EXECUTION: STUDENT LEADERSHIP

SERVICE SUPPPORT- Medevac is currently not available.

COMMAND AND SIGNAL- NO CHANGE

PAX	MIN	MAX
ENEMY	1	1
CIVILIAN	1	1
TOTAL	2	2

Figure 32. Notional CSAR Scenario

The squad's primary goal is to rescue a downed pilot. Their secondary goal is to complete a presence patrol in a local village in order to sustain their support against local enemy fighters. Along the way, the squad has to overcome several challenges.

After leaving the starting point, the squad first encounters a suspected Improved Explosive Device (IED). Despite being an enemy hoax, the squad is still required to perform a series of threat-relevant tasks, such as: Confirming (and communicating) the exact location and description of the device; Clearing all personnel to a safe distance; Cordoning (marking) the area to prevent anyone else from entering; Controlling access to the perimeter; and Checking for secondary devices.

The squad then continues down the path toward the estimated location of the downed pilot. Upon reaching the pilot's location, the squad must physically secure the pilot, cordon off the area, apply first aid, and radio headquarters to request medical evacuation. During this time, a local farmer arrives upon the scene towing a wagon full of goods. Before the helicopter can arrive, the squad then needs to apply escalating force to prevent the farmer – who could be an enemy fighter in disguise – from getting within "danger close" proximity to the pilot.

After the pilot is evacuated, the squad continues down the path toward the village. Along the way, they are ambushed by 3-4 enemy fighters who are equipped with small arms, such as AK-47 rifles. The fighters are largely unskilled and have poor aim. As a result, they cause little (or no) injuries to the squad, but this element provides an opportunity to measure how well the squad maintains their formation and responds to the threat, while maintaining their primary and secondary objectives. After dispatching the enemy fighters, the squad leader issues a Situation Report (SITREP) to headquarters, and redistributes ammo among the team. The squad then continues down the path. Along the way, they encounter a second IED, which requires the same set of behaviors that were described previously. Finally, the squad enters the village. At this time, they interact with village leaders – including the mayor, religious leader, and elders.

It is anticipated that during the scenario there will be several points where the scenario is paused and immediate feedback is given. This will provide opportunity for adjustment and recalibration among the team, but requires that opportunities for teamwork measurement occur throughout the entire scenario.

TEAMWORK MEASUREMENT

Based on a review of existing theory and measures, Sottilare et al. (2017) developed a list of behavioral indicators for several teamwork constructs. This set of behavioral markers provides the foundation upon which the research team is developing unobtrusively metrics of teamwork skills. The authors initially decided to target two areas for measure development – task cohesion and physical coordination – which will highlight how different types of data (e.g., communications, scenario interaction data) can be used to measure teamwork skills.

To develop our teamwork measures, the team uses a process based on the Rational Approach to Developing Systems-based Measures (RADSM; Orvis et al., 2013; see Figure 2), which has been successfully used to develop indicators and measures of team states (McCormack, Brown, Orvis, Perry, Myers, 2017).

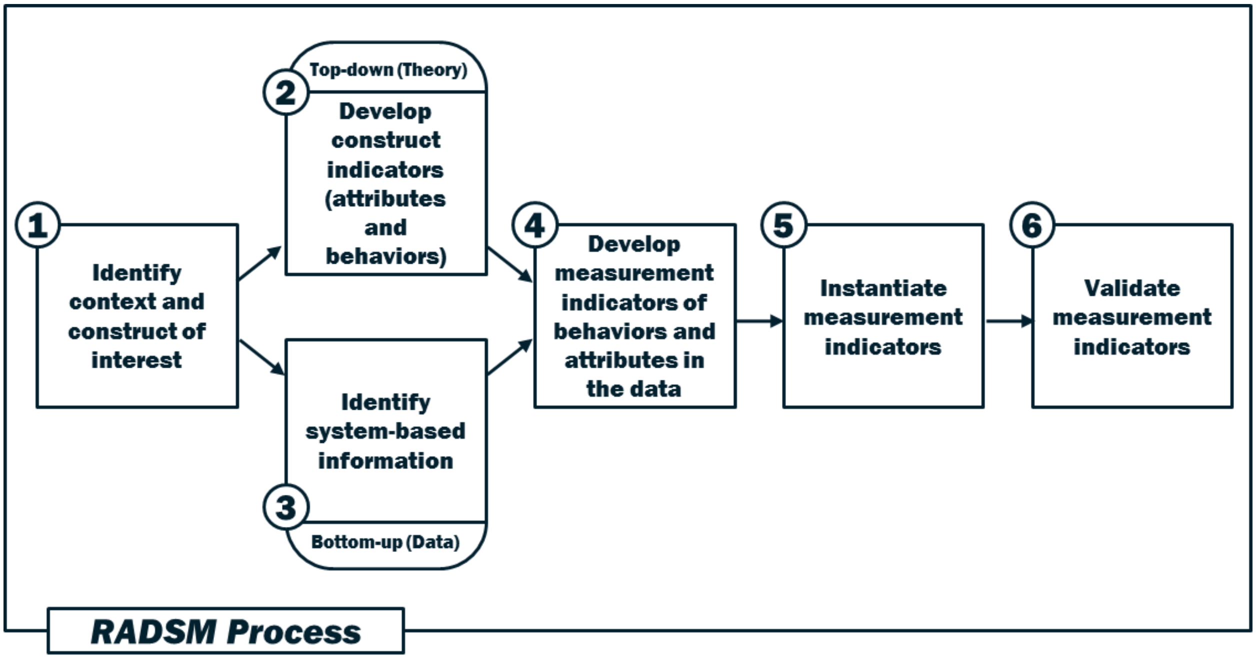

1
Identify context and construct of interest
Top-down (Theory)
2
Develop construct indicators (attributes and behaviors)
Identify system-based information
3
Bottom-up (Data)
4
Develop measurement indicators of behaviors and attributes in the data
5
Instantiate measurement indicators
6
Validate measurement indicators
RADSM Process

Figure 33. The RADSM Process for Measurement Development

The RADSM process consists of several steps, as highlighted in Figure 1, to ensure that developed measures are conceptually sound and contextually relevant. The end result of this process is a set of measures that can be assessed automatically and unobtrusively (that is, not requiring human coding or input) given the data available in the system.

Step 1 is focused on identifying the context and construct of interest for measurement. For the current effort, the context is a teamwork task, described above, performed within VBS3, while the constructs of interest are cohesion and cooperation. Steps 2 and 3 apply top-down and bottom-up approaches, respectively, to measure development. Specifically, in Step 2, the goal is to leverage existing theory to identify behavioral indicators of the constructs that are conceptually grounded. Sottilare and colleagues (2017) have provided a basis for this step. In Step 3, the focus shifts to identifying the available data sources, and specific system-based information, that is available from the environment of interest. The RADSM process is data source-agnostic, supporting data available from a variety of sources. In this case, the goal is to document the various data elements that can be captured from the training scenario. Within VBS3, this data might include text chat, positional data of all entities, sensor actions and results, and weapon fires and remaining ammunitions.

Once the behavioral indicators and list of available data or information is completed, Step 4 consists of bringing these two pieces together to operationalize the indicators using the types of data available in the environment. This transitions the conceptual nature of the behavioral indicators to specific, data-defined performance measures that can be implemented within GIFT. The intent is to develop several operationalized indicators of each teamwork skill, which could each provide unique insight into how the team is doing on that particular skill. It is important to note that any one indicator could be conceptually relevant to a number of teamwork skills, given the conceptual overlap of the teamwork constructs. The goal is to identify a set of indicators, that when used together, do the best job at assessing a unique teamwork skill, such as cohesion. The indicators tied to any one teamwork skill can be implemented and assessed individually, or aggregated to form a single, more comprehensive assessment of a teamwork skill. Table 1 provides examples of what this process looks like when developing measures of task cohesion.

Table 3. Example of RADSM Step 4 for Development of Task Cohesion Measures

Behavioral Marker	How would this be demonstrated?	Data Source(s)	Data Features	Analysis Method(s)
Members are actively working together and pitching in to reach team goals	All team members are communicating with each other	Chat logs	Sender/receiver of chats; number of chat messages sent by person	Compute # of messages sent by each team member; Assess the distribution of communication actions across team members
	Each team member is taking the actions that they are responsible for (e.g., detecting threats in	Movement and action logs; List of team member responsibilities	Who did what action and when	Comparison of user movements/actions against their responsibilities

	their area)			
Occurrences of phrases like "great job everyone", "go team", "you're the best", "good work"; positive affirmations toward the team's work	Team members using these phrases in their chat communications with one another	Chat logs	Sender/receiver of chats; content of chat communications	Dialogue act analysis – sum instances of the use of words and phrases matching those associated with "positive affirmation"

Once the team has compiled a set of operationalized indicators, Step 5 of the RADSM process will focus on implementing these measures in the GIFT environment. During this step, the team specifies the criterion for each measure (e.g., thresholds for effective and ineffective assessments). For example, if the distribution of chat messages across the team is concentrated on one or two individuals, this may indicate low task cohesion and would signal the need for feedback.

Finally, in Step 6, the goal is to validate the measures of the teamwork skills developed in the previous steps. During the development phases of this effort, the primary focus of validation is establishing the face validity of measures. That is, individuals with expertise in teamwork measurement as well as Army SMEs will provide assessment of the utility and accuracy of each conceptual measure. In subsequent efforts, the team will develop and execute controlled experiments of the system using teams of active duty infantry Soldiers to establish and verify the validity of each measure.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This ongoing effort is aimed at training and assessing team performance within the GIFT environment. This serves two purposes: demonstrating that GIFT can be effectively extended from individual training to team training, and demonstrating that reliable and valid measurements of teamwork can be assessed in a virtual team training environment such as VBS3. Our progress to date has shown that there is ample opportunity to deliver rich training experiences through a VE and that there are a plethora of behavioral indicators and measurement opportunities within the scenario. Next steps for this effort include continued development, refinement, and implementation of the scenario inVBS3; development and implementation of the unobtrusive teamwork measures; development of feedback strategies; and validation of the measures through both face validation and rigorous human-in-the-loop experimentation. Future efforts will build upon this work, both in terms of the revised GIFT architecture for supporting team training, but also the scenario and team measures created.

ACKNOWLEDGEMENTS

The research reported in this document was performed in connection with contract number W911NF-17-C-0061 with the U.S. Army Contracting Command - Aberdeen Proving Ground (ACC-APG). The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of ACC-APG, U.S.

Army Research Laboratory or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

REFERENCES

Army Doctrine Reference Publication (ADRP) 6-0, Mission Command. Washington, DC: Headquarters, Department of the Army, 2012.

Bonner, D., Walton, J., Dorneich, M. C., Gilbert, S. B., Winer, E. & Sottilare, R. A. (2015). The Development of a Testbed to Assess an Intelligent Tutoring System for Teams. In Workshop on Developing a Generalized Intelligent Framework for Tutoring (GIFT): Informing Design through a Community of Practice (p. 9).

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., ... & Winer, E. (2017). Creating a team tutor using GIFT. International Journal of Artificial Intelligence in Education, 1-28.

Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. Journal of Applied Psychology, 101(9), 1266–1304.

McCormack, R.K., Brown, T.A., Orvis, K.L., Perry, S., & Myers, C. (2017). Measuring Team Performance and Coordination in a Mixed Human-Synthetic Team Training Environment. In the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2017, No.1).

Orvis, K. L., Duchon, A., & DeCostanza, A. (2013, January). Developing Systems-based Performance Measures: A Rational Approach. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2013, No. 1). National Training Systems Association

Sottilare, R., Brawner, K., Goldberg, B. & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: US Army Research Laboratory. May 2017. DOI: 10.13140/RG.2.2.12941.54244

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. International Journal of Artificial Intelligence in Education, 1-40.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. Human Factors, 49, 243–256

ABOUT THE AUTHORS

***Dr. Robert K. McCormack** is a Principal Mathematician and Lead for the Team and Organizational Performance Capability at Aptima. He has expertise in the areas of unobtrusive measurement, computational linguistics (NLP), machine learning, epidemiological modeling, and human sociocultural modeling and analysis. Dr. McCormack received a Ph.D. and M.S. in Mathematics from Texas Tech University, and a B.A. in Mathematics and Computer Science from Austin College*

***Ms. Tara Kilcullen** is a Program and Customer Engagement Lead at Aptima, Inc. In her role, she supports business strategy, planning, and execution, defines market needs for technology requirements, product maturation and successful transition of science and technology (S&T) research programs, and leads defense programs that differentiate Aptima as an industry leader in Modeling, Simulation and Training. Ms. Kilcullen holds B.A.'s from the University of Pittsburgh and an A.S. from Full Sail University as well as several certifications.*

Dr. Anne M. Sinatra *is part of the adaptive training research team within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab, she is lead on the Team Modeling vector and works on the Generalized Intelligent Framework for Tutoring (GIFT) project. Her background is in Human Factors and Cognitive Psychology.*
Dr. Tara Brown *is a Senior Scientist and Lead for the Instructional Strategy and Support Capability at Aptima. She has expertise in unobtrusive measurement, training, and learning in individuals and teams. Dr. Brown holds a Ph.D. and M.A. in Organizational Psychology from Michigan State University and a B.S. in Psychology from Wright State University.*

Dr. Jeffrey M. Beaubien *is Lead Scientist for Aptima's Learning and Training Systems Division. He also serves as Aptima's Institutional Review Board (IRB) chair. Dr. Beaubien has over 15 years of experience designing and evaluating a broad range of training programs—including classroom-based training, interactive multimedia instruction, distance learning, simulation-based training, game-based training, and mobile learning—for clients in the healthcare, commercial aviation, homeland security, and military sectors. Dr. Beaubien holds a PhD in Industrial and Organizational Psychology from George Mason University, an MA in Industrial and Organizational Psychology from the University of New Haven, and a BA in Psychology from the University of Rhode Island.*

Towards validating a Mission Command Team Training Model in GIFT for Military Populations

Jeanine A. DeFalco[1], Robert Davis[2], Michael Boyce[3],
Erik Kober[2], Ben Goldberg[3]
[1]Oak Ridge Associates University/Army Research Laboratory, West Point, NY/Orlando, FL
[2] United States Military Academy, West Point, NY
[3] Army Research Laboratory, Orlando, FL

INTRODUCTION

While team performance has been the primary focus of team research (Sottilare et al., 2017), this analytical works directionally in a backwards manner, beginning with the end product assessment of successful output to determine the starting points of behavioral, attitudinal, and cognitive constructs that gave rise to that output. Additionally, team performance research has given rise to a breadth and scope of constructs identified and defined in the literature that are numerous, overlapping, and directionally unclear. As such, part of the complexity in unpacking team training models lies in the fact that team product outcomes are the result of numerable variations of institutions with tasks that require unique solutions and outcomes. Therefore, working backwards from a performance outcome approach lends itself to a great many possible model configurations that are almost unwieldly to empirical test.

For the purposes of this paper, we are shifting our lens from team performance outcomes to team formation. Using an existing team model based on the Mission Command paradigm of the US Army, we seek to examine the structural elements that are necessary for effective team formation modeled after this paradigm. While our approach is domain specific, it is our expectation that our analysis on team formation will have broader industry applications.

Specifically, our proposed team training model for GIFT is an adaption of Belbin’s theory of team roles, where the notion of balance of team roles is expanded to incorporate the effects of power/knowledge discourse (Foucault) and adaptive capacity. This approach is based both on qualitative observations conducted at the United States Military Academy (USMA), as well as a review of the literature on research related to team performance (Kjaergaard, Leon, Venables, & Fink, 2013; Sottilare et al., 2017), team role theory (Belbin, 1981; Fisher, Hunter, Macrosson, 1998; Hamada & Sugawara, 2013; Skvoretz, 2016; Liubchenko & Sulimova, 2017), and team learning beliefs and behaviors (Kjaergaard, Leon, Veneables, & Fink, 2013; Massenberg, Spurk, & Kauffeld, 2015; Van den Bossche et al, 2006; Veestraeten, Kundt, & Dochy, 2014).

Accordingly, this paper will first discuss the qualitative observations of team formation observed within the MS200 course in the Department of Military Instruction at USMA that gave rise to an identified Military Command Abdication Narcissistic (MCAN) model of team formation that can be used to inform team training modeling in GIFT. After a discussion of contextualizing the face validity of this model within the broader instructional aims of the cadets as future members of the US Army, we will discuss the Revised Team Role Theory (RTR) components that can serve as a framework for implementing the MCAN team training model. Lastly, we will briefly propose a methodology to validate this framework through a mixed methods research agenda.

MCAN MODEL

We identify the emerging and established military populations as those who are seeking to obtain and those who have already obtained full membership within the US Army. The first population is identified as the cadet learners at USMA working in teams to accomplish learning objectives in preparation for serving in the role as Platoon Leader in the US Army upon graduation. The second population consists of team-sized elements conducting operations at the tactical-level of the US Army. This paper will focus on the first population with an understanding that this cadet population is being trained for incorporation into the second population.

Mission Command model of teams

In analyzing the dynamics of teams in our identified second population, it is recognized that these teams within the operational forces conduct operations at a tactical level functionally under the umbrella of Mission Command as defined by the US Army. Within Mission Command, it is understood that the unit will fight to achieve a small number of key tasks until the point of either being destroyed or heavily attrited. Specifically, that dictates junior leaders will assume responsibilities in the next role in the event a superior becomes incapacitated. In order for this process to work effectively, not only do all members of the team need to have previously demonstrated sufficient competencies in their assigned roles, but a level of trust has to be developed across the entire organization where a tactical unit will still continue “to follow” if a subordinate leader assumes control and essentially must seamlessly adopt a new role within the team. Therefore, as part of developing a team training model within GIFT oriented towards military instruction within a cadet population, elements of role adoption, role execution, and role adoption are key variables that need to be operationalized and assessed in order to support a cadet’s readiness to function within a Mission Command team model.

USMA cadets in Department of Military Instruction

At USMA, the cadet learner cohort is consistently presented with challenges within the Military Science environment. To begin with, the population consists of second-year students with minimal experiential knowledge that consists of the most basic individual military tasks. USMA’s central mission is to educate, train and inspire the Corps of Cadets so that each graduate is a commissioned leader of character committed to the values of Duty, Honor, Country and prepared for a career of professional excellence and service to the nation as an officer in the United States Army. This also includes preparing cadet learners for their future roles as Platoon Leaders in the Army’s Operational Force. Many variables within the venue delineate it from what would be likened to a “normal” college experience.

Academically, the workload is immense comparative to a standard undergraduate curriculum track. For example, it is mandated that a cadet learner execute between 21 and 22 credit hours per semester of their sophomore year. Militarily, the cadet begins their immersive 47-month USMA experience where they have exposure to military development and mentorship that spans the moment they arrive on Reception Day as a freshman until they depart to rapidly integrate into the Army Operational Force. Specific to this discussion involving salient variables of successful team training dynamics, the authors of this paper maintain that the Military Science 200 classroom within the Department of Military Instruction can indeed be categorized as a team unit. Importantly, the Military Program seeks to instill in Cadets the foundational military competencies necessary to win in the US Army, inspiring them to professional excellence and service to the Nation. To accomplish this, the Military Program provides a framework for military education, training, and leader development focused on the roles and principles of being a future tactical Army small unit leader (Platoon Leader. Nested in this higher purpose of the Military Program, the Military Science Program synchronizes across two of the four domains: Military and Academic. Specifically, the Military Science Program looks to develop the small unit leaders' abilities to efficiently and effectively plan, prepare, execute and assess complex tactical missions by way of Troop Leading Procedures and Mission-type orders.

To begin, the cadet learner is quickly immersed in a military environment through the span of their first summer period prior to officially entering into the Corps of Cadets and beginning academic studies. During this period, cadets are exposed to rigorous challenges such as hiking 12 miles with a personal equipment load of 45 pounds, uncomfortable conditions such as constantly being exposed to stifling mid-summer heat often surpassing 95 to 100 degrees while conducting training, and being trained on the most basic military tasks such as rifle marksmanship, combat lifesaver training, and land navigation. Cadets navigate through these experiences individually and collectively, enduring shared hardships alongside one another and rapidly developing their military experiential knowledge base.

The initial summer venue serves as a lab comparable to executing a "hard science" academic degree lab to conduct experiments or test hypotheses. From there, the entire population, segregated into two cohorts of approximately 600 cadets, executes the MS100 curriculum, transferring their initial military experiential knowledge and applying it to fundamental components such as understanding the basic land navigation techniques such as "handrailing." It is important to note that the pedagogical structure of the MS100 curriculum centers on providing foundational declarative knowledge.

Once complete with both the initial lab and classroom experience, the next summer lab experiment, known as Cadet Field Training (CFT) becomes increasingly more difficult where they have to execute military training events both as members of squad and team leaders, navigating various experiences that includes a multi-day field training exercise (FTX) where the cadets remain exposed to the elements and have to conduct multiple small-unit operations such as an ambush or platoon attack. After the lab concludes, the collective population reconvenes to execute the MS200 curriculum.

Unique to the MS200 curriculum versus the MS100 curriculum is that the pedagogical structure completely changes. Cadet learners are forced to learn and retain procedural knowledge consisting of varying conceptual frameworks such as the model to approach Enemy Analysis. The Enemy Analysis framework consists of understanding Composition, Disposition, Strength, and Capabilities. Simultaneous to understanding and anchoring themselves to this framework, they are learning how to craft the narrative to communicate this generated analysis as well as learning where to input the information into the Operations Order, a standardized written medium the Operational Force utilizes to communicate mission-type military orders, essential to the true essence of Mission Command. The facet of shared hardships is an example of one element of their assumed roles as emerging military member.

Other salient elements that emerge from this dynamic include heuristic evaluations of their peers' competencies both inside and outside of the classroom, shared beliefs in their goal orientation in accomplishing assignments, discourse negotiations in problem solving, and adaptability in shifting or adapting to new role assignments within a team when a deficiency is noted or occurs. While the content frameworks are beyond the scope of this paper, taking a closer look at the dynamics of team formation as it relates to completing classroom assignments within MS200 becomes starting point for developing a Mission Command team model that can be employed in GIFT, and can further guide the construction of interventions to correct two commonly occurring dysfunctional team models: the abdication and narcissistic models.

Deviations from Mission Command: Abdication and narcissistic models

The abdication and narcissistic models are two team models that have been identified as dysfunctional and ineffective within MS200, yet adopted by cadets upon being assigned a team assessment task. As open dialogue and group activity is a central pedagogical approach to learning, cadet learners in the course are implicitly and explicitly making their own continual assessments of their peers to determine their competency with course content. These heuristic competency assessments ultimately translate to how cadets self-select and form teams within the classroom. Noticeably, when there is a balance of competency and trust present among self-selected team members, the rudimentary elements of a Mission Command type team are in place. This in turn leads to a successful result in team assessment outcomes, and arguably provides a tangible model of how teams should effectively function in their post-USMA placements. This, unfortunately, is not the only team configuration that emerges. Instead, there is observational evidence that two other team types form that deviate significantly from the Mission Command model. These two other team configurations have been identified as abdication and narcissistic team constructs.

An abdication team construct emerges when a self-selected team of underperforming cadets come together to minimally accomplish an assessment team task. This occurs when cadets create teams where there may or may not be a balance of competencies, but the intent of the team is to accomplish only what is minimally required to pass the assessment with the least amount of effort. In this model, while the team members might trust each other to do their assigned work, they abdicate any responsibility to put forth effort to essentially fight, or more appropriately, struggle, to succeed in their assessment task.

In the narcissistic team construct, this dysfunctional configuration occurs when there is an imbalance of competencies and an absence of trust amongst the cadets. In this model, the overachieving cadet believes their competency is superior to their peers and seeks out groups with substandard partners to insure he or she can produce all the required work independently. While the key tasks might be successfully accomplished, the team itself fails to work as a cohesive unit and in this way fails as a team assessment.

While the dynamics of this MCAN model has been identified primarily within the confines of the USMA classroom, it is still a viable starting place from which to devise a team training model as the patterns of behavior that are exhibited in the classroom at USMA may very well carryover -- if not intervened upon – into the US Army more generally. In this way and within this context, then, designing a team training model devised on the initial observable dynamics and data that emerge within this course is a valid approach. What follows, then, is identifying the behavioral, cognitive, and attitudinal markers that shape the MCAN model so the proposed design of the GIFT MCAN model has clearly articulated possible points of adaptive interventions that can be devised for team training. Accordingly, what follows is an analysis of the relevant behavioral, cognitive, and attitudinal markers that factor into the MCAN model that we term the Revised Team Role theory (RTR) derived in part from Belbin’s (1981) Team Role Theory, Foucault’s notion of power and discourse, and adaptive capacity adopted from ecology and society literature.

A REVISED ROLE THEORY

While researchers are generally moving towards behavioral markers with more objective measures of psychological constructs (Wiese et al., 2015), this approach is limiting in that it does not account for pre-performance team formation elements that should be included in team training modeling. While behavioral markers may be effective to evaluate the cumulative success of a team and the outcomes of team performance, it does not include other markers that inform behavioral performances, such as the function of role adoption in team formation, individual competencies and beliefs, power dynamics in discourse, and adaptive capacities, which could be used as a point of intervention during GIFT team training. Accordingly, this paper suggests unpacking team training through a more comprehensive lens where markers are derived based on the Revised Team Roles Theory (RTR), an adaption and expansion of Belbin's (1981, 1992) original Team Roles theory.

Revised Team Role Theory: Role adoption

Belbin's theory of Team Roles (1981, 1992) maintained that a team's performance could be predicted depending upon the knowledge of each team member's team role. Identifying the role profiles of each team members assigned to specific role types, and assuming there was the requisite balance of types in a team, Belbin maintained you could predict that a team would be high performing. For RTR, this first element of role adoption is akin to Belbin's theory: teams emerge as individuals either are placed in, or self-selected to, roles on a team in order to problem solve and execute activities. The roles assigned and subsequently assumed by the individual starts from a place of competencies. If the individual has the competency to fill a specific role, they assume that role as part of the team. However, at this point, the RTR diverges from the Belbin's model.

The basic premise of RTR is that the notion that successful team outcomes is dependent upon a cumulative hierarchy of role adoption, role execution, and role adaption. Within institutions, teams with specific roles are designed to solve a particular problem or task. The ideal role adoption occurs when an individual's competencies align with the parameters of the specific defined role within the team. For this alignment to happen, competencies and traits must also be taken into consideration in the initial team formation, as these elements will influence the process of team performance. The vetting of competencies and traits happens at two levels: in initial team formation when an individual's competency meets the required role to be filled. The second vetting occurs once team formation is in place, and other team members vet each other heuristically so to individually assess the competencies and traits of team members, and determine the balance of power within a group.

Individual heuristic and more formal evaluations of traits and competencies is an ongoing process in a team, and revisions of prior conceptions of individuals can change as team members either confirm or dispel initial perceptions of competencies through their performance. In this way, perceptions of competencies of is the cornerstone to establishing trust. If you do not believe that your team member is competent to succeed in their assigned task, you will not trust them. However, if they demonstrate competency in spite of your prior belief, then trust can be established, and through the assessment of individual competencies within a team, collective cohesion can be established. This first phase is critical to effect team performance. If the team cannot function because there are failures of competencies or a lack of trust, task will not be effectively or efficiently executed, and communication will be compromised. In short, the role execution phase will be flawed.

Referring back to the MCAN model, one can see how the narcissistic and abdication models of teams emerges based on this first phase of role adoption and competency evaluation. If one member of the team determines that the other members are not competent, they will not trust their team mates to successfully perform their tasks, and accordingly will work and make decisions independently. If, however, collectively the team assesses that there is a lack of competency all around, then the team will readily perform at the lowest acceptable level, compromising an effective and successful team outcome in performance.

Revised Team Role Theory: Role execution

In the second phase of RTR -- role execution -- the objective is to solve a problem that requires the competency of more than one individual, otherwise a team would be unnecessary. As previously mentioned, if there are failures of competency or trust, the execution of the roles to address the assigned tasks may occur, but not at a level of optimal effectiveness or efficiency. If, however, competencies are vetted and trust is established, the execution of tasks may still be compromised if communication is compromised when the power dynamics that shape discourse within a team takes the shape of power as domination rather than power to shape ideas and solutions (Karlburgh, 2005).

Power dynamics are distinct from organizational citizenship behavior (OCB), which according to Organ (1988), is "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of the organization" (p. 4), later redefined by Organ (1997) as "performance that supports the social and psychological environment in which task performance takes place" (p. 95). For this discussion, we are accepting Foucault's premise that power is "everywhere" and that power is not inherently good or bad. Rather, power is a strategy that limits words and actions, but can also open up new ways of acting and thinking (Foucault, 1980).

For example, if an individual in a team engages in a strategy of discourse that seeks to dominate and dictate the shape of ideas and decisions, this erodes trust within the team, dismantles collective efficacy, and impedes a team's ability to use discourse to open up new ways of acting, thinking, and problem solving. Using power to dominate can originate either from an explicit or implicit role hierarchy within a team, where there is an understanding that some roles are more equal and awarded superior rights than others. But power to dominate can also emerge based on the character traits or an individual or set of individuals. Accordingly, understanding how power is used in team discourse is a key element to understanding how teams engage in constructive or destructive communication patterns, sheds light on the difficulty of conflict management, and is instrumental in team cohesion and performance.

Going back to our MCAN model, then, a functional representation of role execution can be operationalized when discourse is equitably engaged upon by team members with a predominance of news ways of acting and thinking in comparison to unproductive words and actions. Our narcissistic model would deviate from the MCAN model in that discourse is not equitably engaged upon by all members. Whereas the abdication model would have equitable engagement of discourse, but the discourse would be unproductive in words and actions.

In sum, power dynamics are realized through discourse that emerges during role execution, through resisting or complying with power strategies, and mediated by individual traits, such as personality. In this way, understanding the parameters of the roles adopted by individuals is as important as understanding how traits interact with strategic power negotiations. If roles are rigid, and power dynamics are non-negotiable, then communication and conflict management will be constrained – even if trust and collective cohesion have previously been established.

Revised Team Role Theory: Role adaption

The last element to consider in defining the MCAN model is the notion of role adaption, or adaptive capacity. Seen mostly in the literature of ecology of human societies, adaptive capacity refers to the conditions that enable people to anticipate and respond to change, and recover from and minimize the consequences of change (Adger and Vincent, 2005). For the purposes developing a team training model, adaptive capacity includes the notion of reflexivity, which is a group level construct on the ability for teams to reflect, communicate, and adapt objectives, decision-making and processes, (Widmer, Schippers,

& West, 2009), as well as an individual's ability to shift, change, and adopt roles as needed. However, it also includes traits such as resilience, self-efficacy, innovative thinking, and selective retention (the ability to analyze and reason logically) (Brown 7 Westaway, 2011) that allow for individuals to move in concert beyond their initial adopted role and shift into new ones. In short, adaptive capacity is a key element in the Mission Command model, and including it a team training model is instrumentally important. Key markers for adaptive capacity, then, include the cumulative effect of successful role adoption that includes trust and collective efficacy, successful role execution including constructive discourse, with the additional individual traits that allow for new ways of thinking and acting independently so to reconfigure team roles. In this way, RTR makes plain how team training is an ongoing, hierarchical, cumulative and iterative process – and the necessary components to configure in a MCAN model for GIFT.

PROPOSED METHODOLOGY TO VALIDATE MCAN MODEL & RTR

As part of the ongoing project in skill decay that is currently in development with the Department of Military Instruction at USMA, the authors of this paper propose a mixed method approach to validating the cumulative, hierarchical MCAN model of team training. Qualitative observations on team dynamics will be conducted in the classroom, coding affect and behavior using the BROMP method while cadets are engaged in team assignments. Further, while cadets are engaged in using GIFT to complete team assignments, log files of interactions and communications will be captured and analyzed. Depending upon the actions/interactions and consequences of observed behavior, the next phase of validating the MCAN model would include a quasi-experimental study that would integrate self-survey instruments, such as self-efficacy, HEXACO personality test, with periodic surveys to evaluate the heuristic beliefs of cadets over the course of a semester. Structural equation modeling will be used to test our cumulative hierarchical MCAN model using data from team assignments completed both via face-to-face and through GIFT.

CONCLUSION

This paper proposed how to best model effective team tutoring for both emerging and established military populations. As a derivative of a concurrent effort to address how to best support content mastery and remediate skill decay on an individual level, the authors identified a target team model, MCAN, as well as articulated a cumulative, hierarchical framework (RTR) to identify behavioral, cognitive, and attitudinal markers that can be used to build the MCAN model in GIFT. While this MCAN model and RTR framework is devised from qualitative observations and a review of the relevant literature, future work in this area includes executing a mixed method approach to empirically validate this model to obtain evidence towards adopting this comprehensive design architecture for military team training in GIFT.

ACKNOWLEDGEMENTS

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number **W911NF-17-2-0152**. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

Brown, K., & Westaway, E. (2011). Agency, capacity, and resilience to environmental change: lessons from human development, well-being, and disasters. Annual review of environment and resources, 36.

Fisher, S. G., Hunter, T. A., & Macrosson, W. D. K. (1998). The structure of Belbin's team roles. Journal of Occupational and Organizational Psychology, 71(3), 283-288.

Foucault, M. (1980), Power/Knowledge: Selected Interviews and Other Writings 1972–1977, London: Harvester Press, p.104.

Foucault, M, (1981). Archaeology of Knowledge and the Discourse on Language (1969) (trans. AM Sheridan Smith, 1972), 135-140 and 49. See also M Foucault 'The Order of Discourse' in R Young (Ed) Untying the Text: A Post-Structuralist Reader (1981).

Hamada, D., & Sugawara, T. (2013). Autonomous decision on team roles for efficient team formation by parameter learning and its evaluation. Intelligent Decision Technologies, 7(3), 163-174.

Kjærgaard, A., Leon, G. R., Venables, N. C., & Fink, B. A. (2013). Personality, personal values and growth in military special unit patrol teams operating in a polar environment. Military Psychology, 25(1), 13-22.

Lai, J. Y., Lam, L. W., & Lam, S. S. (2013). Organizational citizenship behavior in work groups: A team cultural perspective. Journal of Organizational Behavior, 34(7), 1039-1056.

Liubchenko, V., & Sulimova, I. (2017). Examining the attributes of transitions between team roles in the software development projects. Eastern-European Journal of Enterprise Technologies., 1(3 (85)), 12-17.

Massenberg, A. C., Spurk, D., & Kauffeld, S. (2015). Social support at the workplace, motivation to transfer and training transfer: a multilevel indirect effects model. International Journal of Training and Development, 19(3), 161-178.

Organ, D. W. (1988). Organizational citizenship behavior: The good soldier syndrome. Toronto: Lexington Books. Organ, D. W. (1997). Organizational citizenship behavior: It's construct clean-up time. Human Performance, 10, 85–97.

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. Journal of Management, 41(3), 769-788.

Senior, B. (1997). Team roles and team performance: is there really a link?. Journal of occupational and organizational psychology, 70(3), 241-258.

Skvoretz, J., & Bailey, J. L. (2016). "Red, White, Yellow, Blue, All Out but You" Status Effects on Team Formation, an Expectation States Theory. Social Psychology Quarterly, 79(2), 136-155.

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. International Journal of Artificial Intelligence in Education, 1-40.

Veestraeten, M., Kyndt, E., & Dochy, F. (2014). Investigating team learning in a military context. Vocations and learning, 7(1), 75-100.

Widmer, P. S., Schippers, M. C., & West, M. A. (2009). Recent developments in reflexivity research: A review. Psychology of Everyday Activity, 2(2), 2-11.

ABOUT THE AUTHORS

***Dr. Jeanine A. DeFalco** is an Adaptive Training Research Scientist and Post-Doctoral Research Fellow with the Army Research Laboratory, Human Research & Engineering Directorate Training Technology Office (RDRL-HRR) Orlando, Florida, working out of the United States Military Academy at West Point, NY. She received her PhD in Psychology from Columbia University, specializing in Human Development/Cognitive Studies in Education with a concentration in Intelligent Technologies. Jeanine's current research includes developing and testing pedagogical models for the Generalized Intelligent Framework for Tutoring to determine the relationship of creative and analogical reasoning in accelerated expert problem-solving in critical care medical education.*

***CPT Robert Davis** has been a professor of Military Science over the past year at the United States Military Academy at West Point. He recently graduated from Fordham University, receiving a Master's of Science in*

Education. He is an Armor Officer who has had various military assignments, including positions within the 82nd Airborne Division and 101st Airborne Division (Air Assault). He has served on three combat deployments, once in Iraq, two times in Afghanistan.

Dr. Michael Boyce *is a research psychologist with ARL's adaptive training research program. For the past 3 years his emphasis has been in using technologies like GIFT to accurately assess learner knowledge and performance. Located at the United States Military Academy at West Point, his goal is to better inform the research progress of GIFT through interactions with a military student population. He received his Ph.D. in Applied/Experimental Human Factors Psychology from the University of Central Florida in 2014.*

LTC Erik Kober *has been the Chief of Military Science at the United States Military Academy since 2016. He graduated from the United States Military Academy in 1997. He is a Senior Army Aviator who has had a variety of military assignments to include principle duty assignments to Fort Bragg, and Fort Hood, multiple and varied, deployments to include Bosnia-Herzegovina, Afghanistan, and Iraq (x3), and command experiences including Troop (C/1-6 CAV) and Battalion (HHBn(P), XVIII ABC) Command. Erik holds a Military Masters of Arts and Science from the Advanced Military Studies Program (AMSP) at the Army's School of Advanced Military Studies (SAMS), Command and General Staff College at Fort Leavenworth, KS, as well as a Masters of Business Administration from Webster University.*

Dr. Benjamin Goldberg *is an adaptive training scientist at the Army Research Laboratory's Human Research and Engineering Directorate. He leads research focused on instructional management within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).*

THEME V: INSTRUCTIONAL MANAGEMENT

Instructional Management in the Generalized Intelligent Framework for Tutoring: 2018 Update

Benjamin Goldberg[1]
U.S. Army Research Laboratory[1]

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT) serves as a flexible domain-agnostic architecture used to author, deliver, and evaluate computer-based tutoring systems. An end state objective of the GIFT program is to establish a set of defacto best practices that guide the development processes when building adaptive training solutions across military, industry, and academic domain applications. To drive this need, a research vector dedicated to instructional management functions was established. This research vector is used as a road-mapping function to establish capability needs and potential R&D paths to meet recognized gaps. Serving as a framing discussion, we provide an introduction to ongoing work described in the instructional management focused chapters to follow. In addition, we briefly describe new pedagogical functions being developed that have yet to be reported.

Instructional Management Research Vector

In 2015 members of the GIFT team published a research outline that examined specific goals and interests associated with instructional management in ITS type environments (Goldberg, Sinatra, Sottilare, Moss & Graesser, 2015). The authors identified the following dimensions as critical benchmarks in driving capability enhancements:

- Guidance and Scaffolding: focuses on identifying a set of pedagogical best practices that adhere to the tenets of learning and skill development. The challenge is identifying methods that generalize across domains and task environments, and providing tools flexible enough to create scaffolding that can be represented in domain-agnostic terms. Current research aims at creating logic to manage timing, specificity, and modality determinations of intervention content at the individual level.

- Social Dynamics and Virtual Humans: focuses on the social component of learning, and building tools and methods that adhere to the social cognitive tenets of how individuals interact to instill knowledge and solve problems (Bandura, 1986; Vygotsky, 1978). From an adaptive instructional management standpoint, social dynamics is concerned with: (1) using technology to replicate interactive discourses common in learning and operational settings, (2) using technology to create realistic and reactive virtual humans as training elements in a simulation or scenario, and (3) using technology to create social networks for the purpose of supporting peer-to-peer and collaborative learning opportunities.

- Metacognition and Self-Regulated Learning (SRL): focuses on instructional management practices that aim at building habits linked to successful regulation of learning practices and that promote metacognitive applications. This approach to instructional management varies from traditional guidance and scaffolding techniques as it focuses on behavior and application of strategy, rather than on task dependent performance. This research area is of interest as it is based around GIFT supporting SRL, and the efficacy of defining and modeling persistent metacognitive strategies that can be applied across domain applications. The goal is to embed instructional supports that promote situational awareness, and guide learners in planning, monitoring, and reflection based activities.

- Personalization (Occupational and Non-Cognitive Factors): focuses on the use of learner dependent information to personalize a training experience. This can involve personalizing content based on interests, with the goal of inducing a higher level of motivation when the context of a learning event is framed within a use case the learner cares about. In addition, the personalization dimension is also interested in identifying ways to automatically personalize training interactions based on occupational factors that are unique to their upcoming assignment or cur-rent job description. All of these instructional management practices require research to identify mechanisms for easily implementing personalization techniques, along with empirical evidence supporting their application for wide GIFT application.

The dimensions reviewed above provide a means for organizing and prioritizing efforts to enhance GIFT's current instructional management support. The dimensions should be ever evolving, as the needs and requirements of the end user is ever changing. To meet a near-term push to modernize the use of live, virtual and constructive simulations to train collective and team-oriented tasks across the Army, a major focus on instructional management research moving forward needs to be focused on team development and cohesion, as well as application of adaptive training applications in live environments through mobile device technologies. Each of these new problem spaces will be expanded upon as future programs mature.

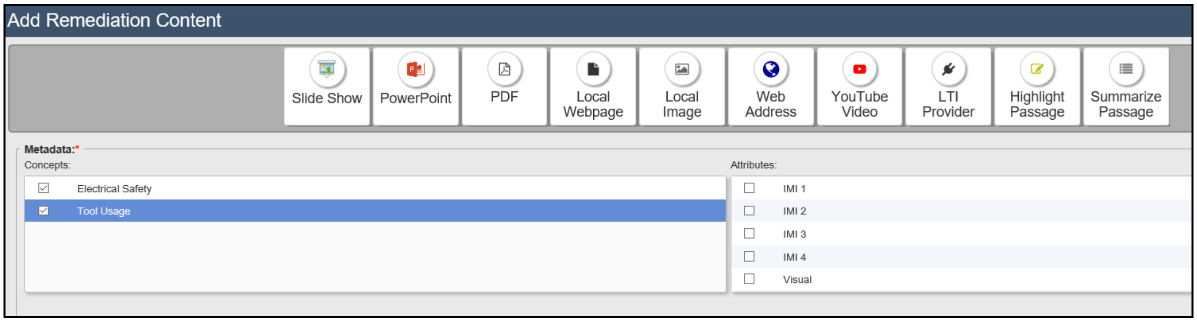
In the remainder of this chapter, we present the 2018 current state of practice for instructional management in GIFT, as those piece parts are the ultimate methods rolled out to the community at large. Following, we review ongoing efforts and how they apply to future enhancements that aim to meet the goals of the overarching instructional management capability dimensions. We end the review with new instructional management efforts that are based on new training concepts centered on worked examples in game-based environments and mobile computing technologies.

2018 INSTRUCTIONAL MANAGEMENT PRACTICE IN GIFT

Enhancements to the Baselines

In the latest public release of GIFT, there have been many updates to the baseline that need to be noted. First, in an effort to extend the remedial capabilities of the Engine for Management of Adaptive Pedagogy (EMAP) to go beyond the passive delivery of new content and information, the previously reported ICAP activities framework was established in GIFT's Adaptive Courseflow object (Chi, 2009; Rowe, Pokorny, Goldberg, Mott & Lester, 2017). The ICAP-Inspired EMAP course-object now supports a configurable 'Remediation' phase (see Figure 1). In this block of the interface an author is tasked with configuring available content and feedback strategies dedicated for remediation purposes only. During this portion of the authoring experience, GIFT requires authors to specify metadata that corresponds with the concept that activity or content targets, and the classification of Constructive, Active or Passive determinations as they relate to Chi's specified activity levels.

This new remedial content addition is now available to all GIFT users. However, it must be noted that in its current state, selection of remedial content is managed by a policy set to randomly select among the ICAP configured resources. Ongoing work, which is reported below by Rowe et al. (2018), will establish the first set of data-driven policies within the domain of COunter-INsurgency (COIN) based on a probabilistic tutorial planning approach.



Metadata:*

Concepts:

☑ Electrical Safety

☑ Tool Usage

Attributes:

□ IMI 1

□ IMI 2

□ IMI 3

□ IMI 4

□ Visual

Figure 1. Remediation Content Configuration Interface in the ICAP-Inspired EMAP Course Object

GIFT Personalization and Management through Learning Tool Interoperability Standards

Next, to support efforts related to GIFT managing interaction across Massive Open Online Courses (MOOCs), development tasks were instituted to make the architecture compliant with the Learning Technology Interoperability (LTI) standards (IMS Global Learning Tools Interoperability Implementation Guide, 2012). The LTI specification establishes application programming interfaces with learning management systems. From this perspective, a learning management system is designated as an LTI consumer, while systems that provide learning activities themselves are considered LTI providers.

For GIFT, two instances of LTI integration were implemented. First, GIFT was established as an LTI provider, where a learning management system can direct a MOOC learner to a GIFTCloud configured lesson for adaptive pedagogical delivery. As an example, GIFT is utilized within a course managed by the site edX.org, where an established lesson incorporates GIFT lesson activities, with completion scores communicated back to edX following execution (Aleven et al., 2017). Next, GIFT was modified to serve as an LTI consumer, where GIFT can call upon LTI providers for support in lesson execution. In this instance, GIFT can now direct a learner to an LTI compliant application to support instruction or practice on specified concepts. As an example, GIFT can now direct a learner to a Cognitive Tutor application within the GIFT lesson flow, where learner and pedagogical modeling controls are handed to that LTI client. Following completion, a score is provided back to GIFT for tracking purposes.

One of the recognized shortfalls of this integration is the reported measure back. Currently, it is a value between 0 and 1, which is used to classify the performance for all assessments performed within that provider application. At the moment, that is not enough granularity to inform intended competency tracking functions GIFT’s overall aim strives to support. With new development efforts to support GIFT as an LTI consumer, new pedagogical paradigms are now made available. Through these mechanisms, GIFT can now re-direct a learner to an LTI provider within the flow of a GIFT configured lesson, which makes GIFT the managing application that guides the ultimate experience. However, seeing as the data provided following an LTI provider interaction are not granular enough to inform complex competency modeling techniques, future research efforts examining how best to manage LTI oriented data feeds is needed.

Enhancements Still Under Development

Establishing Policies in the ICAP-Inspired Engine for Management of Adaptive Pedagogy

With an infrastructure in place to support the ICAP-Inspired EMAP instance described above, the next step is establishing data-driven policies that will dictate run-time pedagogical decisions. To support this development task, experimentation using the Amazon Mechanical Turk platform is being prepared. This will enable the collection of a data set that will ultimately be used to generate a set of simulated students based on the distribution properties of the collected data points. This will enable replicating multiple instances of learner interactions to garner enough data to establish valid policies to inform the ICAP remediation determinations. The methodology to build the simulated student data set is described in last year's GIFTsym proceedings (Rowe, Pokorny, Goldberg, Mott & Lester, 2017), with a breakdown of the testbed development to support this effort described in this year's proceedings (Rowe et al., 2018). Following the creation of policy specifications, a reinforcement learning backend will be established to enable policy weight adjustments as evidence is collected on the utility of specified remedial materials.

NEW INSTRUCTIONAL MANAGEMENT EFFORTS

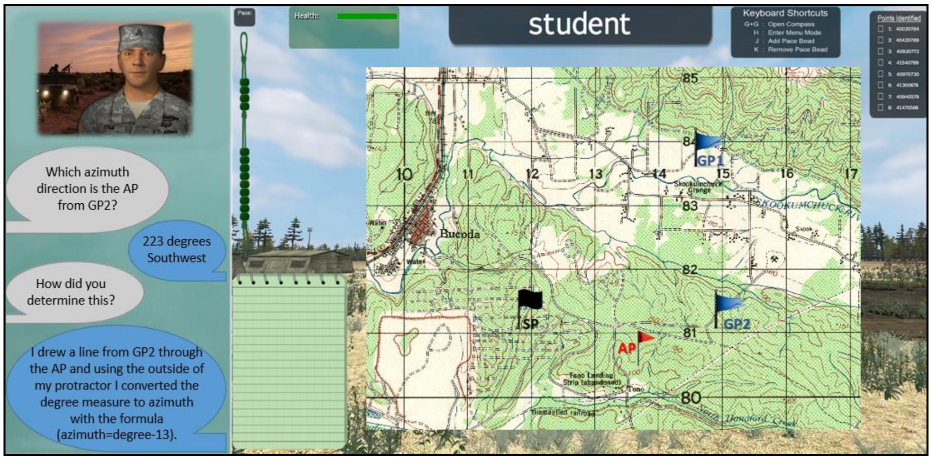
As an extension to last year's update at the 2017 GIFTsym, this section is used to present new efforts currently being worked in the GIFT program that have not yet been reported upon. Each effort is currently in the early stages of implementation, with future experimentation planned across each capability. What is important to note as a grounding function is that each project presented is being applied within the domain of Land Navigation. The domain was selected due the amount of content and scenarios available to train the knowledge, skills, and abilities (KSAs) associated with land navigation execution, as well as excellent support from Subject Matter Experts (SMEs) that will guide assessment and remediation policies.

For initial implementation, the following mechanisms are being researched: (1) using structured interviews in GIFT to facilitate scaffolded worked examples as it relates to procedural tasks that incorporate discrete inputs required to execute a task (e.g., plan a route from one point on a map to another), (2) using mobile app technologies and cloud-computing to guide self-regulated training exercises by blending the physical environment with didactic instruction and personalized assessments (e.g., conducting terrain association exercises), and (3) using metacognitive modeling techniques to track learner competencies across disparate training applications and using persistent models to drive feedback interventions. Each project will be explained in more detail below.

Scaffolded Worked Examples across Procedural Tasks

Worked examples provide a means to guide novice learners through procedural activities, where each step within that activity can be discretely defined for the purpose of guiding execution. In this instance, a system can provide the solution path to a defined problem, with directed engagement with students at specific steps within that process for the purpose of assessing understanding and correcting errors and misconceptions. This pedagogical approach has proven effective across many domains, most of which provide well-defined procedural tasks that require consistent execution to obtain an appropriate solution (Durlach & Spain, 2012). From this perspective, GIFT's survey authoring system is being used as a basis to establish structured interviews for the purpose of using worked example instructional methods. These interviews associate with a set of procedurally related questions that are commonly applied across a set of tasks. For each question, a specific concept or sub-concept can be targeted, with contextualized responses based on the scenario that serve as the assessment criteria. With this framework in place, specific steps within a solution path can be remediated, where focused interventions address direct misconceptions and impasses that result from the learner's input to a step.

As an example in the domain of land navigation, scenarios are designed in a game-based environment to train all concepts associated with dead-reckoning procedures (i.e., navigating from one point to another using a compass and a map). Trainees are responsible for locating points on a map, determining an azimuth to guide the direction by which they walk, determining an estimated distance, and identifying land features to help them orient as they walk. If the learning objective of the training event is to provide multiple opportunities to apply dead-reckoning procedures, then each discrete task can have an associated structured interview that can guide that interaction. Each task requires the same steps, with each input having new contextualized responses based on where they are on the map and where they are supposed to go. Once these interviews are in place, new logic can be established to infer a confidence state in a learner's ability to provide the correct response on each step within the interview. With a high confidence rating, GIFT has the ability to adapt the pedagogical approach by modifying the complexity of the task. Rather than prompt the trainee for inputs on the required steps, the task can re-orient and instruct the trainee to navigate to the next point with a specified time constraint, thus increasing the difficulty and leaving the trainee to execute on their own accord.



Which azimuth direction is the AP from GP2?

223 degrees Southwest

How did you determine this?

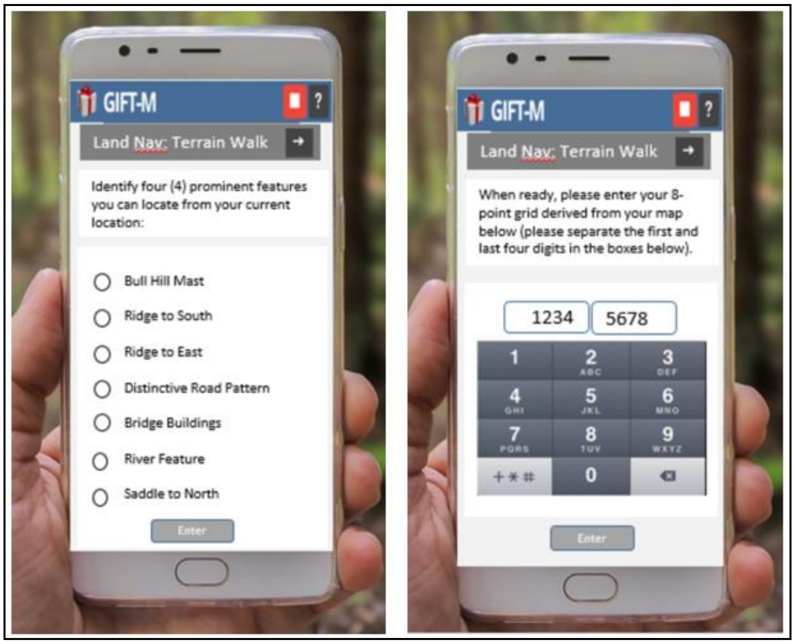
I drew a line from GP2 through the AP and using the outside of my protractor I converted the degree measure to azimuth with the formula (azimuth=degree-13).

Figure 2. Sample Question from GIFT Worked Example Structured Interview

This new instructional management concept has led to some structural changes to GIFT's Domain Knowledge File (DKF), as well as to the survey authoring system. To support direct numeric inputs that orient with map grid points, azimuth directions, and estimated distances, GIFT can now deliver a survey with an open numeric input response with configurable assessments based on exact inputs, or inputs that fall within a defined range. Next, GIFT's concept structure in the DKF will be leveraged to associate a specific question with a specific sub-concept so that remediation and feedback can be contextualized on the procedure step that scores below-expectation. In addition, new pedagogical logic will need to be developed that can adjust the conditions and standards of a defined DKF Task, based on the outcomes of the tasks completed before it. In this example, observing effective execution of two tasks in a row under the scaffolded worked example can lead to a pedagogical shift to increase the complexity by removing the help functions.

Live Training with Mobile Intelligent Tutoring Functions

Another effort being worked with land navigation serving as a guiding domain is the first development of a GIFT mobile application. In this instance, GIFT is leveraging real-time positional and movement data to trigger training events in a live environment through the delivery of contextualized content, tasks and assessments (Goldberg & Boyce, 2018). The notion here is to extend the training space into the actual operational environment and embedding structured learning activities that utilize the elements of the space they are occupying. As an example, the first mobile application being developed is to support an exercise called a Terrain Walk. During this exercise, a trainee completes a specified course where designated spots along the path are used to train directed concepts that associate with land navigation fundamentals. In the traditional sense a Terrain Walk is completed by a live instructor with a group of trainees. To support a self-regulated delivery approach, the idea is to replace the instructor with a smart phone, where each trainee receives a personalized experience.



Identify four (4) prominent features you can locate from your current location:

○ Bull Hill Mast
○ Ridge to South
○ Ridge to East
○ Distinctive Road Pattern
○ Bridge Buildings
○ River Feature
○ Saddle to North

When ready, please enter your 8-point grid derived from your map below (please separate the first and last four digits in the boxes below).

1234 5678

Figure 3. GIFT Mobile App Example Interactions for Terrain Walk

To support this implementation, GIFT has been configured to consume cellular network traffic data to monitor the exact location of an individual as they navigate through an environment. With this new data type, GIFT's DKF can be configured to use location data to inform task start triggers that associate with a task, the concepts linked to that task, and its respective assessments used to infer performance and competency. When a trigger is recognized, GIFT can now deliver content, task directions, and deliver assessments through survey items (see Figure 3). The DKF applies timing functions to guide the delivery of content and items to assist in making the user experience an enjoyable one. Following completion of the first iteration of the GIFT Mobile App to support a Terrain Walk, there will be a designated data collection this summer at the United States Military Academy.

Metacognitive Training across a Network of Simulations

The third new effort using land navigation as a guiding function is extending the learner modeling techniques in GIFT to support metacognitive training across a network of training environments. This approach is based on prior work aiming to establish a hierarchical approach to learner modeling that focused on cognitive skills, cognitive strategies, and metacognitive abilities (Rajendran, Mohammed, Biswas, Goldberg & Sottilare, 2017). This approach was originally developed within the domain of COIN using the game UrbanSim. Now, the learner model framework is being re-applied to land navigation, where approach will manage interactions across three distinct training events that focus on a crawl/walk/run modeling of training (Goldberg, 2017). In this example, the hierarchical student model will be used to infer KSAs as trainees interact with a virtual sand table to learn terrain association concepts, interact with a virtual game to rehearse dead-reckoning procedures, and interact on a live land navigation course. This approach requires the first implementation of a persistent learner model that can track experiences across a number of scenarios and lessons and use those recorded experiences to personalize future interactions through GIFT supported pedagogical functions. This effort is just starting, with much to share in future reporting.

FUTURE CONSIDERATIONS

As mentioned above in the introduction, team intelligent tutoring is a desired capability moving forward across the Department of Defense. With that said, a majority of the instructional management functions built in GIFT as of now are dedicated to the individual learner. Future research is required to implement pedagogical approaches to managing team interactions across the planning, execution, and review phases of a training exercise. Currently, there is much written on how to monitor and measure team development (Sottilare et al., 2017), but there is little contribution to the literature on instructional management techniques that associate with technology-based interventions. To this end, a pedagogical framework is required to associate with feedback and scenario adaptations that are based on team and task structures. Current chapters in the soon-to-be released GIFT Recommendations books will explore some notional theoretical approaches, with sports psychology playing a role in their instantiation.

CONCLUSION

In this chapter, we present current and future instructional management functions that are being built into GIFT. This review covers the last twelve months of development, with the introduction of new capabilities being rolled into the publicly available baseline, while future capabilities reviewed are being developed to support data collections and future extensions to be included in subsequent releases. With GIFT continually evolving to include more AI driven methods, future enhancements to GIFT’s instructional management functions will continue to mature that focus on data-driven agent methods, as well as exploring new approaches to manage team structures.

REFERENCES

Aleven, V., Baker, R., Blomberg, N., Andres, J.M., Sewall, J., Wang, Y. & Popescu, O. (2017). Integrating MOOCs and Intelligent Tutoring Systems: edX, GIFT, and CTAT. *In proceedings of 5th Annual GIFT Users Symposium*. Orlando, FL.

Bandura, A. (1986). Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, N.J.: Prentice Hall.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73-105.

Durlach, P. J., & Spain, R. D. (2012). Framework for instructional technology. In V. G. Duffy (Ed.), *Advances in applied human modeling and simulation* (pp. 222-231). Boca Raton, FL: CRC Press.

Goldberg, B., Sinatra, A., Sottilare, R., Moss, J., & Graesser, A. (2015). Instructional Management for Adaptive Training and Education in Support of the US Army Learning Model-Research Outline: DTIC Document.

Goldberg, B., Davis, F., Riley, J. & Boyce, M. (2017). Adaptive Training across Simulations in Support of Crawl-Walk-Run Model of Interaction. In *Proceedings of the 2017 International Conference on Augmented Cognition*. Vancouver, British Columbia, Canada. July.

Goldberg, B. & Boyce, M. (2018). Experiential Intelligent Tutoring: Using the Environment to Contextualize the Didactic. Paper presented at the Proceedings of the *11th International Conference on Foundations of Augmented Cognition*, Las Vegas, NV.

IMS Global Learning Tools Interoperability™ Implementation Guide (Final Version 1.1). (2012, March 13). Retrieved from https://www.imsglobal.org/specs/ltiv1p1/implementation-guide

Rajendran, R., Mohammed, N., Biswas, G., Goldberg, B. & Sottilare, R. (2017). Multi-level User Modeling in GIFT to Support Complex Learning Tasks. *In proceedings of 5th Annual GIFT Users Symposium*. Orlando, FL.

Rowe, J., Pokorny, B., Goldberg, B., Mott, B. & Lester, J. (2017). Toward Simulated Students for Reinforcement Learning-Driven Tutorial Planning in GIFT. *In proceedings of 5th Annual GIFT Users Symposium*. Orlando, FL.

Rowe, J., Spain, R., Pokorny, B., Mott, B., Goldberg, B. & Lester, J. (2018). Design and Development of an Adaptive Hypermedia-Based Course for Counterinsurgency Training in GIFT: Opportunities and Lessons Learned. *In proceedings of 6th Annual GIFT Users Symposium*. Orlando, FL.

Sottilare, R., Burke, S., Salas, E., Sinatra, A., Johnston, J. & Gilbert, S. (2017). Designing Adaptive Instruction for Teams: a Meta-Analysis, *International Journal of Artificial Intelligence in Education*, 28(*2*), 225-264.

Vygotsky, L. S. (1978). *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

ABOUT THE AUTHORS

Dr. Benjamin Goldberg *is an adaptive training scientist at the Army Research Laboratory's Human Research and Engineering Directorate. He leads research focused on instructional management within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).*

A Blended Approach to Adaptive Learning

Barbara Buck, Ph.D.[1], Matt Genova[1], Robert Sottilare, Ph.D. [2], Benjamin Goldberg, Ph.D. [2]
The Boeing Company[1], U.S. Army Research Laboratory[2]

INTRODUCTION

Adaptive training is often considered the gold standard for addressing the unique training needs of individual users. These unique needs can result from different backgrounds, different experiences, different learning goals, different personal motivations for learning, and different degrees of engagement in the overall learning experience. Adaptivity is the ability of a system to alter (change) itself to better fit or function in a given situation. In order to optimize the learning experience for a unique person, a learning system should adapt to the individual learner or team for the specific situation, much like a human mentor or instructor would adapt to the individual needs of a student.

The goal of an Intelligent Tutoring System (ITS) is to provide automated instruction equivalent to that of a skilled human tutor. ITS development has gained momentum since the 1980's, with numerous automated tutors being developed and applied in both university and Department of Defense settings (Bloom, 1984, Lesgold, Lajoe, Bunzo & Eggan, 1988, Anderson, Corbett, Koedinger & Pelletier, 1995, Hunt & Minstrell, 1994, Graesser & Person, 1994, Cohen, Kulik & Kulik, 1982).

Adaptive training content can be time-consuming and expensive to develop, deliver, and manage. If the adaptive solutions are ever to gain widespread acceptance within the educational and training community, we must find cost-effective ways to develop, deploy and manage content. The U.S. Army Research Laboratory (ARL) has been developing one such solution, the Generalized Intelligent Framework for Tutoring (GIFT). The GIFT program is an ARL effort to develop a framework for personalized, on-demand, computer based instruction to improve the speed and quality of Soldier training (Sottilare, Brawner, Sinatra & Johnston, 2017). In a separate effort, Boeing has been involved in a program of research and development to create an adaptive learning authoring and content delivery system. The Boeing ITS provides a user-friendly authoring environment designed to rapidly create and deliver a rich personalized student-centered learning experience through the modeling of system knowledge, problem-solving rules, and real-time assessment of student performance. The learning experience provides dynamic scenario sequencing, tailored student feedback and student performance summary based on the perceived student strengths and weaknesses (Perrin, Buck, Dargue, Biddle, Stull, & Armstrong, 2007, Perrin, 2009).

In this paper, we will present an aggregate prototype of adaptive learning that leverages these two distinct implementations: ARL's GIFT solution and Boeing's ITS solution. The product of combining these efforts is an integrated adaptive learning prototype. This presentation will describe our efforts to create a seamless adaptive learning experience on the part of the student, as well as plans to conduct an effectiveness study using the adaptive learning methods.

GIFT Framework

GIFT is an open-source, modular architecture developed to ease the burden of authoring, delivering, managing, and evaluating adaptive instruction across a broad array of domains (e.g., cognitive, affective, psychomotor, and social). As an adaptive instructional system (AIS), GIFT guides learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of specific domain learning objectives. GIFT is composed of tools, methods, interfaces, and processes that capture and reinforce best instructional practices, effective learning strategies, and tactical actions for both individual learners and teams. Emerging capabilities include: user dashboards, data analytics, automated content curation, automated after action reviews, and standard messaging for reuse and interoperability.

GIFT has several modules which model and act on data about the learner, instructional decisions, and domain content:

- Domain Module - The primary function of the domain module is to create, maintain and assess domain sessions. This module hosts or points to content used during instruction and contains a domain course file which is an XML file containing information needed to assess the learner's progress toward proficiency for the concepts (learning objectives) identified by the course author.

- Learner Module - The primary function of the Learner module is to determine the learner's state (e.g., real-time performance, real-time emotional, or long term domain competency).

- Sensor Module - The primary function of the sensor module is to read and filter sensor data to determine/predict learner states. There are several sensors integrated with GIFT to provide data about the learner: Microsoft Kinect, Zephyr Bioharness, Affectiva Q Sensor, and others.

- Pedagogical (Instructional) Module - The primary function of the pedagogical module is to use information about the learner's state to generate recommendations (e.g., next course to take) and select instructional strategies (e.g., prompt learner to reflect) to enhance learning. Instructional strategies are passed to the domain module for implementation.

- User Management System (UMS) Module - The primary function of the UMS module is to manage a user session. It is responsible for storing information about the user such as biographical details, in addition to maintaining information about domain sessions. It does not, however, keep scoring records of user's training history. That is handled by the Learning Management System.

- Learning Management System (LMS) Module - The primary function of the LMS module is keep track of a learner's instructional experiences and achievements as part of a history of learning. The GIFT LMS saves the scores of every assessment during every lesson experienced in GIFT.

- Tutor Module - The primary function of the Tutor module is to provide an interface that allows interaction between GIFT and the learner. Often referred to as the tutor-user interface (TUI), this is not a formal module, but is an interface capability.

- Gateway Module - The primary function of the gateway module is to interface with external environments (e.g., game-based simulations). The Gateway Module has interfaces with several applications such as: Distributed Interactive Simulation (DIS) networks, Virtual BattleSpace (VBS) serious game, Augmented REality Sandtable (ARES), Microsoft PowerPoint, Tactical Combat Casualty Care (TC3)/Virtual Medic, and the SCATT Pro Marksman Training Application.

A component of GIFT being utilized specifically for this project is the Engine for Management of Adaptive Pedagogy (EMAP; Goldberg, Hoffman & Tarr, 2015). The EMAP is an underlying pedagogical framework in GIFT based on Merrill's Component Display Theory (CDT; Merrill, 1994). The CDT structures learning across four primary interactions: (1) learning the declarative and procedural RULES of a domain and its associated concepts; (2) seeing EXAMPLES of those rules applied across various contexts for better understanding of the interacting components; (3) RECALLING those associations on your own based on testing approaches; and (4) PRACTICING the application of those rules within dedicated scenarios and problem sets. The EMAP then applies personalization strategies within each of those four interactions based on individuals differences stored in GIFT's learner model (e.g., prior knowledge, motivation, self-regulatory ability, grit, etc.). The EMAP also supports automated remediation loops based on performance outcomes in both the recall and practice interactions. The EMAP configurations are housed in GIFT's adaptive courseflow object, which is the integration point for the resulting testbed developed utilizing the Boeing ITS functions.

Description of the Boeing ITS

Boeing's approach to a learner-centered adaptive training implementation has evolved over the course of the past few years. Initial implementations focused on creation of an architecture and authoring solution in support of intelligent tutoring. The product of this effort was Web-based, SCORM®-conformant computer-based training. Details of this approach is provided below.

The Boeing ITS implementation features 3 components (illustrated in Figure 1): a Student Model, an Instructional Model, and an Expert Model. The Student Model implements a profile of dynamically-maintained variables, each corresponding to one learning objective. These variables are evaluated over a number of observations. As a result, changes due to learning are reflected across exercises, as the score increases due to correct performance, or decreases as errors are made. The amount that scores are changed can be weighted according to the degree to which the action reflects mastery of the learning objective. The amount of change is also adjusted according to the degree of support provided to the student by the ITS in selecting this action.

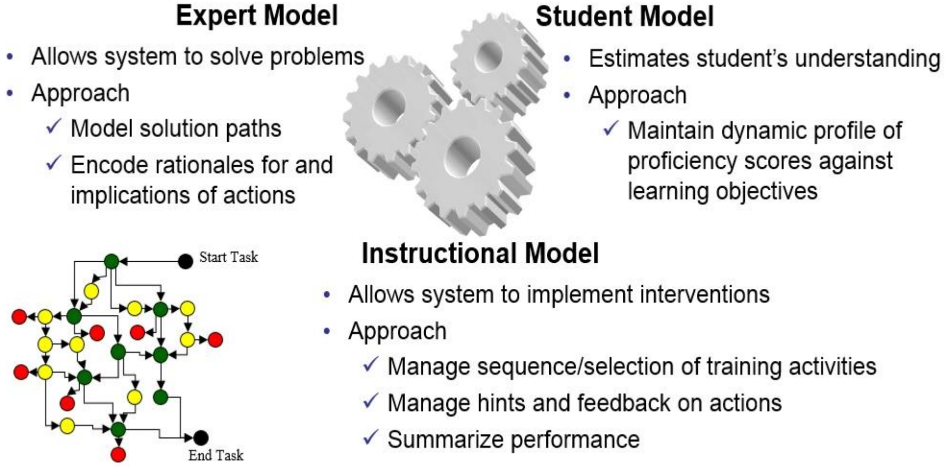

Expert Model
• Allows system to solve problems
• Approach
✓ Model solution paths
✓ Encode rationales for and implications of actions
Student Model
• Estimates student's understanding
• Approach
✓ Maintain dynamic profile of proficiency scores against learning objectives
Instructional Model
• Allows system to implement interventions
• Approach
✓ Manage sequence/selection of training activities
✓ Manage hints and feedback on actions
✓ Summarize performance
Start Task
End Task

Fig. 34. Overview of ITS modeling approach

The Instructional Model responds to student requests for help or student errors with information on problem-solving strategies. The specificity of the information increases as additional requests are made or additional errors occur. The Instructional Model is also tasked with providing within-scenario feedback to guide the student, as well as performance summaries across all learning objectives at the end of the lesson scenario.

The Expert Model is based on cognitive task analysis technique known as PARI, for Precursor, Action, Results, and Interpretation (Hall, Gott & Pokorny, 1995). PARI provides methods to elicit detailed information from experts on how they represent a given state of a solution (what issues have been resolved and what issues remain), optimal and alternative paths to a solution, and their strategies for selecting actions at each step along those paths. The Expert Model directly encodes these solution paths. For each path, the model also captures the expert's summary of the situation (representation of the problem) and the rationales for the possible next steps. Additional details of the ITS architecture and implementation have been published elsewhere (Perrin, 2009).

Details of the Integrated Prototype

As part of a three-year cooperative research and development agreement, Boeing and ARL have been working to develop an integrated adaptive prototype in which we combine the Army's GIFT adaptive learning framework with the Boeing adaptive learning capabilities. The prototype is based on instructing a student on a basic aircraft maintenance task with aspects of troubleshooting and part replacement. In order to perform the task correctly, the student must understand some basics of electrical safety, as well as multimeter usage. Once they have demonstrated an understanding of those basic concepts, then they are taught the fault diagnosis and repair procedure. Basic lesson flow within GIFT is presented in Figure 2.

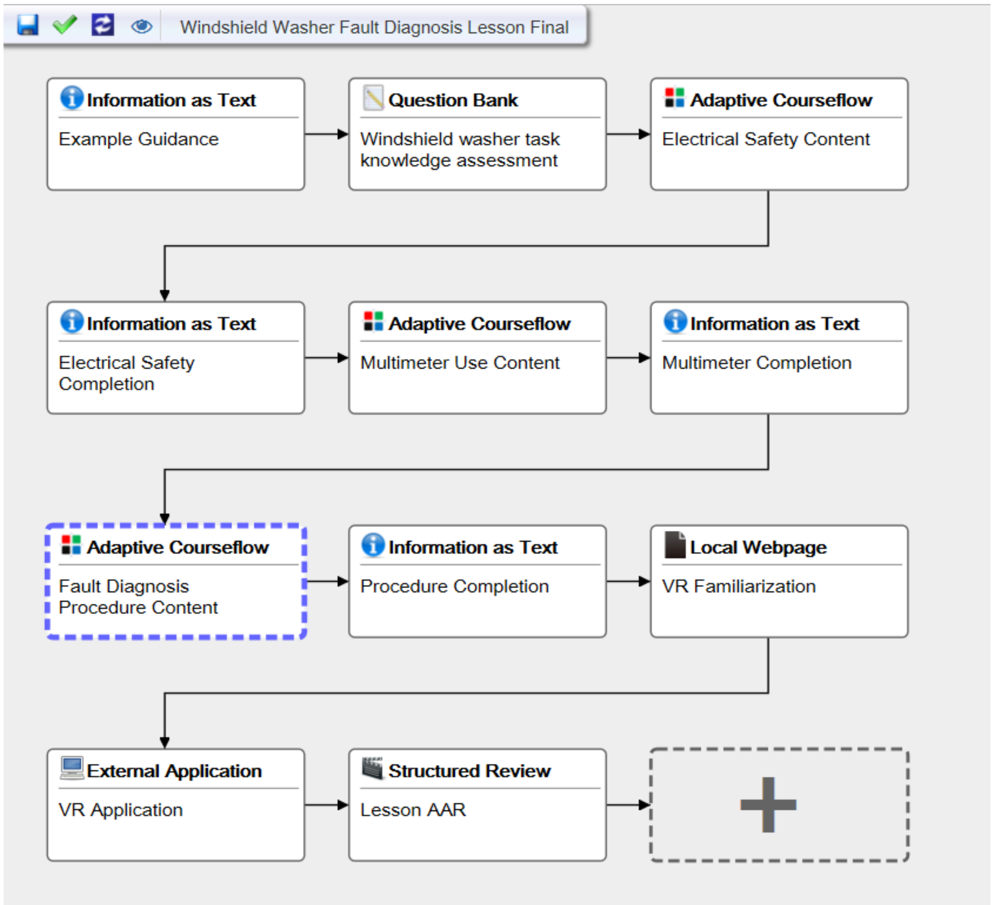


Fig. 35. Lesson flow for the integrated adaptive prototype.

The initial step in the adaptive learning lesson is knowledge assessment based on the course concepts of electrical safety, multimeter use and fault diagnosis procedural knowledge. We employed the Question Bank knowledge assessment functionality within GIFT to assess student understanding on those concepts and to characterize them as a novice, journeyman or expert on each of the three concepts. GIFT then uses those characterizations to sequence course content to the student and to adapt course content based on ongoing student parameter characterization as they move through the lesson content. Students are presented with content for the corresponding Adaptive Courseflow Modules (as described above) for each course concept based on their assessed knowledge level. Basic concept rules and examples content was delivered via PowerPoint presentations within the GIFT Adaptive Courseflow Modules. Knowledge checks were presented in GIFT using a subset of the initial Question Bank questions. If the student was deemed proficient, then the Boeing ITS capability provided the practice lesson content for selected learning concepts, launched from within the Adaptive Courseflow Module. While progressing through the practice module for each leaning concept, the ITS adapts within-lesson content to maximize a student's ability to successfully pass the practice portion of the lesson module on the initial attempt. This adaptation included within-lesson remediation on basic concepts if needed. This step is in addition to the normal GIFT content sequencing. GIFT functionality sequences the student through the rules, examples, knowledge check and practice components of each course concept's Adaptive Courseflow Module, and when all are successfully completed, launches the final practice module.

The final evaluated practice module is an external application using Boeing's virtual maintenance training capability (Jacquin, 2016). As part of the final practice assessment, students don a virtual reality (VR) headset, and using two 3D VR hand controllers, they are able to navigate to various places on the aircraft, perform the required troubleshooting tasks while adhering to required safety protocols, diagnose the fault and replace the faulty part (see Figure 3). Automated real-time performance assessment and adaptive learning capabilities within the virtual maintenance training system score the student on targeted learning objectives, provide on-demand student assistance to help locate components, and provide scoring to determine whether the student passes or fails the practical assessment. These final scores on the practical assessment are passed back to GIFT in order to evaluate whether or not the student successfully completed the entire lesson.

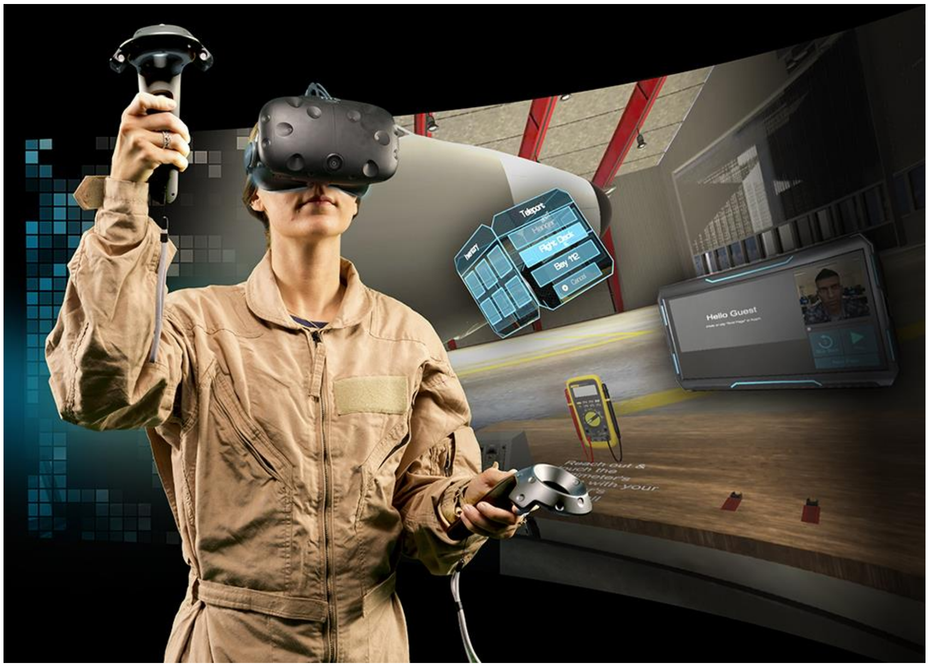


Fig. 3. Maintenance trainee performing a task in the virtual maintenance trainer.

At present, the first iteration of the integrated prototype is complete. Current efforts are focused on development of a test plan for the conduct of an adaptive training effectiveness study. Once the design is complete, any required modifications will be made to the adaptive training prototype in support of the effectiveness study.

Initial Test/Study Plan

Plans are in work to evaluate the effectiveness of the Boeing/GIFT prototype using cadets at West Point. The goal is to assess various manipulations of overall curriculum adaptation in an effort to determine which elements are best utilized to optimize student performance. To determine these efficiencies, we are using multiple measures, including: time to competence as measured by performance outcomes, training transfer and knowledge retention.

The plan is to evaluate the combined GIFT/Boeing prototype across a counter-balanced 3x2 experimental design (see Figure 4). The first independent variable is ITS Methods, with three defined conditions: (1) GIFT alone with personalization through the EMAP, (2) Boeing alone, with focused ITS interactions, and (3) GIFT/Boeing prototype that leverages both pedagogical methods. The second independent variable is prior-knowledge, with classification determined by outcomes on pre-test measures. Prior-knowledge will be scored on a concept by concept basis, with GIFT bypassing content on training materials a participant is showing mastery in. One potential option is to randomly assign students to one of those three groups, and then to divide them into high/low competency based on their initial knowledge assessments. Competency is only one of the potential personalization variables that we could use, as GIFT's pedagogical configuration can support strategy determinations across any individual difference deemed worthy to inform personalization. Our initial prototype did not include personalized measures of motivation-based adaptation or personalized feedback based on individual performance. Those are other options we are considering implementing once the study design is finalized.

		Prior-Knowledge	
		High	Low
ITS Methods	GIFT Alone	X	X
	Boeing Alone	X	X
	GIFT/Boeing	X	X

Fig. 4. Preliminary Effectiveness Study Experimental Design.

The outcomes of this study will inform modifications to both the GIFT architecture and Boeing adaptive training approach. These recommended changes will be based on the results of comparative evaluations across performance measures, along with observations and log-data associated with student interactions and behaviors across all content, assessments and scenarios. New requirements will be defined to better meet the needs of students, with the final year of the CRADA dedicated to instituting those changes.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Throughout the process of blending of two adaptive training solutions into one aggregate prototype, we have learned a number of lessons. There are similarities in the two approaches, as both concepts emphasize development of expertise based on optimizing the learning experience by adapting to the student. While both rely on performance assessment to adapt the lesson, implementations of how each used performance measurement to adapt was different. This led to a number of challenges when merging the two approaches into the combined prototype. On the positive side, we were able to successfully merge these capabilities into a lesson that was seamless from the perspective of the student. We relied on GIFT to perform the initial knowledge assessment, and to determine a starting point within the lesson. For simplicity sake in the initial prototype, we did not attempt to integrate individualized student traits such as motivation or engagement into our pedagogical decisions. We employed the GIFT adaptive modules to sequence through student need-based training, but then employed Boeing's ITS lessons to provide within-module assessments and practice, enabling remediation and practice at a more finite level than that provided by GIFT alone. We also demonstrated the ability to launch an external Unity-based VR practice module from GIFT, and showed that performance within that practice environment could be reported back to GIFT upon completion of the practice exercise.

Along with the positive points, we did identify a number of challenges throughout the course of our development. What follows is a summary of the lessons we have learned along with way.

- With any approach to adaptivity, there are challenges in the implementation of these concepts within a complex task environment. When combining two methods into a single learning solution, there are some additional complexities. For example, the Boeing approach to student assessment and adaptivity was different that the GIFT implementation. In some instances, both adaptation rules could run in parallel, but in other instances, we had to reconcile the different approaches.

- Long-term student record persistence is currently not implemented within GIFT. It would be nice to have the ability to customize lesson content based on a previous lesson learning record, but as of now, all lessons are stand-alone.

- Within a single lesson, we did not have the ability to remediate back to a previous adaptive learning module once it was determined to be mastered by the student. The implementation of our lesson involved completing individual modules and then completing an integrated external simulation exercise which combined aspects of multiple learning concepts. It would be nice to have the ability to remediate the student back to the individual adaptive module if they failed a concept during the final practice. Or, as mentioned previously, to record that failure as a persistent record and then be able to re-launch the lesson and repeat those modules where the student struggled during the final assessment.

- There was no GIFT standard for communicating with external applications. Interfacing external applications with GIFT required the creation of custom gateway modules which involved the implementation of message passing and parsing. Certain naming and scoring conventions between the external application and GIFT domain knowledge files were not intuitive. There was a lot of trial and error to make the process flow as desired. As GIFT becomes more pervasive and others attempt to interface with their existing applications, it would be beneficial to have a more standard approach to communicating with external applications.

- We had a number of usability issues as we initially began to author in GIFT. Some of those were due to bugs in the tool, while others were attributed to complexities in working with external applications. Specifically, we had issues running GIFT behind a proxy. In order to run the authoring tool behind the Boeing firewall, we had to disconnect from the internet and run it in offline mode. We also had difficulties due to size limitations in importing and exporting large lesson files. Users of GIFT could benefit from improved documentation or lessons on how to author.

- Limitations of older technologies within The Boeing-developed applications (e.g. Flash-based lesson playback) made for some complexities in how those lessons were integrated and displayed within GIFT.

We have learned through experience that there are strengths and weaknesses of different approaches to modeling students, providing feedback, and adapting content. As we continue to develop and test the overall effectiveness of adaptive learning in the coming year, we hope to capitalize on the best of each approach in creating a mutually beneficial joint solution.

REFERENCES

Anderson, J.R., Corbett, A.T., Koedinger, K.R., and Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences. 4(2)*: 167-207.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* pp. 4–16.

Cohen, P.A., Kulik, J.A. and Kulik, C-L. (1982). Educational Outcomes of Tutoring: A Meta-analysis of Findings. *American Educational Research Journal*. 19:237.

Graesser, A. C.; Person, N. K. (1994). Question Asking During Tutoring. *American Educational Research Journal, 31*, 104–137.

Hall, E.M., Gott, S.P., & Pokorny, R.A. (1995). A procedural guide to cognitive task analysis: The PARI methodology. *Technical Report No. AL/HR-TR-1955-0108*. Brooks AFB, TX: AFMC.

Hunt, E. and Minstrell, J. (1994). A collaborative classroom for teaching conceptual physics. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice. Cambridge: MIT Press*

Jacquin, G. (2016). V-22 Maintenance System. St. Louis, MO: *Boeing Technical White Paper*.

Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1988). SHERLOCK: A coached practice environment for an electronics troubleshooting job. *Technical Report: University of Pittsburgh, Learning Research and Development Center.*

Perrin, B. (2009). Intelligent Tutoring Systems: Facilitating Learning While Holding to Standard Practice. *Paper presented at the International Training and Education Conference,* Brussels, Belgium.

Perrin, B., Buck, B., Dargue, B., Biddle, E., Stull, T., & Armstrong, C. (2007). Automated Scenario-Based Training Management: Exploring the Possibilities. Orlando, FL: *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.*

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). *Technical Report No. DOI: 10.13140/RG.2.2.12941.54244.* Orlando, FL: US Army Research Laboratory.

ABOUT THE AUTHORS

***Dr. Barbara Buck** is a research psychologist in Innovation & Product Management, part of the Boeing Global Services' Training and Professional Services organization. Her recent efforts have focused on adaptive training, including integration adaptive training methods into lower-cost desktop solutions as well as more complex gaming and live simulation exercises. She has been instrumental in developing the Boeing adaptive authoring capability, and has conducted effectiveness research to evaluate the efficacy of adaptive training approaches, virtual reality applications and physiological measures of performance. Barbara holds a Ph.D. and Master's Degree in Cognitive and Engineering Psychology from the University of Illinois.*

***Matthew Genova** is a software engineer at Boeing whose current focus is developing virtual reality and augmented reality applications. He has developed adaptive learning software and was instrumental in creating the custom gateways enabling Boeing adaptive learning technology to interface with GIFT. He is part of the Innovation & Product Management team in the Boeing Global Services' Training and Professional Services organization. He holds a Master's Degree in Computer Engineering from Washington University in St. Louis.*

***Dr. Robert Sottilare** leads adaptive training research within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He is ARL's technical lead for the Center for Adaptive Instructional Sciences (CAIS).*

***Dr. Benjamin Goldberg** is an adaptive training scientist at the Army Research Laboratory's Human Research and Engineering Directorate. He leads research focused on instructional management within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).*

Design and Development of an Adaptive Hypermedia-Based Course for Counterinsurgency Training in GIFT: Opportunities and Lessons Learned

Jonathan Rowe[1], Randall Spain[1], Robert Pokorny[2], Bradford Mott[1], Benjamin Goldberg[3], and James Lester[1]
North Carolina State University[1], Intelligent Automation, Inc.[2], U.S. Army Research Laboratory[3]

INTRODUCTION

There is broad recognition that intelligent tutoring systems (ITSs) are effective for enhancing student learning across a range of domains (Anderson, Corbett, Koedinger, & Pelletier, 1995; VanLehn, 2011; Ma, Adescope, Nesbit, & Liu, 2014). By leveraging computational models of adaptive pedagogical decision-making, ITSs create personalized learning experiences that are dynamically tailored to individual students. However, ITSs are resource-intensive to create. The amount of engineering effort required to develop one hour of ITS instruction is often estimated to be approximately 200 hours (Aleven, McLaren, Sewall, & Koedinger, 2009). To address this bottleneck, there have been several initiatives to devise tools for supporting ITS authors in creating adaptive training at reduced time and cost (Aleven et al., 2009; Sottilare, Baker, Graesser, & Lester, in press). These efforts hold the promise of making ITSs available across a broader range of subjects and contexts, enhancing the depth of current adaptive learning experiences, and enabling instructional designers and subject matter experts to create novel adaptive training solutions without requiring programming expertise.

Over the past several years, the Generalized Intelligent Framework for Tutoring (GIFT) has emerged as an important initiative to address the authoring challenges raised by ITSs. GIFT is an open source service-oriented framework of software tools, methods, and de-facto best practices for designing, developing, and evaluating adaptive training systems. GIFT provides instructors with a suite of web-based tools for rapidly creating intelligent tutors, and it is linked to several ongoing research efforts to devise methods for automating key elements of the adaptive training authoring process (Rowe et al., 2016). Many of these tools are available through GIFT's Course Creator, which provides a drag-and-drop interface for devising adaptive training experiences across a range of domains. The Course Creator is also continuously improving with new capabilities and usability enhancements released several times a year. As GIFT transitions from the research lab to real-world use, these tools will be subject to new authorial demands and scalability challenges, which makes it a valuable test case for understanding how ITS technologies mature and scale.

In this paper, we describe our experiences and lessons learned from using GIFT to create an approximately 2-hour adaptive hypermedia-based training course for counterinsurgency (COIN) and stability operations. The course builds upon the UrbanSim Primer, which presents a range of multimedia training materials providing direct instruction on doctrinal concepts of COIN that accompanies the UrbanSim simulation-based training environment. The course serves as a showcase of recent enhancements to GIFT's Engine for Management of Adaptive Pedagogy (EMAP) that support adaptive assessment and remediation. Specifically, remediation features in GIFT are based on Chi's ICAP framework (2009). ICAP describes several modes of student engagement with learning materials, including passive, active, constructive, and interactive modes. The ICAP framework predicts that the interactive mode (e.g., peer dialogue) is more effective for learning than the constructive mode (e.g., writing an explanation), the constructive mode is more effective than the active mode (e.g., reading and highlighting a passage), and the active mode is more effective than the passive mode (e.g., reading a passage without doing anything else). But, there are tradeoffs between these pedagogical strategies, such as instructional time required and cognitive load imposed. We are utilizing the COIN hypermedia-based training course to gather data on student responses to passive, active, and constructive remediation activities, which is part of a broader research program on utilizing reinforcement learning to automatically induce intelligent tutoring policies for instructional remediation in GIFT.

Our course is notable in its scope: it utilizes nearly 40 adaptive course flow objects, more than 150 media objects, online videos, pre-post surveys, embedded assessments, adaptive feedback messages, glossaries, and other features of GIFT. Further, we are preparing the course for deployment to hundreds of users through a crowdsourcing study with Amazon Mechanical Turk, which requires preparation for remote deployment to dozens of concurrent users in a fashion that integrates seamlessly with tools and workflows from commercial crowdsourcing providers. We describe how this course was created with the GIFT Course Creator and associated ICAP-inspired functionalities; we describe methods for implementing key ITS features such as immediate feedback and scaffolding in hypermedia-based training with GIFT; and we describe challenges, solutions, and opportunities we have encountered from our experiences creating the course. Our findings point toward future directions for enhancing GIFT's capacity to reduce the authorial cost of creating ITSs and transitioning toward wider scale use.

RESEARCH CONTEXT

Tutorial planning, a critical component of adaptive training, controls how scaffolding and instructional interventions are structured and delivered to learners. Devising computational models that scaffold effectively, i.e., determining when to scaffold, what type of scaffolding to deliver, and how scaffolding should be realized, is a critical challenge for the field of ITSs. Recent years have seen growing interest in data-driven approaches to tutorial planning (Rowe & Lester, 2015; Williams et al., 2016; Zhou, Wang, Lynch, & Chi, 2017). In particular, reinforcement learning techniques have shown promise for automatically inducing tutorial policies that optimize student learning outcomes and do not require pedagogical policies to be manually programmed or demonstrated by expert tutors. These techniques are complementary to advances in ITS authoring, including authoring tools implemented in GIFT, to address challenges inherent in constructing adaptive training materials.

Reinforcement learning is a category of machine learning that centers on devising software agents that perform actions in a stochastic environment to optimize some concept of numerical reward (Sutton & Barto, 1998). In reinforcement learning, the agent induces a control policy by iteratively performing actions and observing their effects on the environment and accumulated rewards. Tutorial planning can be formalized as a reinforcement learning task by conceptualizing the tutor as the agent: the tutor seeks to enact pedagogical decisions (i.e., actions) that will affect its environment (i.e., the trainee and his/her learning environment) in order to optimize student learning outcomes (i.e., rewards). In our case, the pedagogical decisions are choosing between ICAP-inspired remediation activities, and the tutorial planner’s objective is to optimize student learning in an adaptive hypermedia-based training course for COIN.

To investigate a reinforcement learning framework for ICAP-inspired remediation in GIFT, we plan to obtain a large dataset consisting of trainee responses to different types of instructional remediation activities as well as pre-post learning outcomes. The purpose of the dataset is to serve as a training corpus for inducing and evaluating reinforcement learning policies for tutorial planning (Rowe & Lester; Wang et al., 2017). Reinforcement learning techniques are data-intensive, so in order to collect sufficient data, we have devised a training course that is designed to meet three objectives: (1) the course contains numerous opportunities for learners to receive instructional remediation, which will serve as the training data for reinforcement learning; (2) the course is deployable through online crowdsourcing platforms, which will facilitate broad distribution to many learners efficiently; and (3) the course enacts an exploratory (i.e., random) remediation policy in order to broadly sample the space of possible pedagogical decisions. To meet these objectives, we developed an adaptive hypermedia-based training course in GIFT that builds upon materials from the UrbanSim Primer.

UrbanSim Primer

The UrbanSim Primer is a hypermedia-based learning environment that was developed by the USC Institute for Creative Technologies to provide direct instruction on COIN doctrine and principles. Major topics include the importance of population support, processes for intelligence gathering, and issues in successful COIN operations. The UrbanSim Primer's training materials are divided across seven lessons that interleave hyperlinked video, audio, text, and diagrams delivered using Adobe Flash.

In our project, we focus on a subset of training materials from UrbanSim Primer Lessons 1-4. We have extracted video, audio, and text content from the UrbanSim Primer, and we have reconfigured these materials for web-based presentation using GIFT. Specifically, GIFT enables the delivery of UrbanSim Primer materials via web browsers, it enables interleaved training materials that include embedded assessments and instructional remediation, and it supports automatic logging of learner actions within the training course.

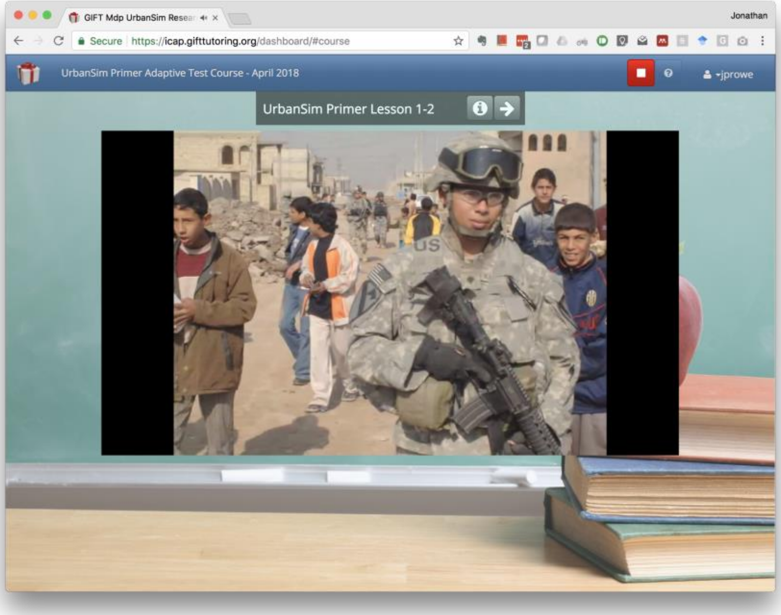

GIFT Mdp UrbanSim Resear
Jonathan
Secure
https://icap.gifttutoring.org/dashboard/#course
UrbanSim Primer Adaptive Test Course - April 2018
jprowe
UrbanSim Primer Lesson 1-2
US

Figure 1: Screenshot of UrbanSim Primer training video presented in GIFT.

An Implemented Adaptive Hypermedia-Based Training Course for COIN

We have designed an adaptive hypermedia-based training course based on the UrbanSim Primer using a branch of GIFT Cloud that supports recent enhancements to the GIFT EMAP to support ICAP-based instructional remediation functionalities. The course builds upon the doctrinal lessons presented in the UrbanSim Primer and includes a series of short videos, instructional texts, quiz questions, remedial content, and glossaries related to the fundamental principles of COIN and stability operations. Trainee experiences with the COIN training course proceed as follows.

The course begins with a general message that welcomes students to the training course. Following this introduction, participants complete a demographic questionnaire that asks them about their age, years of education, and familiarity with COIN topics and concepts, followed by a goal orientation questionnaire that measures students' task-based and intrinsic motivation to learn (Elliot & Murayama, 2008). Next, students complete a 12-item pretest that measures prior knowledge of COIN principles and doctrine.

After completing the pre-training surveys, participants begin the adaptive hypermedia portion of the course, which is organized into four chapters: (1) Introduction to COIN Operations; (2) Planning COIN Operations; (3) COIN Analysis Tools; and (4) COIN and Human Intelligence. Each chapter contains a series of narrated videos and text-based content that cover lesson topics such as "Identifying the center of gravity in COIN operations", "Defining intelligence preparation for the battlefield", and "Understanding lines of effort in COIN operations." Each lesson is implemented as a series of adaptive course flow objects within the GIFT course.

After each video from the UrbanSim Primer, participants complete a brief multiple-choice quiz. Quiz questions consist of single concept and multi-concept review items that align with the course's learning objectives. Single concept review questions require learners to recall and apply concepts presented within the lesson. Multi-concept review questions require learners to demonstrate a deeper understanding of course material by integrating concepts from multiple lessons. The course uses a micro-sequencing adaptive training approach (Durlach & Spain, 2014) to "gate" progress according to learners' demonstrated level of mastery. Learners who correctly answer a quiz question are allowed to advance to the next question or lesson, whereas learners who incorrectly answer a question receive ICAP-inspired supplemental remediation.

When a learner receives supplemental remediation following a missed question, GIFT prompts the learner to either: (1) *passively* re-read the narrated content that was just presented in the lesson video; (2) re-read the video content and *actively* highlight the portion of text that is most relevant to the quiz question that was missed; or (3) re-read the text and *constructively* summarize content related to the quiz question. The active and remediation prompts also include expert highlighting/summaries that students can use to self-evaluate the accuracy of their responses. The course also includes a "no remediation" prompt that only provides students with minimal feedback before being asked to re-answer the quiz question. The course uses a random assignment policy at the item level to determine whether students receive passive, active, constructive, or no remediation after each incorrect item response. Students continue to receive supplemental remediation until they demonstrate concept mastery (i.e., correctly answer the quiz question).

In addition to the ICAP-inspired remediation prompts, the training course also monitors how long students engage with the video-based lessons and provides prompts to those participants who advance through the videos too quickly or too slowly. For example, participants who click past a video before it ends receive the following message, "It appears that you clicked past the video before enough time elapsed for it to play in entirety. Please do not rush through the training materials, or else you may not achieve the course learning objectives." Conversely, participants who spend too much time dwelling on the video (defined as more than 5 minutes on a video page) will receive the following message, "It appears that you spent an unusually long amount of time on this video. Please attempt to complete the training materials at an efficient pace." The maximum video length is approximately 1.5 minutes.

Upon finishing the final lesson, participants complete a series of post-training surveys. These include a multiple-choice posttest to measure retention of foundational COIN concepts and a short questionnaire to collect opinions about the training experience. After completing these activities, participants receive a debriefing message and are thanked for their participation. In addition, participants who access the course through an online crowdsourcing platform (e.g., Mechanical Turk) receive a unique code that they can provide to the crowdsourcing vendor to receive payment for participation. The code is randomly generated by a customized survey implemented as the final course object in the training course.

In order to collect data on learner interactions with ICAP-inspired remediation activities, we plan to conduct a human subjects study with a sample of 300-500 participants recruited through Amazon Mechanical Turk. A short description of the study will be posted on the Mechanical Turk website. Individuals interested in completing the training course will first complete an electronic informed consent before being hyperlinked to the published training course hosted on the cloud-based version of GIFT. Once in the course, participants will proceed through the course activities described above. At the end of the training course, participants will receive a unique 7-digit code that they must enter into the Mechanical Turk site to receive payment for their participation. Using the data gathered from the Mechanical Turk study, we will begin to investigate data-driven models of tutorial planning using reinforcement learning techniques.

DESIGNING AN ADAPTIVE HYPERMEDIA-BASED TRAINING COURSE IN GIFT

To develop an adaptive hypermedia-based training course that can be delivered through the web, we made extensive use of the GIFT Course Creator. The GIFT Course Creator is a web-based GUI authoring tool that enables instructional developers to construct training workflows that encode sequences of online learning activities using a drag-and-drop interface. The Course Creator enables instructional developers to specify *fixed course flows*, which are course-object sequences that unfold the same way for every learner, as well as *adaptive course flows*, which utilize the GIFT EMAP to drive macro-adaptive pedagogical decisions about content sequencing based on student performance. A key component of our work on the adaptive hypermedia-based COIN training course is utilization of an enhanced version of EMAP that supports ICAP-inspired remediation functionalities. Specifically, the course includes 39 adaptive course flow objects, each linked to a range of supporting media files including videos, text passages, feedback statements, ICAP-inspired remediation prompts, and quiz questions that align with course concepts. In this section, we briefly describe how these adaptive courseflow objects are configured to provide direct instruction, embedded assessment, immediate feedback, and adaptive remediation on COIN concepts.

In GIFT, adaptive courseflow objects are deeply grounded in Component Display Theory (Merrill, Reiser, Ranney & Trafton, 1992). Component Display Theory describes a process for learning the rules of a domain, examining relevant examples, testing recall of knowledge, and engaging in guided practice. These four types of learning activities delineate quadrants in an adaptive courseflow object: Rules, Examples, Recall, and Practice. During a typical interaction with an adaptive courseflow object, the learner begins by viewing multimedia training materials associated with a set of target domain concepts; this is the learner's experience of the Rules Quadrant. After viewing these materials, the learner transitions to the Examples Quadrant in which she views additional training materials that illustrate examples of the target concepts. Afterward, the learner transitions to the Recall Quadrant, where her understanding of the target concepts is assessed through a series of quiz questions. After successfully completing the quiz, the learner optionally transitions to the Practice Quadrant, where she interacts with an external training simulation to apply her relevant knowledge in a hands-on manner. In our course, we do not currently make use of the Practice Quadrant.

In the ICAP-enhanced version of the GIFT EMAP, the four quadrants are augmented with an additional fifth quadrant: Remediation. The Remediation Quadrant houses logic and training materials for presenting instructional feedback and ICAP-inspired remediation to learners with below-threshold performance in the Recall Quadrant. In other words, when learners miss too many embedded quiz questions, they receive immediate feedback and remediation. The Remediation Quadrant is populated with multimedia training materials that are distinct from those presented in the Rules and Example Quadrants. Remediation materials can be conceptualized in terms of three categories: (1) Constructive-response remediation, (2) Active-response remediation, and (3) Passive-response remediation.[1] Constructive- and active-response remediation materials are created using built-in GIFT authoring templates, whereas passive remediation materials can be constructed with a range on supported media types, including videos, text passages, web pages, and slide decks. In the case of our course, all remediation materials are text based, and we specifically utilize text-based local web pages to implement Passive-response remediation. At present, the presentation of these three different types of remediation is performed according to a uniform random policy. This control policy will be replaced with an adaptive policy induced using reinforcement learning following the completion of the Mechanical Turk study, subsequent data analysis and model creation.

In our course, we utilize adaptive courseflow objects to provide immediate feedback and remediation after each quiz question. We devise a unique adaptive courseflow object for each embedded quiz item in the course. Each adaptive courseflow object contains a Recall Quadrant with a single question, as well as a Remediation Quadrant with four associated media files: a passive-response remediation intervention, an active-response remediation intervention, a constructive-response remediation intervention, and a non-remediation intervention. Each of these remediation media files contains a feedback statement about the quiz question that the learner must have missed prior to receiving the remediation. Because our course presents multiple embedded quiz questions after each video from the UrbanSim Primer, a subset of adaptive courseflow objects contain links to YouTube videos in their Example Quadrants. However, not all adaptive courseflow objects contain these videos, or contain Example Quadrant media files at all. In these cases, we repurpose the Example Quadrant to present a local webpage containing positive feedback about the quiz question that the student just answered; when a learner transitions to a new adaptive courseflow object, she must have just answered a quiz question correctly, so we provide positive feedback. The Rule Quadrant of each adaptive courseflow object is generally not used in our course, except to present transition text in a handful of locations. For an illustration of all of the materials and media files associated with each adaptive courseflow object in our course, please see Figure 2.

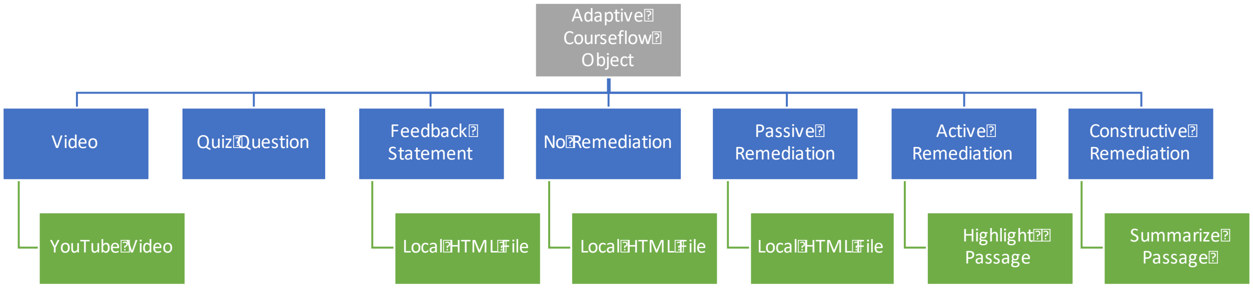


Figure 2. Overview of training materials associated with a single adaptive courseflow object.

[1] The fourth category of ICAP, interactive-response remediation, is not currently supported by the GIFT EMAP.

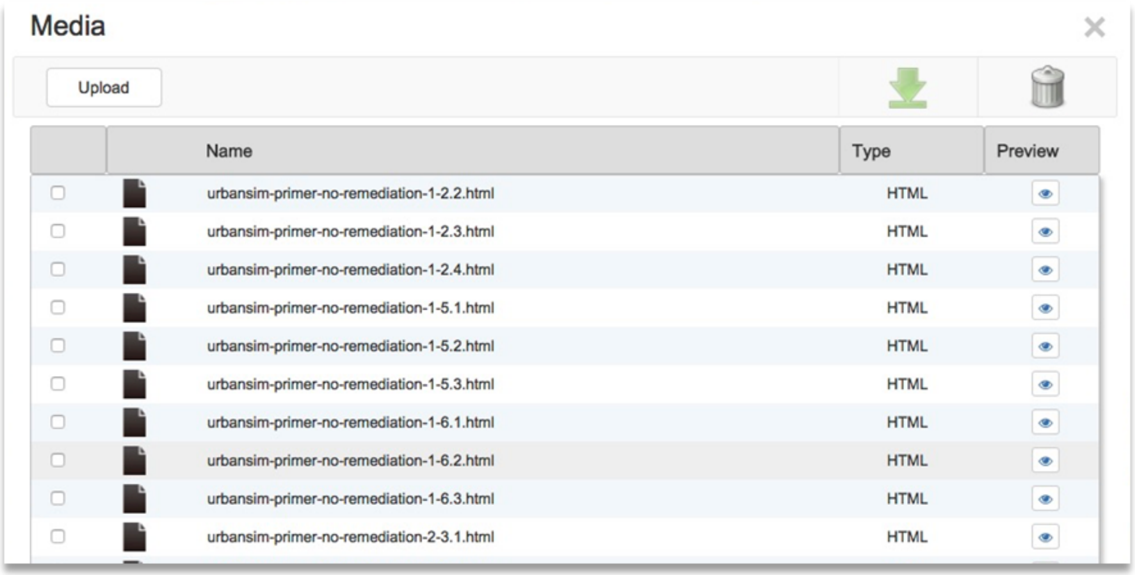
BEST PRACTICES AND LESSONS LEARNED

As noted above, the course is relatively large in comparison to many GIFT courses that have been created to date. The course includes approximately 40 adaptive course flow objects, more than 150 media objects, several pre-post surveys, numerous embedded assessments, adaptive feedback messages, glossaries, and other features of GIFT.

As we developed the course, we found that designing a large training course in GIFT required significant preparation and planning. A key practice that we utilized was to develop a course prototype outside of GIFT prior to constructing the training course inside of GIFT. In our case, we developed a rough course prototype in PowerPoint, which served two functions: (1) It provided the team with an easily editable instructional design map of the training course, including an overview of the course flow for each chapter and subsequent lessons, and (2) It allowed the team to quickly edit and refine the training content, embedded assessments, and remediation content before implementing the full user-ready version in GIFT. We also found that the prototype served as a useful reference for authoring remediation prompts. For each lesson, we created a series of slides that showed the quiz question that aligned with the lesson, transcripts of the narrated text from the video, text for the passive remediation prompts, text and suggested highlighting for the active remediation prompts, and content for the constructive remediation prompts. Organizing all of this information in a format that could be rapidly generated, easily edited, shared between collaborators, and which did not require perfect precision in specifying courseflows played a vital role in the early stages of authoring the adaptive training course.

A second lesson we learned was that during course development and revision, there were several occasions in which we needed to revise course content (e.g., quiz questions and prompts) to improve the clarity of the training materials. These changes were based on upon user feedback from pilot testing of the course. We found it helpful to keep track of these revisions in the PowerPoint prototype of the course, which allowed the project team to easily track changes made to the course over the development cycles.

A third lesson we learned was that it is important to implement a naming and organization convention for the media files used in the course. As a best practice, we used an object + lesson naming scheme (e.g., Remediation 2-3 Constructive; Remediation 2-3 Active, etc.) to provide structure and consistency among all of our training assets. This organizational scheme was particularly useful in managing the feedback statements, passive remediation files, and no-remediation files associated with each adaptive courseflow object. This allowed us to quickly review which objects were included in each course object's Remediation Quadrant. It also helped us manage the large number of training assets saved in the course's media folder. As previously noted, our course includes over 150 content files. During the authoring process, there were many occasions in which we needed to either preview or edit passive and/or non-remediation files associated with the course. In the current implementation of GIFT, the only way to preview and update these files is by accessing the file through the media content organizer, which lists all of a courses' media objects (see Figure 3). Using a pre-established naming convention allowed us to quickly locate and replace old course objects when we needed to make changes to the training course, which occurred several times during the iterative course development and refinement process.



	Name	Type	Preview
□	urbansim-primer-no-remediation-1-2.2.html	HTML	👁
□	urbansim-primer-no-remediation-1-2.3.html	HTML	👁
□	urbansim-primer-no-remediation-1-2.4.html	HTML	👁
□	urbansim-primer-no-remediation-1-5.1.html	HTML	👁
□	urbansim-primer-no-remediation-1-5.2.html	HTML	👁
□	urbansim-primer-no-remediation-1-5.3.html	HTML	👁
□	urbansim-primer-no-remediation-1-6.1.html	HTML	👁
□	urbansim-primer-no-remediation-1-6.2.html	HTML	👁
□	urbansim-primer-no-remediation-1-6.3.html	HTML	👁
□	urbansim-primer-no-remediation-2-3.1.html	HTML	👁

Figure 3. Training Assets in the media folder of GIFT's Course Creator.

A fourth lesson is the importance of developing a large hypermedia course such as this one in an iterative fashion. As a best practice, we developed the course one chapter at a time and conducted internal pilot testing between development cycles to ensure the course workflow and remediation materials were being implemented properly. During our pilot testing sessions, we examined extreme, correct, and incorrect responses to the quiz questions to ensure the course logic was correct, and we examined whether the remediation prompts were being executed correctly in order to tune course parameters and functionality. In addition to reviewing the behavior of these system level features, we also used testing as an opportunity to make any changes to the visual design of the course, such as making changes to font sizes and line spacing in our remediation prompts and messages prior to developing the rest of the course's media objects; a change in the presentation style of one feedback message could potentially propagate to more than a hundred additional files if an author is not careful about phased development.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Adaptive training systems show considerable promise for enhancing student learning across a range of domains. Recent advances in ITS authoring tools, as well as data-driven tutorial planning, are showing significant progress toward reducing the effort required to create personalized learning experiences. Reinforcement learning is a natural formalism for automatically inducing tutorial planning models to drive pedagogical decisions about instructional feedback and remediation. In order to utilize reinforcement learning techniques for data-driven tutorial planning, we have constructed an adaptive hypermedia-based training course in GIFT that is based on the UrbanSim Primer to teach foundational principles and doctrine on COIN operations. We utilize ICAP-inspired enhancements to GIFT's EMAP to provide immediate feedback and remediation during the adaptive training course. Based on our experience creating the course, we have identified several best practices and lessons learned for adaptive course creation in GIFT. These include the importance of external prototyping, carefully tracking course revisions, devising consistent file-naming schemes, and emphasizing iterative design and development throughout course creation.

As a next step, we will deploy the adaptive training course in a human subject study using the Amazon Mechanical Turk crowdsourcing platform in order to collect a training corpus for investigating reinforcement learning-based tutorial planning. Following the study, we will utilize the dataset to induce control policies for adaptively personalizing remediation and feedback decisions to individual learners. In the future, we plan for these models to be incorporated back into the run-time adaptive training course and evaluated with a new cohort of learners in order to evaluate the effectiveness of reinforcement learning techniques for data-driven tutorial planning in GIFT.

There are several promising avenues for future enhancements to GIFT. One recommendation is to include advanced previewing capabilities within the GIFT Course Creator. In particular, adding features that allow authors to preview adaptive course flow objects, and in particular, Remediation Quadrant materials, would be highly valuable. Currently, course authors can access and edit the content of the constructive and active remediation prompts, but they cannot preview how these prompts appear at run-time when they are presented by GIFT. The same previewing functionality would be useful for passive remediation content as well, such as local web pages, particularly if they could be previewed directly from adaptive courseflow objects in the Course Creator.

Enhancements related to viewing and managing large numbers of media files would also be helpful to course creators. Including a feature that allows course authors to quickly view all of the media file labels attached to an adaptive courseflow object would significantly facilitate authoring for large courses. Currently, authors have to open each adaptive courseflow object and individually click on each quadrant to see which media files are linked to each quadrant. Including a feature that could quickly export or summarize this information at a high level would eliminate this process and would be a valuable tool for evaluating and refining the training course.

ACKNOWLEDGMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory under cooperative agreement W911NF-15-2-0030. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167-207.

Aleven, V., Mclaren, B.M., Sewall, J., & Koedinger, K.R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education*, 19(2), 105-154.

Chi, M.T.H. (2009). Active-Constructive-Interactive: A conceptual framework of differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73-105.

Durlach, P. J., & Spain R. D. (2013). Framework for instructional technology: Methods of implementing *adaptive training and education*. (Technical Report 1335). Arlington VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613.

Ma, W., Adesope, O.O., Nesbit, J.C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901.

Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *Journal of the Learning Sciences*, 2(3), 277-305.

Rowe, J., Frankosky, M., Mott, B., Lester, J., Pokorny, B., Peng, W., & Goldberg, B. (2016). Extending GIFT with a Reinforcement Learning-Based Framework for Generalized Tutorial Planning. *Proceedings of the Fourth Annual GIFT User Symposium* (GIFTSym4), pp. 87-97.

Rowe, J. P., & Lester, J. C. (2015). Improving Student Problem Solving in Narrative-Centered Learning Environments: A Modular Reinforcement Learning Framework. *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (pp. 419-428), Madrid, Spain.

Sottilare, R.A., Baker, R., Graesser, A.C. & Lester, J. (In press). Special Issue on the Generalized Intelligent Framework for Tutoring (GIFT): Creating a Stable and Flexible Platform for Innovations in AIED Research. *International Journal of Artificial Intelligence in Education*.

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

Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., & Heffernan, N. (2016). Axis: Generating Explanations at Scale with Learnersourcing and Machine Learning. *Proceedings of the 3rd ACM Conference on Learning@ Scale* (pp. 379-388), Edinburgh, UK.

Zhou, G., Wang, J., Lynch, C. & Chi, M. (2017) Towards Closing the Loop: Bridging Machine-induced Pedagogical Policies to Learning Theories. *Proceedings of the 10th International Conference on Educational Data Mining*, (pp.112-119), Wuhan, China.

ABOUT THE AUTHORS

***Dr. Jonathan Rowe** is a Research Scientist in the Center for Educational Informatics at North Carolina State University. He received the Ph.D. and M.S. degrees in Computer Science from North Carolina State University, and the B.S. degree in Computer Science from Lafayette College. His research is in the areas of artificial intelligence and human-computer interaction for advanced learning technologies, with an emphasis on game-based learning environments, intelligent tutoring systems, user modeling, educational data mining, and computational models of interactive narrative generation.*

***Dr. Randall Spain** is a Research Psychologist in the Center for Educational Informatics at North Carolina State University. He earned his Ph.D. in Human Factors Psychology from Old Dominion University in 2009 and has been actively conducting training and human factors research for the Department of Defense and the Department of Homeland Security for the past 10 years. His research uses principles, theories, and methods of applied psychology (human factors, educational psychology, personnel psychology, experimental psychology, and psychometrics) to evaluate the impact of advanced training technologies on learning and performance.*

***Dr. Robert Pokorny** is Principal of the Education and Training Technologies Division at Intelligent Automation, Inc. He earned his Ph.D. in Experimental Psychology at University of Oregon in 1985, and completed a postdoctoral appointment at University of Texas at Austin in Artificial Intelligence. Bob's first position after completing graduate school was at the Air Force Research Laboratory, where he developed methodologies to efficiently create intelligent tutoring systems for a wide variety of Air Force jobs. At Intelligent Automation, Bob has led many cognitive science projects, including adaptive visualization training for equipment maintainers, and an expert system approach for scoring trainee performance in complex simulations.*

***Dr. Bradford Mott** is a Senior Research Scientist in the Center for Educational Informatics at North Carolina State University. Prior to joining North Carolina State University, he served as Technical Director at Emergent Game Technologies where he created cross-platform middleware solutions for video game development, including solutions for the PlayStation 3, Wii, and Xbox 360. Dr. Mott received his Ph.D. in Computer Science from North Carolina State University in 2006, where his research focused on intelligent game-based learning environments. His* *current research interests include computer games, computational models of interactive narrative, and intelligent game-based learning environments.*

Dr. Benjamin Goldberg *is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL) Human Research and Engineering Directorate (HRED), Simulation and Training Technology Center (STTC) in Orlando, FL. He has been conducting research in modeling and simulation for the past five years with a focus on adaptive learning and how to leverage artificial intelligence tools and methods for adaptive computer-based instruction. Currently, he is the LITE Lab's lead scientist on instructional strategy research within adaptive training environments. Dr. Goldberg is a Ph.D. graduate from the University of Central Florida in the program of Modeling & Simulation.*

Dr. James Lester *is Distinguished Professor of Computer Science at North Carolina State University, where he is Director of the Center for Educational Informatics. His research centers on transforming education with technology-rich learning environments. With a focus on adaptive learning technologies, his research spans intelligent tutoring systems, game-based learning environments, affective computing, and tutorial dialogue. The adaptive learning environments he and his colleagues develop have been used by thousands of students in K-12 classrooms. He received his Ph.D. in Computer Science from the University of Texas at Austin in 1994. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI).*

Effects of feedback framing and regulatory focus are task-dependent

Ashley H. Oiknine[1,2], Kimberly A. Pollard[3], Peter Khooshabeh[2,3], Antony D. Passaro[3], Benjamin T. Files[3]
DCS Corporation[1], U.C. Santa Barbara[2], US Army Research Laboratory West[3]

INTRODUCTION

The Problem

Training paradigms often depend on performance feedback to enhance motivation, increase engagement, and improve performance. However, the effects of feedback on task performance are mixed (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). These mixed results may be explained by how individuals differ in their reactions to specific types of feedback, but this variability is often difficult to predict. Furthermore, task properties may influence feedback effectiveness. Feedback intervention theory, (FIT; Alder, 2007; Kluger & DeNisi, 1996) states that feedback interventions regulate behavior by changing the focus of attention to a particular discrepancy between performance and standards. Individual differences in goal orientations (i.e., trait regulatory focus) influence attentional focus, as well as intrinsic goals or standards, and therefore likely impact whether and to what extent feedback influences future performance. More study is needed investigating the effectiveness of feedback within the context of individual differences and their interactions with tasks and domains to inform learner models and better implement individually optimized instructional strategies.

Relevance to GIFT

The design of GIFT incorporates users' individual traits to deliver tailored training. One of GIFT's major design principles includes the delivery of individually tailored instructional interventions using empirically based generic instructional strategies (Wang-Costello, Goldberg, Tarr, Contron, & Jiang, 2013). GIFT contains mechanisms to select appropriate feedback for given training tasks. Further refining a model which incorporates task properties and individual responses to feedback would improve GIFT's ability to provide more tailored and effective training. What we present is (1) a particular trait to consider and (2) the implications that task properties may have in determining effective feedback.

In the present work, we looked at the interaction of task affordances and trait regulatory focus as possible predictors of feedback effectiveness to inform GIFT's existing models. Results can be incorporated into learner models but may also require domain module information for proper implementation.

Regulatory Focus and Regulatory Fit

Regulatory focus is a goal orientation construct (Higgins, 1998; Higgins et al., 2001) that contains two distinct motivational orientations that describe an individual's propensity to approach gains or avoid losses: promotion focus and prevention focus. Highly promotion-focused individuals have a tendency to pay more attention to opportunities for gain and are motivated by intrinsic ideals as compared to highly prevention-focused persons whose motivations are rooted in extrinsic obligation and avoidance of loss (Higgins, 1998; Van-Dijk & Kluger, 2004). These propensities may have implications for responses to strategic affordances in tasks, such as eagerness/approach strategies and vigilance/avoidance strategies

(Higgins et al., 2001). Promotion and prevention scores are largely independent of each other (Summerville & Roese, 2008), and can be obtained from questionnaires such as the Regulatory Focus Questionnaire (Higgins et al., 2001).

Regulatory fit theory (Higgins, 2000) predicts that when an individual's regulatory focus is matched with the nature of a goal, object, or reward structure framing (i.e. point-gains for promotion and point-losses for prevention), a more motivated and engaged state is elicited as compared to when they do not align. According to this theory, matching a highly promotion-focused individual with feedback framed in terms of gains should yield a more motivated and engaged learner as compared to a highly prevention-focused individual and vice versa. In addition, the nature of the task itself and its strategic affordances should also influence regulatory fit. To investigate whether regulatory fit theory may be useful to incorporate in learner models, we examined the effects of regulatory focus, feedback framing, and task affordances within the context of two inhibitory control go/no-go tasks that varied in the timing and number of trials.

Inhibitory Control

Inhibitory control involves the ability to override or halt an otherwise automated response, especially when that automated response is wrong or inappropriate. The ability to suppress inappropriate responses is essential for healthy living and functioning. Deficiencies in inhibitory control contribute to the risk of engaging in maladaptive behaviors such as alcohol abuse (Kamarajan et al., 2004), poor sleep hygiene (Todd & Mullan, 2014), drug use (Fillmore & Rush, 2002), and over-eating (Houben, 2011). Individuals with a deficiency in inhibitory control experience difficulties with decision making (Shenoy & Yu, 2011), executive function and working memory (Carlson, Moses, & Claxton, 2004). Some work has shown that inhibitory control can be improved with training (Berkman, Kahn, & Merchant, 2014). For example, one week of inhibitory control training significantly reduced civilian casualties in a simulated hostage situation (Biggs, Cain, & Mitroff, 2015). Inhibitory control can be trained using a go/no-go task, in which participants are asked to press a button in response to a "go" stimulus and withhold a response to a "no-go" stimulus. The simplicity of go/no-go paradigms makes them an excellent testbed for examining the effects of individual traits, feedback framing, and task strategic affordances. The flexibility of go/no-go paradigms allows the same basic task to be performed using different strategic affordances, which may be encouraged via subtle changes of stimulus timing.

Current Research

We tested the effectiveness of regulatory fit as a means of increasing performance on an inhibitory control training task in two experiments using different trial timelines. Based on previous literature that supports regulatory fit's ability to elicit a more motivated state, we predicted in both cases that the training would be more effective when the feedback framing matched the trainee's regulatory focus and the task's strategic affordances. Both experiments showed effects of regulatory focus, but the effects were different in the two experiments. In Experiment 1, the more prevention-focused the individual, the better they learned under the loss-framed feedback condition. In Experiment 2, the more promotion-focused the individual, the worse they learned under a points-free feedback condition. The differences may have resulted from different task affordances in the two experiments: a vigilant strategy (loss-avoiding) in Experiment 1 vs. an eager strategy (gains-seeking) in Experiment 2. Overall, these results highlight the relevance of regulatory focus for learner models, the complexity of regulatory fit (i.e., 3-way rather than 2-way fit), and how influential a small change in a task can be, if it changes the task's strategic affordances.

METHODS

Two similar experiments were run, differing in their number of participants and the timing and number of trials in the training task. Experiment 1 included 103 participants. Data from 10 of those participants were excluded based on pre-specified performance criteria, leaving 93 participants. Experiment 2 included 33 participants, of which 3 were excluded based on the same criteria, leaving 30 participants. Experiment 2 had fewer participants, because it was designed as a small-scale pilot for a future planned experiment. The voluntary, fully informed written consent of participants in this research was obtained as required by Title 32, Part 219 of the CFR and Army Regulation 70-25. All human subjects testing was approved by the Institutional Review Board of the United States Army Research Laboratory.

After completing an online pre-screener, participants were tested for normal visual acuity and color vision and completed a battery of questionnaires including the RFQ. After completing the questionnaires, participants completed the training task. In both experiments, the training task was a speeded go/no-go task with a computer-rendered character holding a gun as the go stimulus and the same character wearing different clothes and not holding a gun as the no-go stimulus. Go stimuli were four times as frequent as no-go stimuli. In both experiments, stimuli were visible for 400 ms and were presented at a randomized location on the screen. Participants had a limited time to press a response button in response to a go stimulus. In Experiment 1, participants were required to respond within 500 ms of image onset, whereas in Experiment 2, participants were required to respond within 1 s. After this deadline, feedback (see below) was displayed for 500 ms. In Experiment 1, the next trial began 500 ms after the end of the feedback, but in Experiment 2 the next trial began between 1 and 2 seconds later (uniform distribution).

Training in Experiment 1 consisted of 30 blocks of 30 trials each, lasting 20-30 minutes total. In Experiment 2, training consisted of 20 blocks of 30 trials each; because the trials were longer, training lasted 20-30 minutes.

After the training task, participants completed questionnaires about the training task. Next, participants completed the transfer task. The transfer task was a desktop simulation of being a passenger/spotter in a vehicle patrol of a middle-eastern-themed town with intermittent fog. As the vehicle proceeded, images would pop into the environment. The task was to classify those images as threats or non-threats, and to press a corresponding response button within 1 s of image onset. Two of the images were the go and no-go images from the training task. The other two were a table either with (threat) or without (non-threat) a table cloth obscuring the view under the table. In total, there were 200 images. Periodically, a diffuse fog would obscure the view to make the task more difficult. There were 5 periods of fog and 5 periods of no-fog, each averaging 1-minute in duration, and the transfer task took 10 minutes total. Finally, the participants completed another set of questionnaires.

The main independent variable of both studies was the framing of feedback in the training task. Participants were randomly assigned to point-gain-based feedback, point-loss-based feedback, or an informative control. In both the point gain and loss conditions, go trials were worth 30-60 points, with faster responses receiving more points and no-go trials were worth 180 points. In the gain condition, participants began with no points, and points were presented as gains. In the loss condition, participants began with the maximum points possible for a block, and points were presented as losses. For example, an average response time on a go trial in the gain condition would earn +45 points, but in the loss condition it would lose 15 points. These scoring systems are mathematically identical, but differ only in their framing. The control feedback showed a green check or red x to indicate correctness, and in the case of a response on a go trial, it also indicated response time.

Many outcome variables were measured; here we focus on two of them to illustrate the different outcomes of the two experiments. These outcomes are change in correct rejection (i.e. successfully not responding on no-go trials) rate over the course of training (i.e. the first 3 blocks vs the last 3 blocks), and accuracy in responding to the character stimuli out of fog in the transfer task. Both of these quantities are typically expressed as proportions, but for analysis they were analyzed as the logarithm of odds ratios (i.e. logit-transformed) in order to better meet the assumptions of linear modelling. Data from both experiments were combined and analyzed using a linear model with predictors of prevention strength, promotion strength, feedback condition (dummy coded), and experiment (1 or 2). The model included interaction terms for each strength with condition and experiment, condition with experiment, and the three-way interactions of strength, condition and experiment. Coefficients are reported with uncorrected 95% confidence intervals, and p-values are reported both uncorrected and corrected for multiple comparisons using false discovery rate (FDR).

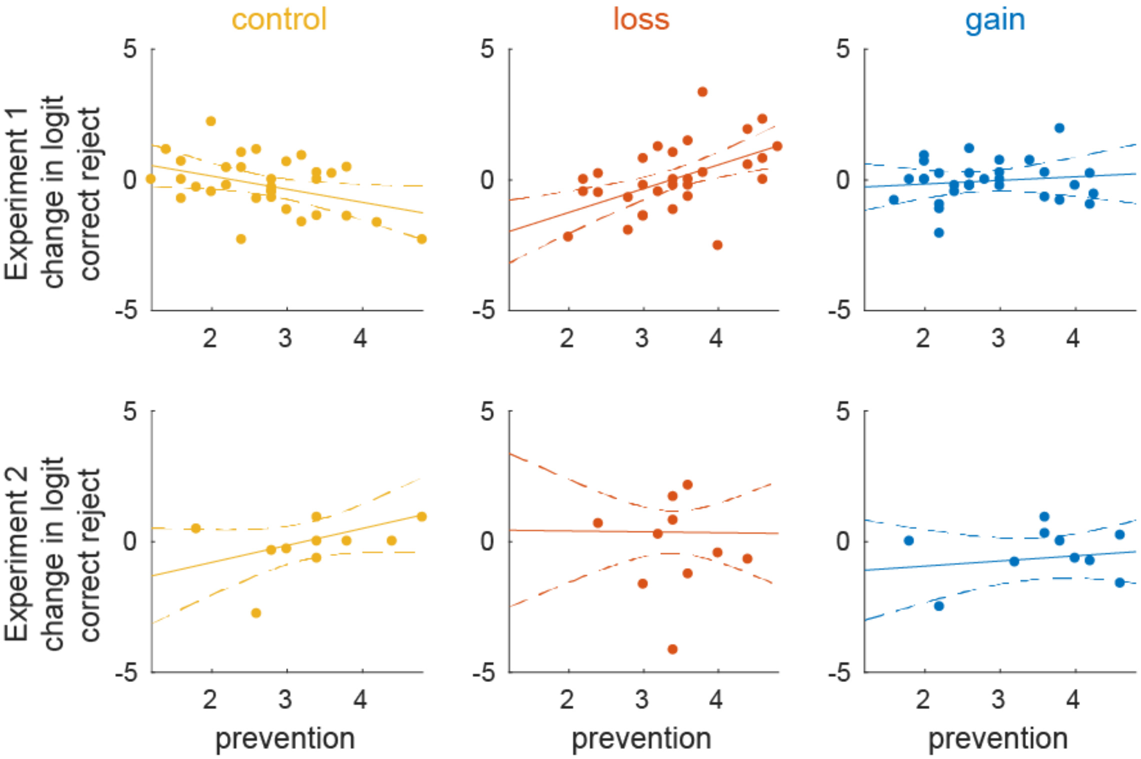


Figure 1. Change in logit correct rejection rate in control, loss and gain conditions across Experiments 1 & 2. Circles show individual participant results. Solid lines show expected values, and dashed lines show 95% confidence regions of the expected values.

RESULTS

Regression coefficient estimates with 95% confidence intervals appear in Table 1. There were no statistically significant effects of promotion strength on change in the logit correct rejection rate; however, there were effects and interactions involving prevention score, loss framing, and the experiment (1 or 2). The experiment term interacted with the effect of prevention strength, *B* = 1.15 [0.21, 2.10]

T(105) = 2.42, *p* = .017 (*p* = .065 FDR), the loss condition *B* = 10.86 [1.7, 20.02] *T*(105) = 2.35, *p* = .021 (*p* = .070 FDR), and their interaction, *B*=-2.10 [-3.78, -0.41], *T*(105)=-2.47, *p* = .015 (*p* = .064 FDR). These interactions are visualized with slice plots (Figure 1) showing how expected change in the logit correct rejection rate varies with prevention strength under the three conditions and in the two experiments when promotion strength is held constant at the sample average.

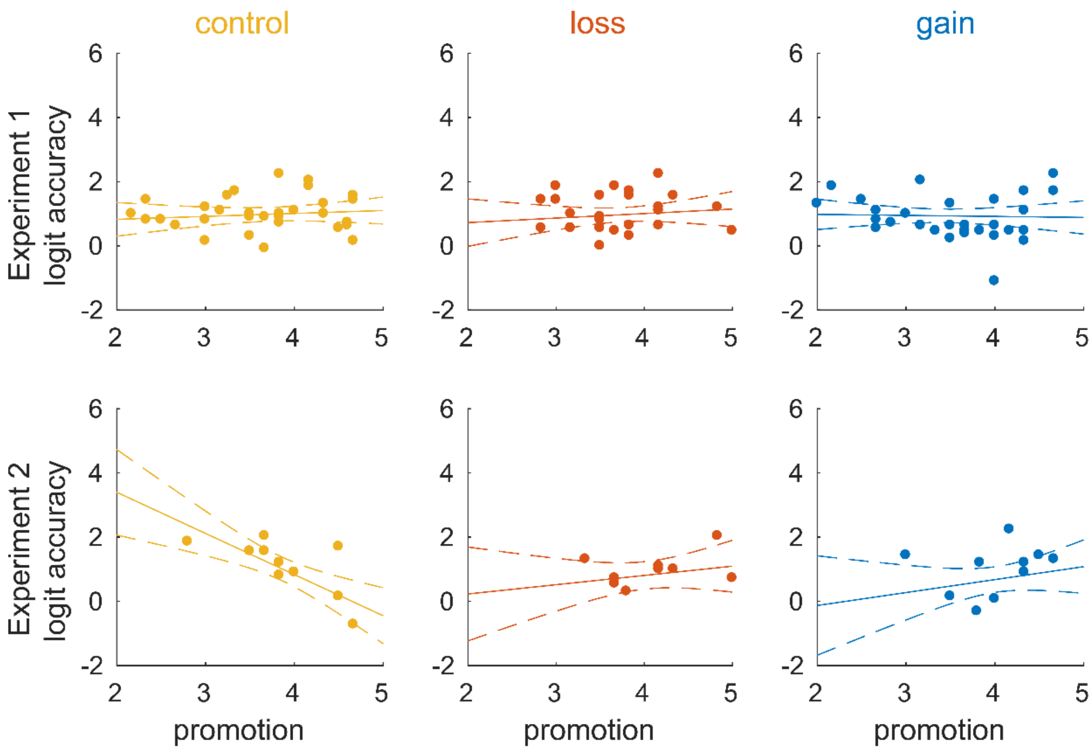


Figure 2. Logit accuracy on the trained stimulus with no fog in control, loss and gain conditions across Experiments 1 & 2. Circles show individual participant results. Solid lines show expected values, and dashed lines show 95% confidence regions of the expected values.

In the analysis of logit accuracy on the transfer task (Figure 2), the experiment factor interacted with promotion strength, *B* = -1.37 [-2.12, -0.63], *T*(105) = -3.66, *p* < .001 (*p* = .007 FDR), the interactions of promotion strength with loss framing, *B* = 1.52 [0.43, 2.62], *T*(105) = 2.76, *p* = .007 (*p* = .043 FDR), and promotion strength with gain framing, *B* = 1.81 [0.72, 2.90], *T*(105) = 3.28, *p* = .001 (*p* = .014 FDR). These reflect a negative effect of promotion strength on performance in the control condition in Experiment 2 that was not apparent in Experiment 1; moreover, this negative effect is counter-acted in both the gain and the loss conditions by effects in the opposite direction of the coefficient on the control condition.

Table 1. Regression coefficients and statistics

	B	95% CI		tStat	p	FD
Change in logit correct rejection rate						
(Intercept)	1.57					
prev.	-	-0.96	-	-2.16	.03	.091
pro.	-	-0.63	0.41	-0.43	.66	.711
loss	-	-	-	-3.23	.00	.014

gain	-	-6.08	0.82	-1.51	.13	.239
Exp. 2	1.09	-4.36	6.54	0.40	.69	.713
prev:loss	1.41	0.72	2.10	4.06	.00	.003
pro.:loss	0.68	-0.23	1.59	1.47	.14	.244
prev.:gain	0.64	-0.05	1.34	1.83	.07	.158

Continues

Table 1 Continued

	B	95% CI		tStat	p	FD
Change in logit correct rejection rate						
pro.:gain	0.28	-0.48	1.04	0.74	.46	.563
prev:Exp. 2	1.15	0.21	2.10	2.42	.01	.065
pro.:Exp. 2	-	-2.56	0.21	-1.68	.09	.188
loss:Exp. 2	10.8	1.70	20.0	2.35	.02	.070
gain:Exp. 2	-	-	5.67	-0.69	.48	.574
prev.:loss:Ex	-	-3.78	-	-2.47	.01	.064
pro.:loss:Exp.	-	-3.14	0.93	-1.07	.28	.405
prev.:gain:Ex	-	-2.40	0.21	-1.66	.10	.188
pro.:gain:Exp	1.46	-0.57	3.49	1.42	.15	.255
Logit accuracy for trained stimuli out of fog						
(Intercept)	1.04					
prev.	-	-0.38	0.12	-1.03	.30	.414
pro.	0.09	-0.19	0.37	0.65	.51	.587
loss	-	-3.70	0.74	-1.32	.18	.279
gain	-	-2.66	1.05	-0.86	.39	.492
Exp. 2	3.16	0.23	6.10	2.14	.03	.091
prev:loss	0.41	0.04	0.78	2.21	.02	.090
pro.:loss	0.05	-0.44	0.54	0.20	.84	.843
prev.:gain	0.39	0.02	0.76	2.07	.04	.099
pro.:gain	-	-0.53	0.29	-0.60	.55	.606
prev:Exp. 2	0.70	0.19	1.21	2.73	.00	.043
pro.:Exp. 2	-	-2.12	-	-3.66	.00	.007
loss:Exp. 2	-	-9.08	0.77	-1.67	.09	.188
gain:Exp. 2	-	-	-	-2.62	.01	.049
prev.:loss:Ex	-	-1.54	0.27	-1.39	.16	.261
pro.:loss:Exp.	1.52	0.43	2.62	2.76	.00	.043
prev.:gain:Ex	-	-1.06	0.35	-1.01	.31	.414
pro.:gain:Exp	1.81	0.72	2.90	3.28	.00	.014

DISCUSSION

Although both experiments demonstrated effects of regulatory fit between participant regulatory focus and feedback framing, the effects were different. This suggests that regulatory focus could usefully be incorporated into individual learner models, but that these models might also need to be task-dependent. The differences between our two experimental training tasks were in the timing and number of the trials.

The different regulatory fit relationships may result from the different strategic affordances these timing differences offer. In Experiment 1, responses were required to be fast (< 500 ms), and there was no variability in the inter-trial-interval. These factors together may have encouraged a strategy in which responding was essentially automatic unless it was canceled by some inhibitory process. In other words, success on this version of the task may have relied upon the participant adopting a vigilant strategy of avoiding false alarms on no-go trials. In Experiment 2, the slower pace and the unpredictability of stimulus onset might have reduced the automaticity of the go response, so rather than focusing on avoiding errors, participants might have focused on quickly reacting to stimuli and therefore relied on an eager/approach strategy. With only 30 participants, this interpretation should be considered tentative until more data is collected. Taken together, these experiments are consistent with the effect of feedback framing on performance depending on a three-way interaction among individual regulatory focus, feedback framing, and the strategic affordance of the task in question.

The three-way interaction has practical consequences, in that it would lead to recommending different point-based feedback interventions based not only on an individual's regulatory focus but also on the nuances of the task. For example, framing feedback in terms of loss of points appears beneficial for training prevention-focused individuals, but only if the task itself has a prevention-like (e.g., vigilant) strategy. Applying loss-based feedback for prevention-focused individuals in other tasks may not be helpful. In the case of Experiment 2 (eager/approach task strategy), we found that the more promotion-focused an individual, the worse they did in the absence of point-based feedback. Either gain-framed or loss-framed points feedback eliminated this performance decrement. This unexpected result may have come about due to the extra and more variable timing in the second experiment. There may have been just enough time to allow the promotion-oriented participants to interpret either form of point-based feedback as indicative of achievement. This highlights the potential complexity of regulatory fit theory and of its application in practice.

Overall, our findings point toward the need to include regulatory focus as a trait in individual learner models (see also Reinerman-Jones, Lameier, Biddle, & Boyce, 2017), as a potential source of adaptation (Goldberg et al., 2012) in training frameworks. More work is needed to develop an ontology of tasks and their strategic affordances in order to better predict the interaction effects of regulatory focus with different kinds of feedback, and the resulting effects on learner performance. Stronger predictive models could be incorporated into GIFT to support optimal feedback framing selection in different task domains.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This work examined the effects of regulatory focus and feedback framing on performance in two go/no-go training tasks that differed in the timing and number of trials. Three major conclusions stem from this work.

1) Regulatory focus is an important individual trait worth including in learner models for improving training outcomes.

Regulatory focus describes an individual's goal orientation. It is an individual trait tied to reward-based behavioral motivation, and thus is expected to influence how different individuals respond to reward-based training interventions and feedback. Our work revealed significant effects of regulatory focus on how individual trainees responded to feedback framing in a go/no-go paradigm. Trainees' prevention focus or promotion focus, under different feedback conditions and different strategic affordances, predicted performance improvements or decrements. Regulatory focus is simple to measure with a short questionnaire and can be included in learner models. These may be used by learner modules to determine states and thus help pedagogical agents select appropriate feedback framing options to maximize performance.

2) Regarding regulatory fit theory, a 3-way model of regulatory focus x feedback-framing x task strategic affordances may be more predictive of training outcomes than the traditional 2-way model of regulatory focus x feedback-framing.

The timing differences in our go/no-go paradigms yielded outwardly similar tasks that nonetheless differed in their strategic affordances. The first experiment's task encouraged a vigilant (i.e., error-avoiding) strategy by creating a rhythmic, pre-potent response to go stimuli that required inhibitory control to prevent that response in no-go trials. The second experiment's task encouraged a more eager (i.e., achievement-seeking) strategy by rewarding rapid response to go stimuli that were less predictable in their onset. By exploring the relationship between regulatory focus and feedback framing on two strategically different tasks, we uncovered evidence of a 3-way regulatory fit effect. The mechanisms underlying this effect remain to be examined in future work. Measurements of motivation, attention, or other affective or physiological states may shed light on what mediates the 3-way regulatory fit effect.

3) Small differences in training tasks, such as the timing differences in our study, may substantially affect the way that human variability dimensions interact with feedback framing and other personalized training interventions.

The scientific literature shows mixed results for a variety of training interventions, including various points-based reward schemes used for gamifying training tasks (Hamari, Koivisto, & Sarsa, 2014; Hanus & Fox, 2015; Seaborn & Fels, 2015) One possible explanation for this variability is that superficially similar tasks may in fact encourage different strategies, and the most effective feedback framing may depend on the strategy that the task is encouraging trainees to use. In our study, a subtle difference in timing was enough to yield tasks that relied more or less on vigilant vs. eager strategies, even though both were go/no-go tasks with the same visual stimuli and same points-based feedback. This highlights a need to think clearly about what strategies a given training task may afford. It may be beneficial to develop an ontology of strategic affordances of candidate tasks and consult this when designing training interventions that rely on regulatory fit or, by extension, fit with other individual trainee traits or states. Strategic affordance may be a useful variable to include in domain modules in intelligent tutoring systems like GIFT.

REFERENCES

Alder, G. S. (2007). Examining the relationship between feedback and performance in a monitored environment: A clarification and extension of feedback intervention theory. *The Journal of High Technology Management Research*, *17*(2), 157–174. https://doi.org/10.1016/j.hitech.2006.11.004

Berkman, E. T., Kahn, L. E., & Merchant, J. S. (2014). Training-Induced Changes in Inhibitory Control Network Activity. *Journal of Neuroscience*, *34*(1), 149–157. https://doi.org/10.1523/JNEUROSCI.3564-13.2014

Biggs, A. T., Cain, M. S., & Mitroff, S. R. (2015). Cognitive Training Can Reduce Civilian Casualties in a Simulated Shooting Environment. *Psychological Science*, *26*(8), 1164–1176. https://doi.org/10.1177/0956797615579274

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319. https://doi.org/10.1016/j.jecp.2004.01.002

Fillmore, M. T., & Rush, C. R. (2002). Impaired inhibitory control of behavior in chronic cocaine users. *Drug and Alcohol Dependence*, *66*(3), 265–273. https://doi.org/10.1016/S0376-8716(01)00206-X

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., R Billings, D., & Malone, N. (2012). *Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies*. https://doi.org/10.13140/2.1.1531.1367

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). https://doi.org/10.1109/HICSS.2014.377

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Higgins, E. T. (1998). Promotion and Prevention: Regulatory Focus as A Motivational Principle. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 1–46). Academic Press. https://doi.org/10.1016/S0065-2601(08)60381-0

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*(11), 1217–1230. http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1037/0003-066X.55.11.1217

Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: Promotion pride versus prevention pride. *European Journal of Social Psychology*, *31*(1), 3–23. https://doi.org/10.1002/ejsp.27

Houben, K. (2011). Overcoming the urge to splurge: Influencing eating behavior by manipulating inhibitory control. *Journal of Behavior Therapy and Experimental Psychiatry*, *42*(3), 384–388. https://doi.org/10.1016/j.jbtep.2011.02.008

Kamarajan, C., Porjesz, B., Jones, K. A., Choi, K., Chorlian, D. B., Padmanabhapillai, A., … Begleiter, H. (2004). The role of brain oscillations as functional correlates of cognitive systems: A study of frontal inhibitory control in alcoholism. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *51*(2), 155–180.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. http://dx.doi.org.proxy.library.ucsb.edu:2048/10.1037/0033-2909.119.2.254

Reinerman-Jones, L., Lameier, E., Biddle, E., & Boyce, M. W. (2017). *Informing the Long-Term Learner Model: Motivating the Adult Learner (Phase 1)* (No. ARL-TR-8160). US Army Research Laboratory Aberdeen Proving Ground United States, US Army Research Laboratory Aberdeen Proving Ground United States. Retrieved from http://www.dtic.mil/docs/citations/AD1041013

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Shenoy, P., & Yu, A. J. (2011). Rational Decision-Making in Inhibitory Control. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00048

Summerville, A., & Roese, N. J. (2008). Self-Report Measures of Individual Differences in Regulatory Focus: A Cautionary Note. *Journal of Research in Personality*, *42*(1), 247–254. https://doi.org/10.1016/j.jrp.2007.05.005

Todd, J., & Mullan, B. (2014). The Role of Self-Monitoring and Response Inhibition in Improving Sleep Behaviours. *International Journal of Behavioral Medicine*, *21*(3), 470–477. https://doi.org/10.1007/s12529-013-9328-8

Van-Dijk, D., & Kluger, A. N. (2004). Feedback Sign Effect on Motivation: Is it Moderated by Regulatory Focus? *Applied Psychology*, *53*(1), 113–135. https://doi.org/10.1111/j.1464-0597.2004.00163.x

Wang-Costello, J., Goldberg, B., Tarr, R., Contron, L., & Jiang, H. (2013). Creating an Advanced Pedagogical Model to Improve Intelligent Tutoring Technologies.

ABOUT THE AUTHORS

Ashley Oiknine *is a graduate and Visiting Scholar at the University of California, Santa Barbara and holds a BA in Psychology with an Applied Psychology minor. Her background studies and research experience include physiological data collection, cognitive psychology, and research methodology. She is currently a Research Analyst under DCS Corporation for the US Army Research Lab West's Training Effectiveness Group in Playa Vista, CA under the ICE Branch of the FST Division in ARL-HRED. Alongside the Training Effectiveness Group, she has been investigating individual differences of gamified training and virtual environments. Her research interests include human factors, psychophysiology, and neuroscience.*

Dr. Kimberly Pollard *is a Research Biologist in the Synergistic Human-Machine Interfaces Branch of ARL-HRED. She previously worked as an Oak Ridge Associated Universities (ORAU) Postdoctoral Fellow at ARL and as a National Science Foundation (NSF) Graduate Research Fellow at UCLA. Dr. Pollard earned her B.A. in biology at Rice University and her Ph.D. in biology (specializing in behavioral ecology) from UCLA. Dr. Pollard's research focuses on the areas of behavior and perception, including work on virtual training environments, human-robot interaction, sensory perception, social behavior, and individual differences.*

Dr. Peter Khooshabehadeh *is acting Regional Lead of ARL West and a cognitive scientist in the ICE Branch of the FST Division in ARL-HRED. Dr. Khooshabehadeh's background is in spatial, socio-cultural, and emotional cognition. He uses several methods, including interactive virtual environment technology, eye tracking, and psycho-physiological recording, to study human-computer interaction and thinking in several different domains and with different user populations; much of his research investigates individual differences. He received his B.A. from UC Berkeley in cognitive science, with emphases in both computational modeling and cognitive psychology, and Ph.D. in cognitive and perceptual science from UC Santa Barbara, where he was both an NSF and Department of Homeland Security Graduate Research Fellow.*

Dr. Antony Passaro *is a cognitive neuroscientist within the ICE Branch, FST Division in ARL-HRED. Dr. Passaro's background is in cognitive neuroimaging, including EEG, magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging (DTI). Dr. Passaro recently authored a chapter on DTI in a neuroscience textbook for Oxford University Press. He received his B.A. at Rice University in cognitive sciences and his Ph.D. in neuroscience at The University of Texas Health Science Center in Houston.*

Dr. Benjamin Files *is a neuroscientist within the Real-world Soldier Quantification Branch, Future Soldier Technologies (FST) Division at the U.S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED). Dr. Files' background is in neuroscience and vision science, with techniques including psychophysical testing, machine learning and EEG. Dr. Files received his B.A. at the University of California, Berkeley in neurobiology and his Ph.D. in neuroscience at the University of Southern California, where he was a College Doctoral and NIH Hearing & Communication Neuroscience fellow.*

Theme VI:
Domain Modeling

Expanding Domain Modeling in GIFT: 2018 Update

Robert A. Sottilare, Ph.D.
US Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED)
Learning in Intelligent Tutoring Environments (LITE) Lab
Center for Adaptive Instructional Sciences (CAIS)

INTRODUCTION

Building upon last year's domain modeling update (Sottilare, 2017), the purpose of this paper is to educate users of the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, & Holden, 2012; Sottilare, Brawner, Sinatra, & Johnston, 2017) about new and emerging capabilities to represent a broader variety of task domains in Intelligent Tutoring Systems (ITSs) in support of adaptive instruction. Adaptive instruction delivers content, offers feedback, and intervenes with learners based on tailored strategies and tactics with the goal of optimizing learning, performance, retention, and transfer of skills for both individual learners and teams.

GIFT is a tutoring architecture that has evolved over the last six years with three primary goals: 1) reduce the time and skill required to author ITSs, 2) automate best practices of instruction in the policy, strategies, and tactics of tutoring, and 3) provide a testbed to assess the effectiveness of adaptive instructional tools and methods with respect to learning, performance, retention, and transfer of skills. Another overarching goal for GIFT has been to adapt ITSs to provide instruction in militarily-relevant training and educational domains. For training domains, this means psychomotor tasks that involve both physical and cognitive aspects.

Currently, most ITSs are focused on cognitive task domains (e.g., problem solving and decision making) in academic topics that primarily include software programming, physics, and mathematics. While there are many military task domains that involve cognitive skill development (e.g., military planning processes and assessment of battlespace strategies and tactics), many more involve interdependent team processes (e.g., building clearing) and psychomotor skills (e.g., marksmanship). It is for this reason that we desire to extend current capabilities in GIFT to support content delivery, assessment, and remediation processes for more complex team and psychomotor tasks while simultaneously enhancing the effectiveness of individual instruction in cognitive and affective domains. In 2015, Sottilare, Sinatra, Boyce, & Graesser documented domain modeling goals, challenges and approaches to providing adaptive instruction in various domains. The following section describes some of the challenges to expanding domain modeling beyond cognitive tasks and beyond the current model of desktop training.

The following sections examine areas of enhanced, new or emerging capabilities in support of expanding GIFT to a wider variety of task domains.

TUTORING MARKSMANSHIP

While this was reported in last year's update (Sottilare, 2017), it is worth noting that there remains growth potential in the marksmanship task domain. Although, GIFT has now been integrated with PEO STRI's Engagement Skills Trainer (EST) to demonstrate interaction of the learner, there is more to be done to fully demonstrate GIFT as a psychomotor task tutor. The current implementation provides training with stationary targets, assessment of the learner's performance, and remediation of any detected errors by the tutor with respect to the Army marksmanship principles. Ideally, future versions of GIFT will also allow the integration of new weapons (e.g., different rifles and pistols) and their associated expert models. We project that many of the sensors needed to acquire weapon cant and aim points could remain the same depending on the size of the weapon and certainly the breathing harness would not change with a change in weapons.

To ease the process for developing ITSs for psychomotor task domains, we have invested in an agent-based approach to guide authoring of psychomotor tasks (see paper #2 in this year's GIFTSym by Brown, Goldberg, Bell, & Kelsey, 2018). This approach includes automated acquisition of sensor data and uses this data with reinforcement learning to develop expert models for psychomotor tasks.

TUTORING MEDICAL TASKS

Previously, we reported that GIFT had been used to provide tailored training of military tasks using desktop applications (e.g., Virtual Battlespace and Virtual Medic). The degree of transfer of skills from training to operations was limited in these environments since the training primarily exercised cognitive functions. So in 2016, Sottilare, Hackett, Pike & LaViola examined how commercial sensor technologies might be adapted to work with GIFT and support tailored computer-guided instruction in the psychomotor domain for a military medical training task, specifically hemorrhage control. While this concept was well-thought out, the implementation has been hampered by changes in technology, specifically the turnover of commercial smart glasses in the market.

Recently, Julian (2018) applied GIFT to the task of basic robotic surgical skills. The purpose is to help train physicians on both the cognitive and basic knowledge of skills needed to use the most commonly known robotic surgical system, the da Vinci. Two skills were taught in the GIFT-based course: camera control and interrupted suturing. Again, the focus of the instruction was primarily cognitive (knowledge components) and GIFT's ability to support physical measures during practice were limited. Ideally, some type of board or mannequin might be used in combination with sensors to detect the delicate control required for this type of robot-assisted surgery and we are evaluating how this might be accomplished across a variety of tasks. One approach could be embedded training where GIFT is used to stimulate a system (e.g., da Vinci) and the interface used by the learner is the same one used during the actual work task. This type of approach would reduce any negative training introduced by poor attempts to replicate the interface.

Another potential medical domain application of GIFT is being developed at Columbia University and the Morgan Stanley Children's Hospital in New York. The pediatric physicians on staff at the hospital are exploring the use of GIFT to train pediatric residents. The ARL adaptive training team provided a short course on authoring using GIFT in January 2018 and the staff is assembling content for their first course. The intent is to use GIFT to augment the instruction of pediatric residents in a self-regulated (computer-regulated) learning environment.

On the research side of GIFT domain applications, we are engaged in the development of an experimental protocol to investigate accelerated learning models in GIFT for medical military and civilian training (Sottilare & DeFalco, 2018). Data collection has already begun and will involve several user groups from United States Military Academy, Columbia University, University of Wisconsin, and Amazon Mechanical Turk (MTurk).

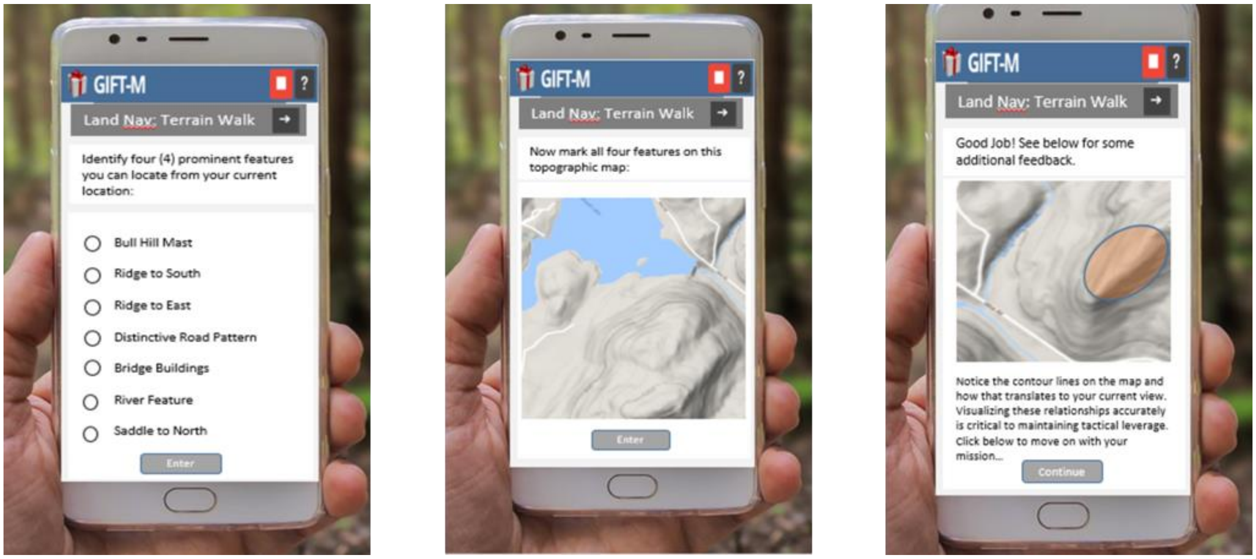
TUTORING PSYCHOMOTOR TASKS WITH TACTICAL BREATHING

Last year, we reported information about an experimental approach involving psychomotor tasks and tactical breathing (Kim, Dancy, Goldberg, & Sottilare, 2017). Tactical breathing is a specific breath-control technique used by individuals to perform a precision action required for a psychomotor task in a stressful environment (Neumann & Thomas, 2009; Neumann & Thomas, 2011). The focus of this research is to examine the relationship between cognitive (e.g., attentional resources) and physiological (e.g., breathing) factors during the execution a psychomotor task (i.e., golf putting). It is not well understood that how the corresponding mechanisms of attentional control interact with the physiological factors as the learner progresses to the learning stage. An experimental protocol has been drafted and the experimental apparatus is being developed to support the measurement of critical factors during task performance. Data collection is scheduled for the Fall of 2018.

TUTORING IN THE WILD (LIVE, AUGMENTED OR MIXED REALITY)

An important aspect of the value of ITSs is associated with their accessibility or ability to go where learners go. Often referred to as *mobile learning,* instruction delivered to portable computing devices (e.g., laptops computers, smartphones or tablet computers) and managed remotely by either human or artificially-intelligent tutors, we are advocating an expanded capability that could be delivered to learners in either live, augmented, or mixed reality environments. We consider this an important design feature for ITSs so that they can support learning as an augmentation in a variety of environments where military personnel might be assigned. To this end, we continue to examine opportunities to link GIFT through interfaces that can expand learner experience and knowledge.

As reported last year, we are continuing to evaluate the application of various hardware platforms (e.g., smartglasses, mobile devices). A large part of our domain application effort this year has been dedicated to providing a proof of concept for land navigation training to USMA. This concept provides learners a means of planning their routes (Virtual BattleSpace) and executing their routes (live environment augmented with a phone-based mobile application (Figure 1).



Identify four (4) prominent features you can locate from your current location:

○ Bull Hill Mast
○ Ridge to South
○ Ridge to East
○ Distinctive Road Pattern
○ Bridge Buildings
○ River Feature
○ Saddle to North

Now mark all four features on this topographic map:

Good Job! See below for some additional feedback.

Notice the contour lines on the map and how that translates to your current view. Visualizing these relationships accurately is critical to maintaining tactical leverage. Click below to move on with your mission...

Figure 1. GIFT Mobile Application for Land Navigation Training

Moving we anticipate more eyes on the *tutoring in the wild* problem space. Last year, the North Atlantic Treaty Organization (NATO) approved a research task group (RTG) to examine existing and emerging augmentation technologies to enhance human performance in both instructional (training and educational) domains as well as work/operational domains. The broad scope of this RTG includes the use of ITSs to deliver and manage instruction as well as support mission essential tasks as a job aid. In addition to the US, this group has garnered interest from eight NATO countries which implies its importance.

TUTORING TEAM DOMAINS: TEAMWORK AND TASKWORK

One way of extending domain-independence in GIFT to the modeling of teams is to separate domain-independent teamwork behaviors from task-specific, domain-dependent behaviors. Salas (2015) distinguishes teamwork, interactions between team members, from taskwork, behaviors demonstrated in executing the task. An examination of teamwork activities (e.g., coaching or conflict management) via a meta-analysis of the team training and performance literature led to the identification of several behavior markers for high performing teams (Sottilare, et al, 2017). Next steps are to seek methods to unobtrusively acquire these behavioral markers in order to identify team states and subsequently assign the ITS to manage them.

Currently, there are no tools or methods available in the public baseline for modeling or tutoring teams in GIFT. We are continuing to develop a model of team tutoring in which we will incrementally provide team instruction through GIFT without human intervention. While the specific approach is not yet set in stone, it might look something like this:

- Configure GIFT to identify hierarchical concepts or learning objectives associated with team taskwork (GIFT can already do this)
- Configure GIFT authoring tools to support the development of team models and associated measures
- Configure GIFT authoring tools to support the development of sub-team and multiple individual models and associated measures, roles, and responsibilities
- Configure GIFT authoring tools to support the development of a team strategy engine based on teamwork (domain independent) best practices.

ACKNOWLEDGMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory. Statements and opinions expressed in this paper do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Brown, D., Goldberg, B., Bell, B., & Kelsey, E. (2018, *in press*). Incorporating psychomotor skills training into GIFT tutors: Supporting outside-the-box authoring. In Proceedings of the 6th GIFT Users Symposium. Orlando, Florida, May 2018.

Kim, J., Dancy, C., Goldberg, B., & Sottilare, R. (2017). A Cognitive Modeling Approach - Does Tactical Breathing in a Psychomotor Task Influence Skill Development during Adaptive Instruction? In *Foundations of Augmented Cognition*. Springer International Publishing.

Julian, D. (2018). Final Report for IDS6938 – Intelligent Tutoring System Design: Basic Robotic Course. University of Central Florida, Orlando, FL.

Neumann, D.L. & Thomas, P.R. (2009). The relationship between skill level and patterns in cardiac and respiratory activity during golf putting. Int. J. Psychophysiol. 72(3), 276–282.

Neumann, D.L. & Thomas, P.R. (2011). Cardiac and respiratory activity and golf putting performance under attentional focus instructions. Psychol. Sport Exerc. 12(4), 451–459.

Salas, E. (2015). Team Training Essentials: A research-based guide. Routledge Publishing. New York & London.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Concept paper released as part of GIFT software documentation. Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED). Retrieved from: https://gifttutoring.org/attachments/152/GIFTDescription_0.pdf

Sottilare, R. & DeFalco, J. (2018). Experimental Protocol (ARL 17-251) - Developing accelerated learning models in GIFT for medical military and civilian training. US Army Research Laboratory, Orlando, FL.

Sottilare, R., Hackett, M., Pike, W., and LaViola, J. (2016). Adaptive Instruction for Medical Training in the Psychomotor Domain. In J. Cohn, D. Fitzhugh, and H. Freeman (Eds.) Special Issue: Modeling and Simulation Technologies to Enhance and Optimize the DoD's Medical Readiness and Response Capabilities of the *Journal for Defense Modeling & Simulation (JDMS).*

Sottilare, R., Brawner, K., Sinatra, A., & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED).

Sottilare, R., Sinatra, A., Boyce, M., and Graesser, A. (2015). Domain Modeling for Adaptive Training and Education in Support of the US Army Learning Model: Research Outline. US Army Research Laboratory (ARL-SR-0325), June 2015.

Sottilare, R. (2017, May). Expanding Domain Modeling in GIFT. In R. Sottilare (Ed., 2017) 5th Annual GIFT Users Symposium (GIFTSym5). Army Research Laboratory, Orlando, Florida. ISBN: 978-0-9977257-1-1.

Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H. & Gilbert, S.B. (2017). Towards a Design Process for Adaptive Instruction of Teams: A Meta-Analysis. In R. Sottilare, A. Graesser, J. Lester & R. Baker (Eds.) Special Issue: Generalized Intelligent Framework for Tutoring (GIFT): Creating a stable and flexible platform for innovations in AIED research. *International Journal of Artificial Intelligence in Education.*

ABOUT THE AUTHOR

***Dr. Robert Sottilare** leads adaptive training research within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He is ARL's technical lead for the Center for Adaptive Instructional Sciences (CAIS).*

The 2018 Research Psychologist's Guide to GIFT

Anne M. Sinatra[1],
US Army Research Laboratory[1]

INTRODUCTION (DON'T PANIC)

When approaching making a sequel, whether it is a movie, book, or even a research paper, one of the goals is to recapture the elements that people enjoyed about the original, while building on it to provide a wholly new experience. A long time ago (4 years), in a galaxy not so far away, I wrote the first version of "The Research Psychologist's Guide to GIFT" (Sinatra, 2014). I started my guide with two important words: "Don't Panic". I also was sure to start my 2016 sequel to the guide (Sinatra, 2016), and the current paper with the same words. These words serve as a reminder that while the Generalized Intelligent Framework for Tutoring (GIFT) can seem overwhelming at times, it is a very powerful tool for a research psychologist, and there are guides such as this and previous ones that will help you so that you do not become overwhelmed. The words also reference the fictitious Hitchhiker's Guide to the Galaxy (Adams, 1979), which had those two words on the cover and sold more copies than any other similar text (due to be reassuring, and slightly cheaper).

The current work does not necessarily replace the previous two guides, but it builds on and updates them, as any good sequel should. GIFT ultimately is a research-based project, and as part of that model the software has been continuously updated. In the current guide, there will be explanations of the software overall, as well as changes that have occurred in GIFT 2017-1, GIFT Cloud, and the upcoming GIFT 2018-1.

WHAT IS THE GENERALIZED INTELLIGENT FRAMEWORK FOR TUTORING?

GIFT is a domain-independent framework for creating intelligent tutoring systems (ITSs) (Sottilare, Brawner, Sinatra & Johnston, 2017). In an ITS, the domain is traditionally highly coupled with the tutor itself. GIFT provides the tools and foundation such that an instructor, researcher, or subject matter expert can bring materials that he or she already has and use it to create new ITSs. If there is overlap between previously created courses or surveys they can be reused in the new course, and the materials (such as PowerPoints or PDFs) can be easily changed out for other existing or created materials. GIFT has been designed primarily for use in creating ITSs, however, its other goals include providing functionality to conduct research, and to be used as a testbed.

Research that is conducted with GIFT can take two forms: experiments specifically focused on the applications of intelligent tutoring, which can utilize remediation, and more traditional linear experiments. A general overview of different types of experimentation with GIFT is also available (Sinatra, 2017). The current paper will focus on the latter of these two types of experiments, the traditional linear psychology experiments that do not utilize ITS remediation. One of the strongest features of GIFT is the ability to put a number of materials of different types (e.g., PowerPoints, PDFs, html, images, and surveys) together in a consistent and fluid sequence that does not require intervention from an experimenter to move on from one file type to the next. This reduces the number of research assistants that are needed to run a study, reduces the possibility of human error, and allows for multiple participants to easily be simultaneously run in the experiment.

GIFT provides the tools to create a "course" or sequence items that are displayed in a specified order. To create an experiment with multiple conditions, the original course can be created, and copied. The appropriate changes can then be made to the relevant materials in each of the additional conditions. Once a participant arrives they can be setup on a computer that is running the necessary condition. Or if the study is being done online, they can be provided with a link that is specific to the experimental condition that they were assigned. GIFT provides tools to create the flow of the experiment, hosts/generates online links for cloud based studies, and provides ways for experimenters to extract their data from the experimental sessions.

There are two different versions of GIFT: desktop based and cloud based. The functionality of these two versions of GIFT is very similar, but there are considerations that should occur when deciding between which version of GIFT to use. In the current paper the differences between them will be briefly discussed, but previous Research Psychologist's Guide papers (Sinatra, 2014; Sinatra, 2016) focused primarily on the desktop based versions. For the purposes of the current paper, the focus will be primarily on the Cloud based implementation of GIFT, and how it can be utilized for experiments.

CREATING AN EXPERIMENT WITH GIFT

An experimenter can author a GIFT course that consists of the different components that he or she wants to present to the participants. As noted above, in order to create multiple conditions, multiple versions of the same course can be created with the appropriate manipulation of interest included in each of them.

GIFT Authoring Tools

The GIFT Authoring Tools are used to create GIFT courses, and have an easy to use drag and drop interface. The Authoring Tools have gone through many iterations and developments through the years, and the interface has been updated significantly since the previous Research Psychologist's Guide paper (Sinatra, 2016). The current version of the GIFT authoring tools is shown in Figure 1. The left side of the screen has course objects that can be added to the course, and the right side of the screen is a timeline that shows the flow of the GIFT course that is authored. The example in Figure 1 is a subset of one of the conditions of a previous experiment (Sinatra, Sims, Sottilare, 2014). The sequence of items displayed are, introductory text, a survey, information text, an interactive PowerPoint, an information text, a survey, introductory text, a survey, and finally information text. The course continues on and would be visible if the screen was scrolled by the experimenter. Each of these items will be automatically presented the participant in sequence without the need for intervention by a research assistant. In the case of PowerPoint, GIFT connects to the instance of PowerPoint that is on the participant computer and opens/closes it as appropriate.

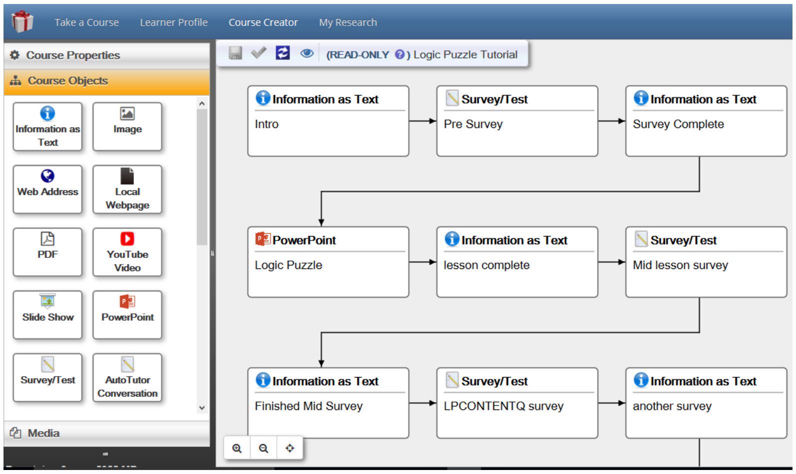


Figure 1. The current version of the GIFT authoring tools.

The structure of the course is created by the experimenter, and can be rearranged as needed. There are a number of "public" courses that are available by default with each new GIFT account. On the main tile login screen of GIFT, instead of clicking to run the course, you can click on "Edit Course" which will open it in the course authoring tool. As demonstrated in Figure 1, the courses may say "Read-Only", but they can still be opened in the authoring tool and the components, order and configurations can be viewed. If you would like to make changes to these courses you can do so by copying them on the course tile page, and making edits to the copy. These courses are a great place to start in order to get comfortable with how GIFT is set up.

Course Objects

In order to begin authoring a new course, the experimenter can access the Course Authoring Tool from the top menu, and then can drag an item from the left side of the screen onto the timeline on the right side. Each individual item (e.g., "Information as Text") can then be selected, and configured. The configuration window for each item will appear on the far right side of the screen when a course object from the timeline is selected.

Many of the course objects are straight forward and self-explanatory. For instance, "information as text" allows the experimenter to provide information (such as instructions) that will simply be presented to the participant. A web address, PDF, or local web page can also be used during the course.

An important distinction is made between the Slide Show object and the PowerPoint object. If you have already existing materials in PowerPoint that you would like to use, or you would like to create your informative material with PowerPoint's interface, you can then use what you created in your GIFT course. The Slide Show object should be used when there is no interaction between the participant and the PowerPoint course itself, and there is no multimedia present in it. The Slide Show object makes individual images of each slide that you upload, and converts it into an easy to navigate slideshow. In order to use it, the original file that is uploaded should be in .pps form (PowerPoint 2003 show form).

One of the advantages of using the Slide Show is that it does not need to open and close the actual PowerPoint program, which means that in the cloud this will run extremely smoothly and not require a connection to be made between the participant's computer and GIFT. If it is necessary to interact with PowerPoint using macros, Visual Basic for Applications or even to include multimedia such as sounds or videos, then a PowerPoint object needs to be used. In this case, when the individual runs the course in the cloud it will require downloading and installing a brief java webstart application to connect the PowerPoint software on the computer to GIFT. This can take time and requires that the participant go an extra step. Additionally, it not only requires that the participant have PowerPoint on their computer, but that it is a version that GIFT is compatible with. Therefore, unless the PowerPoint is interactive, or adaption needs to occur during it, the preferred item to use is the Slide Show.

As a note, with the Slide Show item there are navigational arrows at the top of the slides. One button will be to go back a slide (this option can be taken away if the course author wants it to be), to go forward a slide, and to complete the entire Slide Show. It may be helpful to provide instructions to the participants to let them know the difference between the arrow that goes forward a slide, and the one that completes the show so that they do not accidentally exit out of the show.

In the case of both objects, they will require a .pps (2003 PowerPoint show) file, but GIFT will handle them differently. This PowerPoint file is created by selecting “Save As” and scrolling to ‘PowerPoint 1997 – 2003 Show (*.pps)” in the “Save as type” selection. If a different type of file is used, it will not be able to be uploaded. Additionally, if macros are to be used for a PowerPoint object, the correct version to save is “PowerPoint Macro Enabled Show (*.ppsm)”. PowerPoint objects will require a GIFT compatible version of PowerPoint to be installed on the computer that is being used for participation. The Slide Show Object does not require that PowerPoint be on the computer that is being used. A quick reference table of when it is most advantageous to use each object is displayed in Table 1.

Table 1. Comparison Chart for when to use PowerPoint Objects vs. Slide Show Objects

Requirement	PowerPoint Object	Slide Show Object
PowerPoint with or without images and no interactions		X
Videos or Audio in the PowerPoint Presentation	X	
Visual Basic for Applications or Macros is used	X	
Online presentation of materials on participant’s own computer		X
Assessment or time spent on slides is needed	X	

Survey Authoring

The survey authoring system is a very important tool to a research psychologist. Since the previous version of this guide the survey has system has undergone significant changes and improvements. Usability was the number one focus of the changes, and many of the extraneous details such as setting up survey contexts were removed. However, an important distinction must be made between the Survey/Test object, and the Question Bank object. Additionally, within the Survey/Test object there are three options that can be selected for use when creating a survey. It is important to understand what the functions are of each type of survey to ensure the correct one is authored. Figure 2 provides a display of the three types of options provided when an experimenter pushes “Create New” on the Survey/Test window pane.

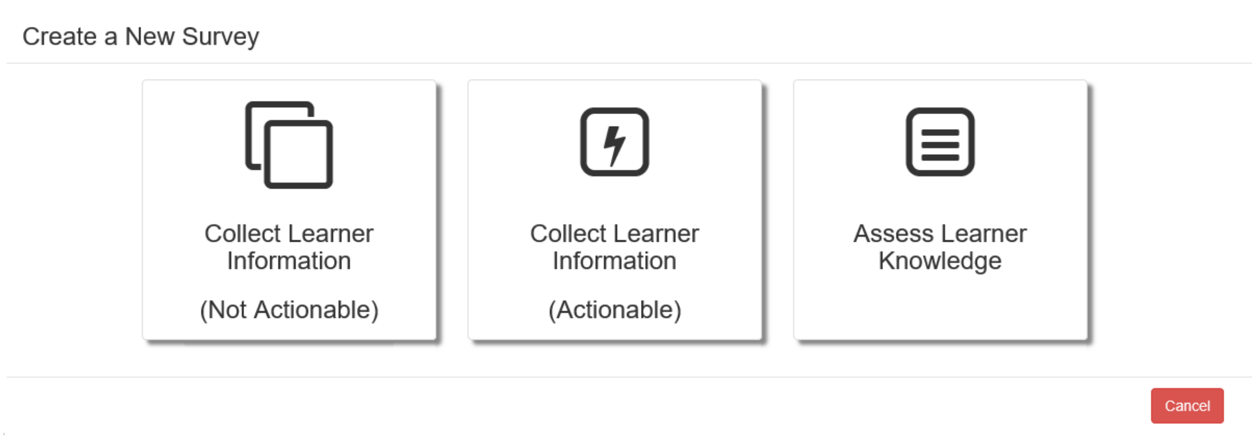
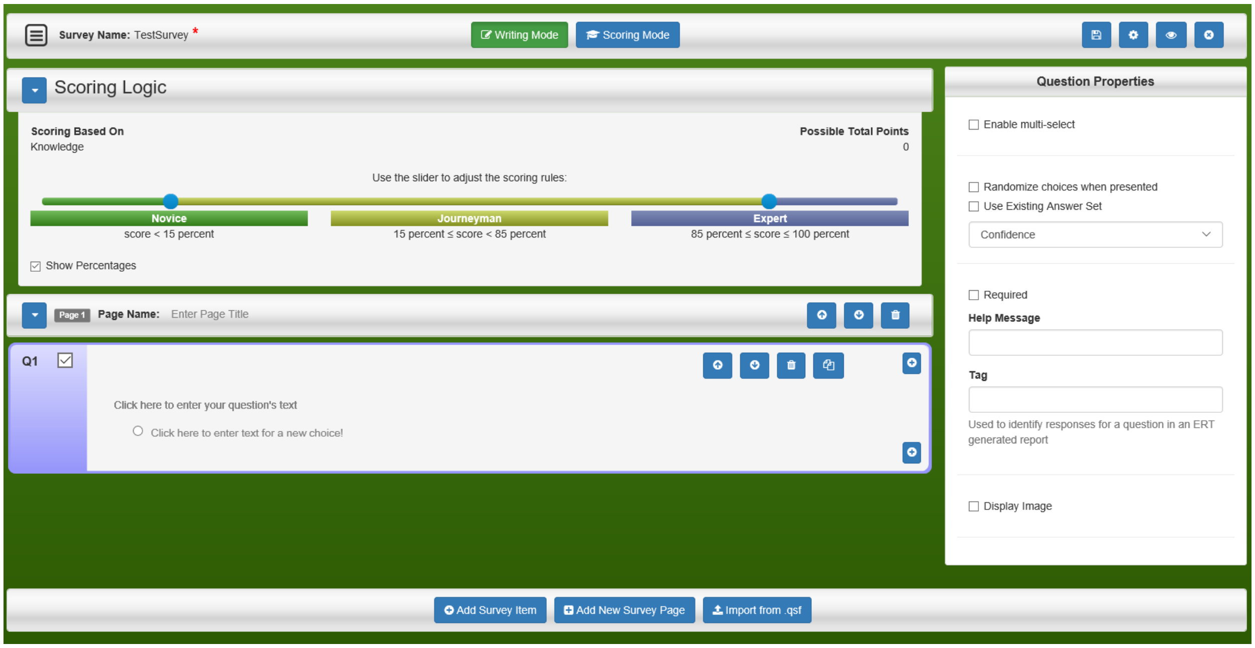


Figure 2. A screenshot of the three types of surveys.

In the case of a demographics survey, in which learner information will be recorded but not be used actively during the experiment the proper selection to use would be “Collect Learner Information”. If the intention was to assess the learner knowledge in the form of a score on a specific series of questions, then the option to select is “assess learner knowledge”. In this case the information will be actively graded during the course, and can be used to make selections about what materials are presented. The survey author also has the ability to note what is considered a novice, journeyman and expert for the specific survey. Figure 3 provides a screenshot example of the “Assess Learner Knowledge” option. Note that the slider at the top of the screen can be used to adjust the percentages correct that are required to be classified in each category.


Survey Name: TestSurvey *

Scoring Based On
Knowledge

Possible Total Points
0

Use the slider to adjust the scoring rules:

Novice	Journeyman	Expert
score < 15 percent	15 percent ≤ score < 85 percent	85 percent ≤ score ≤ 100 percent

☑ Show Percentages

Page 1 Page Name: Enter Page Title

Q1 ☑

Click here to enter your question's text

○ Click here to enter text for a new choice!

Question Properties

□ Enable multi-select

□ Randomize choices when presented

□ Use Existing Answer Set

Confidence

□ Required

Help Message

Tag

Used to identify responses for a question in an ERT generated report

□ Display Image

Figure 3. Screenshot of Assess Learner Knowledge Survey type.

Questions are added using the buttons on the bottom of the screen. After question text is added scoring and correct answers can be added by clicking on the “Scoring Mode” button on the top of the screen, which will add some additional features to the interface. The correct answer can be provided by putting a number point value such as 1 for correct and 0 for incorrect next to the options for each question.

The Question Bank object is similar to a survey, but draws on a larger set of questions that have been authored within the course. If an experimenter wants to provide a randomized selection of questions that are associated with specific concepts, he or she may choose to use the question bank. It is also necessary to define course concepts by clicking on the "Course Properties" tab on the main Course Authoring Tool page (this tab can be seen on the left side of the screen in Figure 1). The issue that may be run into when using a question bank in an experiment, is that unless the same number of questions that are authored is selected as the number of questions given, you cannot be sure of what questions will be provided to the participants. Therefore, this functionality is more advantageous for actual team tutors, or experiments in which the questions themselves are being varied. Table 2 provides a quick guide to when to use the Survey/Test Object as opposed to the Question Bank Object. The question bank is also utilized within an adaptive courseflow object, which his beyond the scope of the current paper.

Table 2. Comparison Chart for when to use Survey/Test Object vs. Question Bank Object

Requirement	Survey/Test Object	Question Bank Object
Present questions in a random order		X
Present questions that are associated with concepts		X
Collecting demographics information	X	
Using a questionnaire or measure that requires a specific order of presentation	X	
Not all generated questions are required to be answered		X
Using an assessment that requires all questions to be answered and to be shown in a specific order	X	
Questions will be reused in Adaptive Course Flow object		X

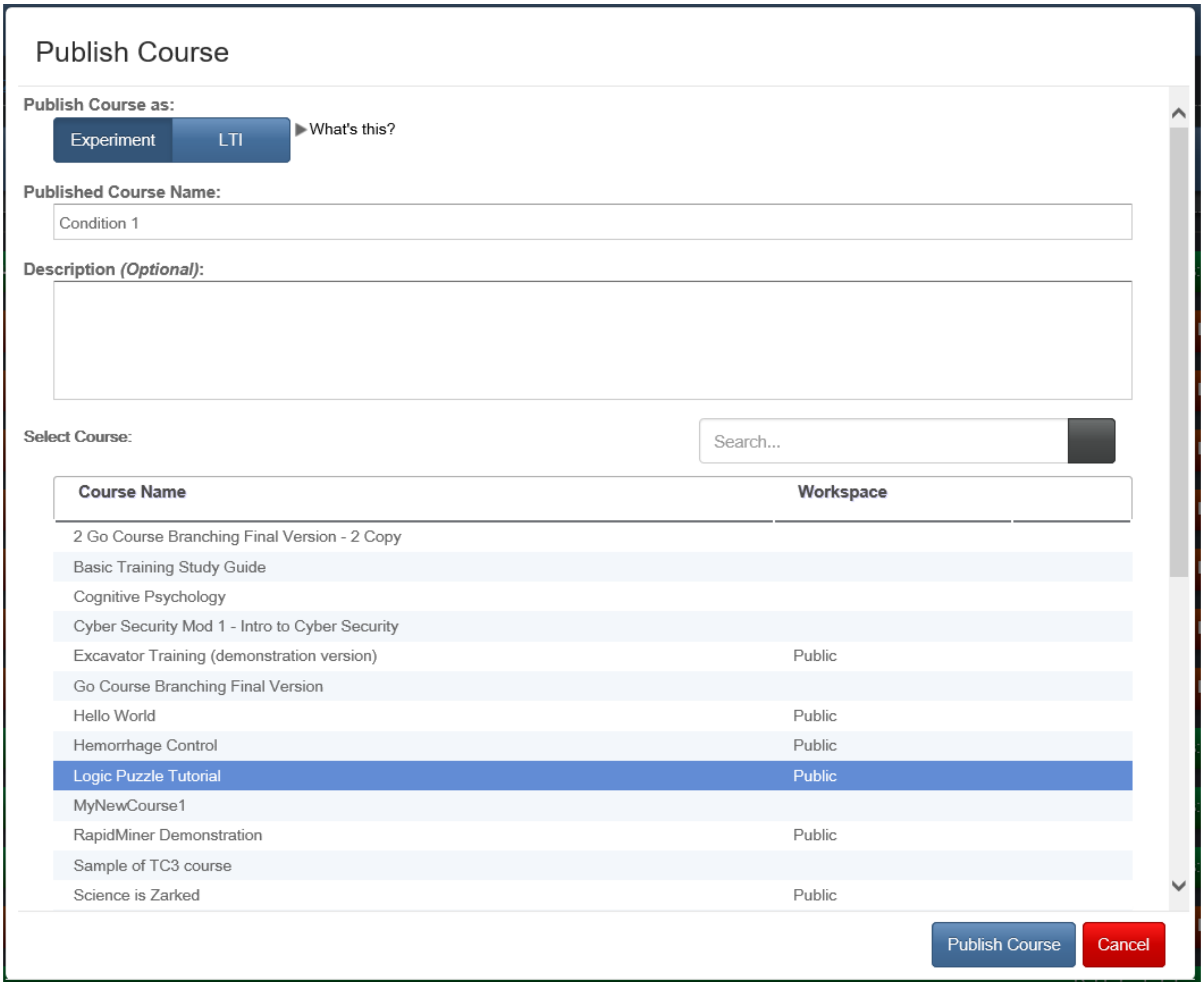
Creating a "Tag" or name for each question is extremely important, as otherwise it will output simply with a number associated with when the question was authored. The place to create the tag is visible on the right side of the screen in Figure 3. Without a tag, it is very difficult to figure out which question response is which. Therefore, when authoring the question be sure to put a short word in the "Tag" box that will let you know what it means so that you can go back later and look at the survey data easily.

EXPERIMENTAL PROCESS

A very important decision that will need to be made is whether the experiment will be run on the cloud or locally on a computer. The way that data is extracted and saved will be different depending on the approach that is used. In the previous guide (Sinatra, 2016) an in-depth explanation is provided about how to use "Experiment Mode" on the local computer. If you wish to use "Experiment Mode" and create experiment specific user IDs, please refer to the previous paper. For the current paper, the approach of "Publishing Courses" and the difference between desktop vs. cloud mode will be discussed.

Publish Course and Participant Management

Once you are happy with the course that you created you will click on "Public Course" at the top of the screen. You will then select the "Publish Course" button on the left side of the screen. This will bring up a screen that allows you to publish the course as an experiment, and add a specific course name and description. At the bottom of this screen you will select the course that you wish to publish. If you have three experimental conditions then you will do this three times, each with a different published course name that will make sense to you the experimenter, and your research assistants. See Figure 4 for a screenshot of the "Publish Course" screen.



Publish Course

Publish Course as: Experiment

Published Course Name: Condition 1

Description *(Optional)*:

Select Course:

Course Name	Workspace
2 Go Course Branching Final Version - 2 Copy	
Basic Training Study Guide	
Cognitive Psychology	
Cyber Security Mod 1 - Intro to Cyber Security	
Excavator Training (demonstration version)	Public
Go Course Branching Final Version	
Hello World	Public
Hemorrhage Control	Public
Logic Puzzle Tutorial	Public
MyNewCourse1	
RapidMiner Demonstration	Public
Sample of TC3 course	
Science is Zarked	Public

Figure 4. Screenshot of the Publish Course screen

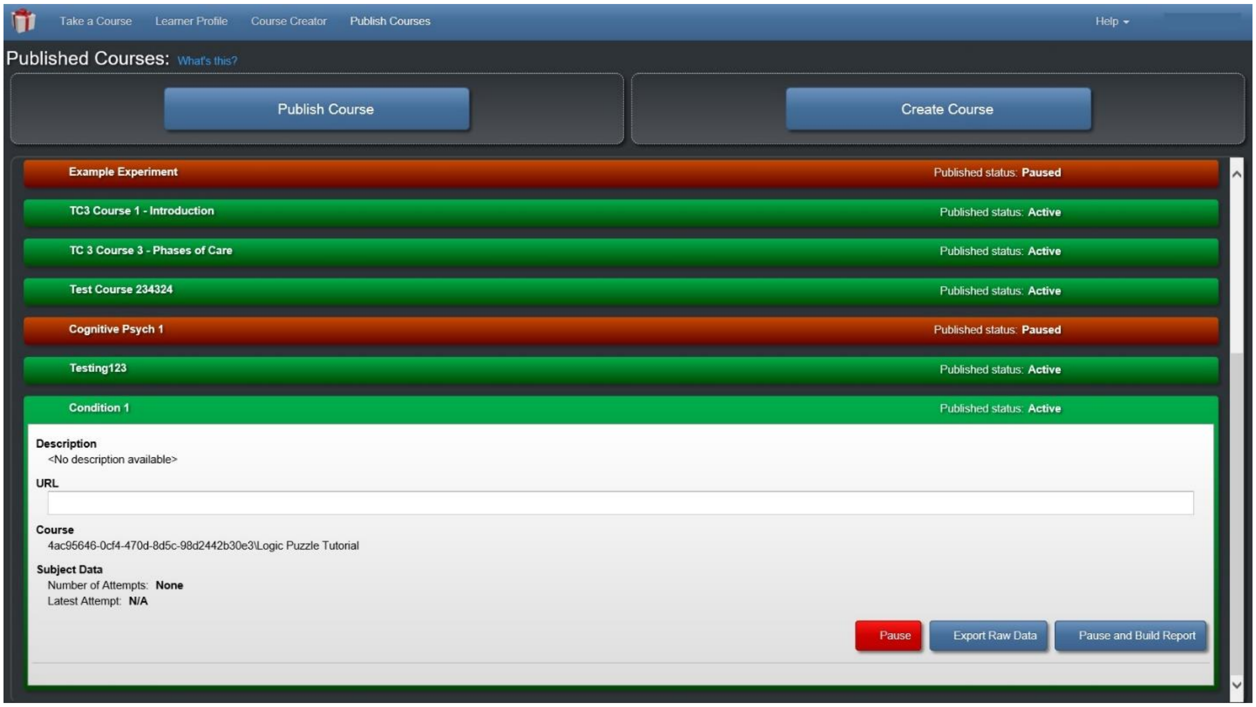
Once you "publish" a course it makes a copy of the course at the current moment in time. Any changes you make to the original course after this point will not populate into the previously published instance of the course. If you need to make a change you will need to publish the course again, using a different course name. It will be important to make sure that you save any collected participant data, as it will not transfer over to your new published course.

Once you create your course you will get a notice that says “Experiment Created!” and has the URL that you can provide to the participants to access it. If you are using the cloud version of GIFT, this will be a cloud.gifttutoring.org URL. If you are using the desktop version, it will be a local URL that can be pasted into the web browser to bring up the course on your computer without needed the participant to log in to a GIFT account.

Since publishing your course does not require a GIFT account or login, you will need to be sure to clearly provide a participant number to each participant. You will also need to author a survey item at the beginning of your experiment that requires the participant number to be entered. This is extremely important, as it is the only way that you will be able to match up the participant with their specific data.

Extracting Data and Building Reports

Each experiment or published course that you have created can be found and accessed on the “Publish Courses” screen. Active experiments are in green, and paused experiments are in red. Once you have clicked on your current experiment, it will provide you with the URL, number of attempts and last attempt time. The interface to interact with your published course/experiment can be seen in Figure 5. The information below refers to the cloud version of GIFT, however, the process is very similar on the desktop version of GIFT and uses the same tools.



Published Courses:

Example Experiment	Published status: Paused
TC3 Course 1 - Introduction	Published status: Active
TC 3 Course 3 - Phases of Care	Published status: Active
Test Course 234324	Published status: Active
Cognitive Psych 1	Published status: Paused
Testing123	Published status: Active
Condition 1	Published status: Active

Description
<No description available>

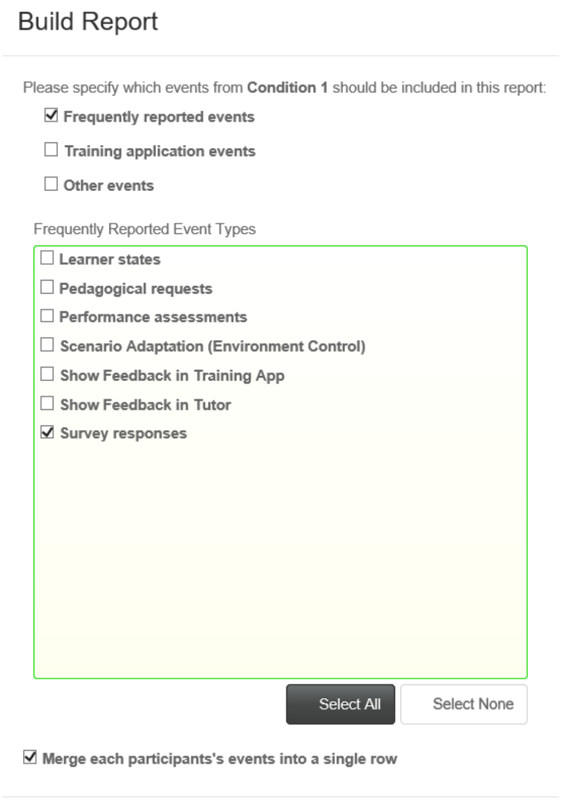
URL

Course
4ac95646-0cf4-470d-8d5c-98d2442b30e3\Logic Puzzle Tutorial

Subject Data
Number of Attempts: None
Latest Attempt: N/A

Figure 5. Published Course interface

Further, it has the options of “Pause”, “Export Raw Data” and “Pause and Build Report”. Each of these are very important functions. “Pause” allows you to temporarily stop collection of data; you might do this if you have collected data from the maximum number of participants that you need. “Export Raw Data” allows you to download the full GIFT log for each participant so that you can later extract it on the desktop version of GIFT. This function is most helpful if you used real-time assessment in a training application like PowerPoint. It is likely that for a straightforward experiment it will not be used. The most important option is "Pause and Build a Report". This allows you to extract data from the participant logs. If you have simply are interested in the input that participants gave in surveys, then you will select "Survey responses". In most cases it will be helpful to also click "Merge each participant's events into a single row". This will arrange it so that each participant is on a different line in the output spreadsheet, and that the question names are on the top of each column. Once you click create report, it will provide a report for you to download, which can then be opened in Excel to be saved as a spreadsheet, and subsequently SPSS if desired. At this point it is very important that you tagged each question with a name so that you can see which responses belonged to each question for later grading or coding. See Figure 6 for a screenshot of the "Build Report" screen with survey responses selected.



Build Report

Please specify which events from **Condition 1** should be included in this report:

☑ Frequently reported events

☐ Training application events

☐ Other events

Frequently Reported Event Types

☐ Learner states

☐ Pedagogical requests

☐ Performance assessments

☐ Scenario Adaptation (Environment Control)

☐ Show Feedback in Training App

☐ Show Feedback in Tutor

☑ Survey responses

☑ Merge each participants's events into a single row

Figure 6. Build a Report Screenshot

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The current paper provides a guide for research psychologists who would like to use GIFT to conduct traditional psychology experiments. Many changes have occurred over the years in GIFT, and one area that works, but is not perfected is participant management and the extraction/organization of data after conducting an experiment. Future research and development in GIFT can include updating the data extraction tools to provide more information to researchers, and clearer options. Additionally, while the current implementation of published courses allows for a question to be asked that can gather the participant number, it does not automatically link a number to the session that is visible to the experimenter. It would be helpful to have an alternate method of logging in that did not require creating a new online GIFT account. It would also be helpful to export surveys and questions in a manner that would allow researchers to match up the output questions with the responses that were provided by participants. This is particularly useful in cases when an experimenter may have forgotten to put a tag on the individual question and are unable to figure out what question was being responded to by the participant.

Even though there are still a number of improvements that could make GIFT more useful as a tool for conducting research, it has made great strides over the years. The authoring tool and survey system user interfaces have been greatly improved since the original version of this guide was created, and there are many more tools and features that are available to researchers. It is expected that GIFT will continue to develop and become an even more powerful tool for research psychologists and others to leverage to conduct experiments.

REFERENCES

Adams, D. (1979). *The hitchhiker's guide to the galaxy.* New York: Pocket Books.

Sinatra, A. M. (2014). The research psychologist's guide to GIFT. In *Proceedings of the 2nd Annual GIFT Users Symposium*, 85 – 92.

Sinatra, A. M. (2016). The Updated Research Psychologist's Guide to GIFT. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym4),* 135 – 144.

Sinatra, A.M. (2017). Utilizing intelligent tutoring systems and the Generalized Intelligent Framework for Tutoring (GIFT) for research. Proceedings of the *International Defense and Homeland Security Simulation Workshop 2017*, 18 – 24.

Sinatra, A. M., Sims, V. K., & Sottilare, R. A. (2014). *The Impact of Need for Cognition and Self-Reference on Tutoring a Deductive Reasoning Skill* (No. ARL-TR-6961). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring. org*.

ABOUT THE AUTHOR

Dr. Anne Sinatra *is part of the adaptive training research team within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab, she is lead on the Team Modeling vector and works on the Generalized Intelligent Framework for Tutoring (GIFT) project. Her background is in Human Factors and Cognitive Psychology.*

ACKNOWLEDGEMENTS

The research described herein has been sponsored by the U.S. Army Research Laboratory. The statements and opinions expressed in this article do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Using the Generalized Intelligent Framework for Tutoring (GIFT) to Support Adaptations in Challenge Levels for Collaborative Problem Solving in Digital and Virtual Reality Team Training Environments

Chris Meyer[1], Zach Heylmun[1], Mike Kalaf[1], & Lucy Woodman[1]
Synaptic Sparks[1]

INTRODUCTION

Team training practices have evolved in parallel with advancements in technology, most notably as advancements in computer technology allowed for the low-cost creation of immersive and realistic environments in which participants can train together in. Environments created with rich immersion, compelling stories, believable characters, and the ability to adapt scenarios to participants' levels of skill are prevalent throughout many industries, of note, the Military and Entertainment industries, and have allowed industries to begin training both individuals and teams in new and effective ways.

It is the Military and Entertainment industries that Synaptic Sparks, Inc. assists with bridging, and through a recent partnership with a local hi-tech escape room company has created a prototype research framework utilizing the Army Research Laboratory's Generalized Intelligent Framework for Tutoring (GIFT) software suite with technologically-advanced escape rooms. Each research effort, GIFT and the hi-tech escape rooms, stand alone in their respective disciplines. GIFT allows for unprecedented Intelligent Tutoring System (ITS) design, integration with external applications and sensors, and provides experimental frameworks with which to test Team Models and Adaptive Instructional Systems. And, the other party provides the only existing software framework compatible with military simulation paradigms and software interfaces into and out of an adaptable, virtual reality environment that currently tests both individuals and teams throughout their scenarios.

Together with these agencies, SSI is continuing to research and develop an integrated framework that uses the intelligent tutoring, adaptive learning, and experimental metrics aspects of the GIFT software suite to experiment and test with active players participating inside of adaptable difficulty scenarios. Grades of scenario-level and individual puzzle-level of challenges serve as the adaptable content implementations within the escape rooms and are resulting in experiment data usable for further research into how best to challenge groups and individuals as they progress through high tech training simulations and games.

INITIAL SCENARIO DESIGN AND GAME THEORY

GIFT requires fundamental constructs in order to be used in conjunction with a to-be-created team training system, namely a Learner (or Learners), Training Content, Adaptive/Remedial Content, a Training Goal, Sensors, Software/Instructor Controls (whether automated or not), and an objective Measure of Training Success.

These fundamental requirements are not necessarily defined only by current GIFT documentation, but a combination of Intelligent Tutoring System requirements, entertainment industry standards for group activities, current and/or upcoming military training readiness standards, and Synaptic Sparks partners' engineering knowledge.

When designing a team training scenario, retail and entertainment environments must follow a rigorous set of scenario design and game theory practices in order to create a positive customer experience. These goals are not unlike training goals, though the end result of satisfied customers is replaced by stringent sets of desired training outcomes in serious gaming domains.

The following general rules of game theory are adhered to by the design team to set the stage for a software suite such as GIFT to perform adequately (represented visually below in Figure 1):

- Entry criteria are established for the customers/learners
- Goals and evaluations are established (possibly unknown to learner)
- Individual and shared means of accomplishing goals are provided to/discovered by learners
- Systems of discovery, inputs, processes, and teamwork are iteratively repeated to satisfy learning goals
- Assessments are either performed throughout or at the conclusion of scenarios
- Conclusions and debriefing (after action reviews) are performed to fully enforce training goals

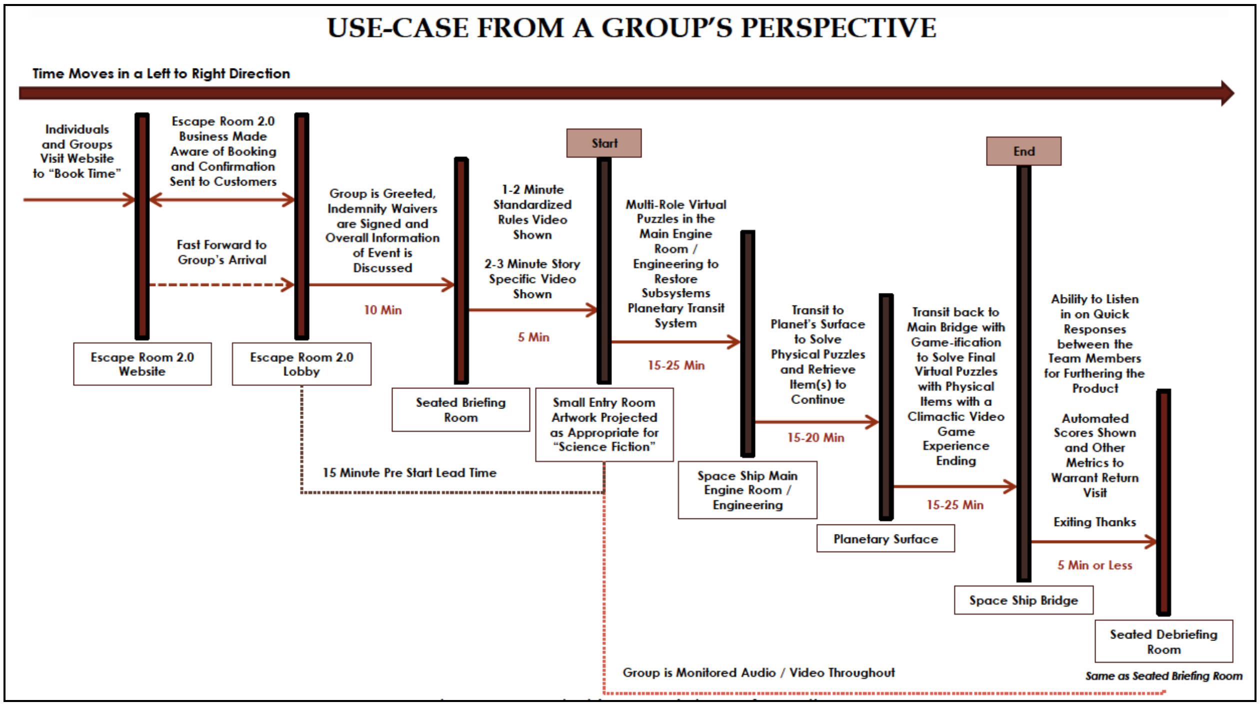

USE-CASE FROM A GROUP'S PERSPECTIVE
Time Moves in a Left to Right Direction
Individuals and Groups Visit Website to "Book Time"
Escape Room 2.0 Business Made Aware of Booking and Confirmation Sent to Customers
Fast Forward to Group's Arrival
Escape Room 2.0 Website
Escape Room 2.0 Lobby
Group is Greeted, Indemnity Waivers are Signed and Overall Information of Event is Discussed
10 Min
Seated Briefing Room
1-2 Minute Standardized Rules Video Shown
2-3 Minute Story Specific Video Shown
5 Min
Start
Small Entry Room Artwork Projected as Appropriate for "Science Fiction"
15 Minute Pre Start Lead Time
Multi-Role Virtual Puzzles in the Main Engine Room / Engineering to Restore Subsystems Planetary Transit System
15-25 Min
Space Ship Main Engine Room / Engineering
Transit to Planet's Surface to Solve Physical Puzzles and Retrieve Item(s) to Continue
15-20 Min
Planetary Surface
Transit back to Main Bridge with Game-ification to Solve Final Virtual Puzzles with Physical Items with a Climactic Video Game Experience Ending
15-25 Min
End
Space Ship Bridge
Ability to Listen in on Quick Responses between the Team Members for Furthering the Product
Automated Scores Shown and Other Metrics to Warrant Return Visit
Exiting Thanks
5 Min or Less
Seated Debriefing Room
Same as Seated Briefing Room
Group is Monitored Audio / Video Throughout

Figure 36 – Escape Room (Team Training) Use-Case Flowchart

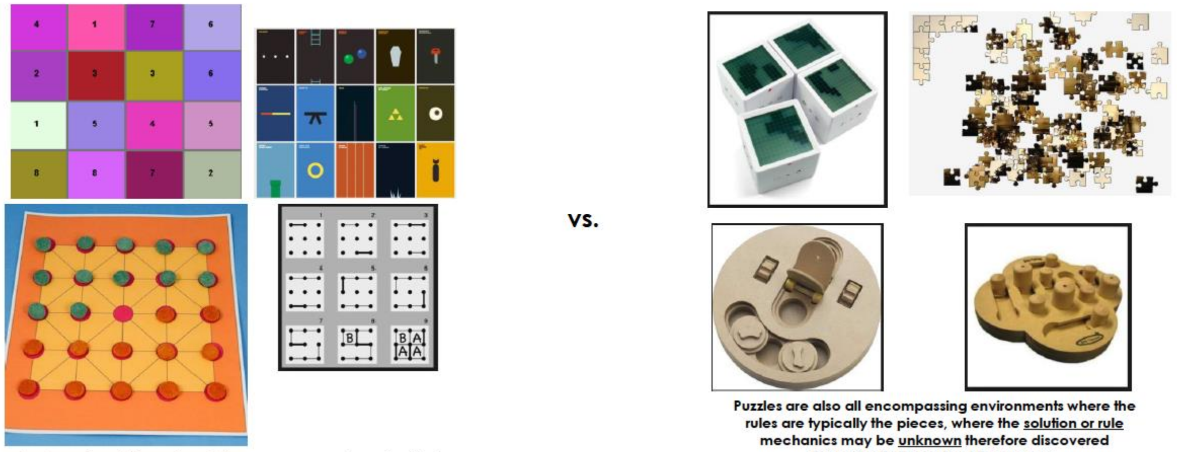
With these similarities between retail entertainment scenarios in a hi-tech, focused, results-oriented environment and serious training environments established, Synaptic Sparks moved to include the GIFT software suite into adaptive challenge determinations based on both individual and team performance throughout a scenario's operational time limits.

GAMES VS. PUZZLES (VIDEO GAMES VS. SERIOUS GAMES): INDIVIDUAL, SMALL TEAM, AND LARGE TEAM PERFORMANCE

With overarching User Stories (and Training Goals) for scenarios defined by a partner company, the Synaptic Sparks team then began to analyze similarities between GIFT monitoring and assessment capabilities and software analytics, difficulty settings, and operator interactions present within the hi-tech escape rooms.

Of first note was the module nature of GIFT training content when compared to escape room (training scenario) puzzles. Unlike traditional games, the hi-tech escape room puzzles maintained key similarities with training content, visually represented in Figure 2 below, namely:

- Victory conditions were not always a 1 or a 0
- Time bounds were nearly always a factor for success
- Individuals could assist or hinder team performance depending on their actions
- Team performance is more than a sum of the parts of individual performance
- Populations of teams exhibited standard deviations of performance (80% of all future performance is within bounds based on past team performance with a data set of significant enough size)
- Discovery of puzzle rules and goals, no matter the proficiency of the learner, is more important to the training and learning process when considering success and knowledge retention than being told the rules and simply executing them given limited training time and first-time exposure to a challenge



For the sake of discussion mini-games are usually embedded Video Games within a larger Video Game.

However; in this context DEV defines a game or mini-game (via Game Theory) as an all encompassing environment where known rules exist to meet a self-serving objective – to "Win"

Puzzles are also all encompassing environments where the rules are typically the pieces, where the solution or rule mechanics may be unknown therefore discovered dynamically throughout the process.

Puzzles also employ a self-serving objective – to "Solve"

Figure 37 - Games Vs. Puzzles: A Visual Representation of Potential Negative Learning Vs. Critical Thinking Development, Concerning Real-Life Tasks

Elements of Social Utility – Encouraging Team Performance and Monitoring Success

When taking into account the similarity of effective training content to correctly-designed puzzles in a hi-tech training scenario, Synaptic Sparks personnel then analyzed customer data to best-define what social utility elements GIFT may be able to use as it evolves to intelligent tutoring adaptations for teams as opposed to individuals.

To generalize and as represented in Figure 3 below, when SSI observed roughly 5,000 participants in groups of up to 6 individuals per team, the **perceived** Social Utility of the team experience was most-relevant to customer satisfaction and team performance, and also learning retention.

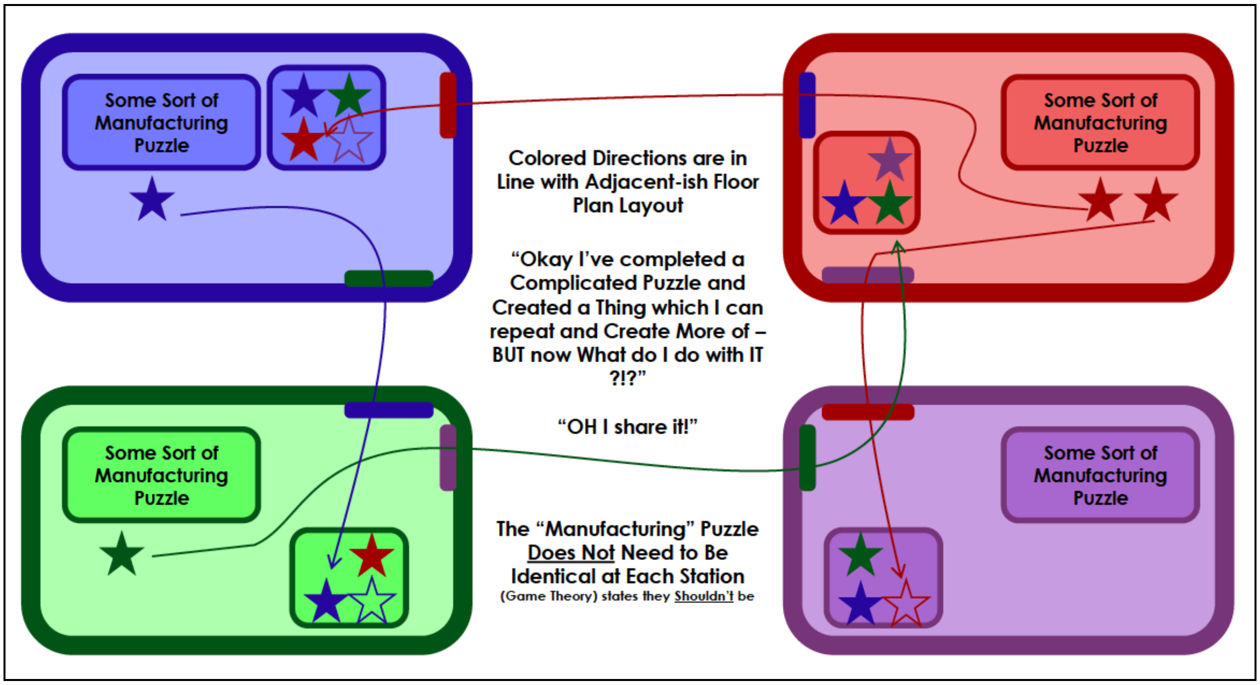

Some Sort of Manufacturing Puzzle
Some Sort of Manufacturing Puzzle
Some Sort of Manufacturing Puzzle
Some Sort of Manufacturing Puzzle
Colored Directions are in Line with Adjacent-ish Floor Plan Layout
"Okay I've completed a Complicated Puzzle and Created a Thing which I can repeat and Create More of – BUT now What do I do with IT ?!?"
"OH I share it!"
The "Manufacturing" Puzzle Does Not Need to Be Identical at Each Station
(Game Theory) states they Shouldn't be

Figure 38 - A Conceptual Diagram of Social Utility for Training Content

During all escape room scenarios, Synaptic Sparks was given license to add software "hooks" into the escape room experience, and direct all metrics to a GIFT software suite monitoring station via methods explained, in-part, in the following section.

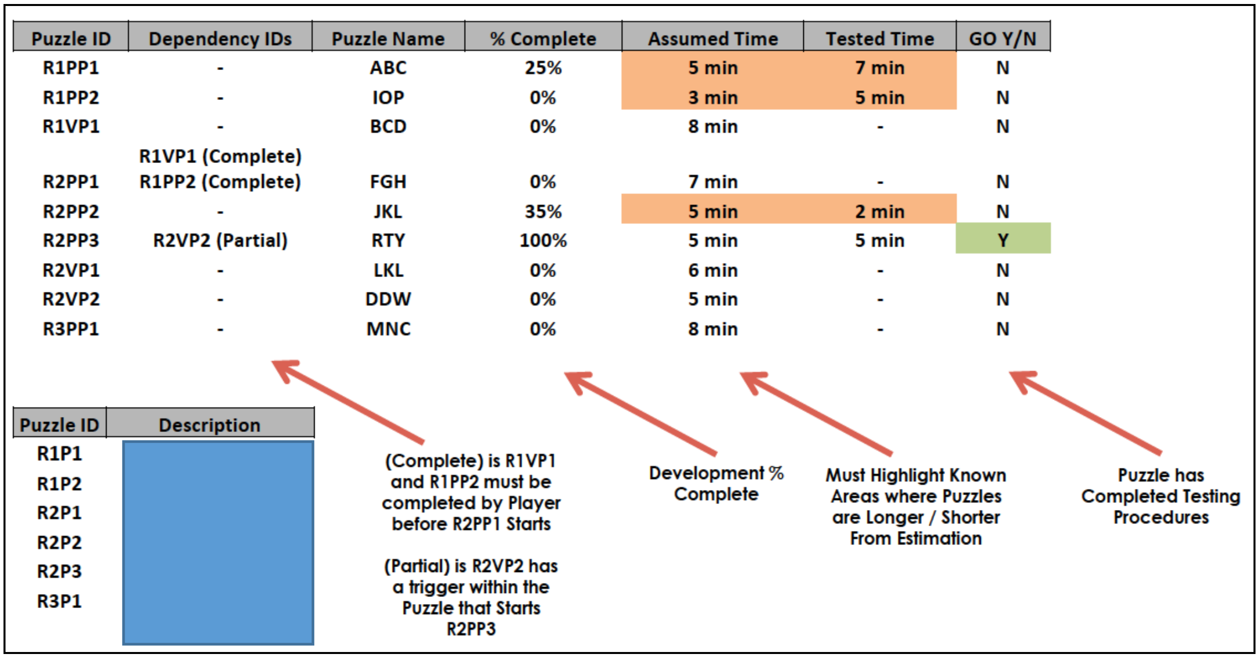
The analytics resulting from the metrics analysis gave new insight into individual, team, and overall performance for groups, and how GIFT can be used to adapt puzzle (and therefore training) content to better serve participants.

UTILIZING THE GIFT SOFTWARE SUITE TO ENHANCE CONTENT ADAPTATION WITHIN A TEAM TRAINING SCENARIO

Synaptic Sparks identified some existing feedback loops between GIFT and the escape room software to create low-cost, high-value experiments. SSI created:

- A GIFT-Monitored Statistic Set Consisting of Elapsed Time Based on Puzzle Start and Completion Time to Adapt Early/Late Puzzle Exposure to Team Performance

- A Failsafe “Easy Mode” Monitoring System to Skip Puzzles for “Overwhelmed” Participants
- Messaging Between Escape Room Manager and Subcomponents, and GIFT (Figure 5)



Puzzle ID	Dependency IDs	Puzzle Name	% Complete	Assumed Time	Tested Time	GO Y/N
R1PP1	-	ABC	25%	5 min	7 min	N
R1PP2	-	IOP	0%	3 min	5 min	N
R1VP1	-	BCD	0%	8 min	-	N
R2PP1	R1VP1 (Complete) R1PP2 (Complete)	FGH	0%	7 min	-	N
R2PP2	-	JKL	35%	5 min	2 min	N
R2PP3	R2VP2 (Partial)	RTY	100%	5 min	5 min	Y
R2VP1	-	LKL	0%	6 min	-	N
R2VP2	-	DDW	0%	5 min	-	N
R3PP1	-	MNC	0%	8 min	-	N

Puzzle ID	Description
R1P1	
R1P2	
R2P1	
R2P2	
R2P3	
R3P1	

(Complete) is R1VP1 and R1PP2 must be completed by Player before R2PP1 Starts

(Partial) is R2VP2 has a trigger within the Puzzle that Starts R2PP3

Development % Complete

Must Highlight Known Areas where Puzzles are Longer / Shorter From Estimation

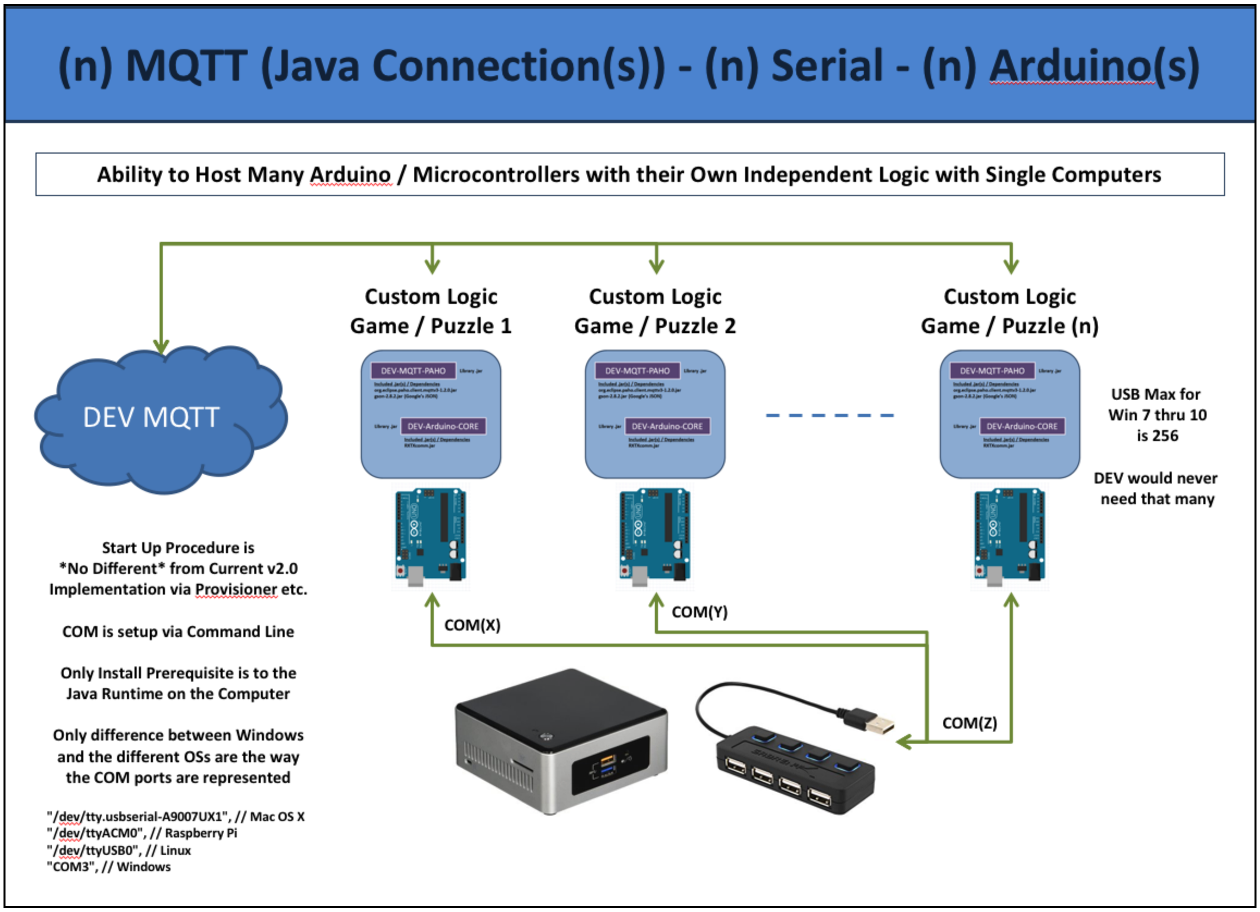
Puzzle has Completed Testing Procedures

Figure 39 - Metrics Visualization for Estimated Vs. Actual Team Performance Monitored by GIFT

These systems allowed the escape room’s software to utilize, as the private company sees fit, GIFT recommendations on content adaption concerning two main elements of the experience; namely the time at which a new puzzle is given to a team, and the difficulty level of a puzzle selected from Easy, Medium, or Hard levels.

Communications between scenarios and the GIFT software suite were established through message broadcasts, broadly defined and represented in Figure 5 below.

The results of these experiments are still being compiled, but a fully autonomous system that can adapt to any team performance is expected by the end of 2Q 2018.



(n) MQTT (Java Connection(s)) - (n) Serial - (n) Arduino(s)

Ability to Host Many Arduino / Microcontrollers with their Own Independent Logic with Single Computers

DEV MQTT
Custom Logic Game / Puzzle 1
Custom Logic Game / Puzzle 2
Custom Logic Game / Puzzle (n)
DEV-MQTT-PAHO
Library .jar
Included Jar(s) / Dependencies
org.eclipse.paho.client.mqttv3-1.2.0.jar
gson-2.8.2.jar (Google's JSON)
Library .jar
DEV-Arduino-CORE
Included Jar(s) / Dependencies
RXTXcomm.jar
COM(X)
COM(Y)
COM(Z)

USB Max for Win 7 thru 10 is 256

DEV would never need that many

Start Up Procedure is *No Different* from Current v2.0 Implementation via Provisioner etc.

COM is setup via Command Line

Only Install Prerequisite is to the Java Runtime on the Computer

Only difference between Windows and the different OSs are the way the COM ports are represented

```
"/dev/tty.usbserial-A9007UX1", // Mac OS X
"/dev/ttyACM0", // Raspberry Pi
"/dev/ttyUSB0", // Linux
"COM3", // Windows
```

Figure 40 - A Sample Messaging Scheme between Puzzles in Hi-Tech Escape Room Sub-System and GIFT

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Currently, the SSI team is performing the same level of integrations as described above for an even more modernized, fully-digital escape room experience in Virtual Reality. While scenarios described above are currently being adapted for various military training scenarios that have a “real world” environment with which GIFT can interact, the new Virtual Reality project is limited only by the art asset content and training knowledge of Subject Matter Experts.

With GIFT operating alongside the hi-tech escape room scenario manager, adaptive content can be served and adapted not only to individuals, but teams as they collaborate together in a proprietary virtual environment.

While the SSI team consists primarily of experienced engineers, this new experimental platform is provided to the ARL GIFT team researchers to further evolve and experiment with as research representatives see fit.

ABOUT THE AUTHORS

***Christopher Meyer** leads the Synaptic Sparks team as the President and CEO of the Board, and holds a B.S. and M.S. in Computer Science obtained from Kansas State University, with minors in Economics and Modern Languages. Chris specialized in Artificial Intelligence studies in Chukyo University, Japan, and has 15 years of experience in business and engineering.*

***Zach Heylmun** holds the Chief Scientist position in Synaptic Sparks, and holds a B.S in Computer Science obtained from Florida State University. Zach specializes in software optimizations and graphics processing after 5 years of professional experience.*

***Mike Kalaf** is the Chief Operating Officer in SSI, and has over 30 years of Modeling, Simulation and Training leading large scale efforts leveraging cutting edge technology. Mike also serves on several non-profit boards dedicated to STEM outreach in the Orlando, FL locale.*

***Lucy Woodman** is currently studying to graduate with a B.S. in Information Technology and assisting SSI in her early career. Lucy specializes in System Administration, Computer Security, and Programming.*

Basic Robotic Course

Danielle Julian
University of Central Florida

INTRODUCTION

The prevalence of robot-assisted laparoscopic surgery (RALS) within both military and civilian hospitals has been steadily increasing in recent years, reaching a total of over 2 million cases worldwide to date (www.intuitivesurgical.com), generating a need for effective training of the unique skills and technological knowledge required to perform such technologically advanced procedures. Of particular importance is the initial learning curve associated with acquisition of these skills by inexperienced surgeons, which has numerous implications, particularly in terms of patient safety (Hopper, Jamison, & Lewis, 2007). The acquisition of purely cognitive skills has been studied extensively, revealing learning curves that typically involve three distinct stages of skill development: a cognitive stage, an associative stage, and an autonomous stage during which expertise is achieved (Anderson et al., 1997).

The purpose of this intelligent tutoring system (ITS) is to help train physicians both the cognitive and basic knowledge of skills needed to use the most commonly known robotic surgical system, the da Vinci. This system could be used to bridge the training gap between online cognitive training materials and hands-on psychomotor skills training with simulators and robots. The ITS could provide novice and intermediate robotic surgeons with intelligent guidance in an easily accessible system to train the cognitive process and procedural steps behind fundamental robotic surgery skills.

The tutor will include cognitive material covering an introduction to surgical robotics, introduction into the da Vinci Surgical System, basics on camera control, and interrupted suturing. This ITS will be developed using the gift framework of tools and provided as a web-based course. The content for the system was collected from multiple practicing robotic surgeons who performed each tutored task using a simulator and explaining their actions, reasoning, and potential mistakes as they performed each exercise. This information was captured as video, instruction sets, and flow charts, which were reviewed for accuracy by surgeons and then used as training content within the GIFT framework. GIFT houses several modules of RALS content that interact with each other to tailor content to learners attributes and provide the content via a Computer-Bases Tutoring System (CBTS).

RELATED RESEARCH

Individualized training has been shown to be highly effective (e.g., one on one instruction) because it allows trainees to receive expert feedback, targeting the skills most in need of acquisition or remediation. However, this form of training is costly in terms of expert time, and therefore limited. This issue has been addressed within the education and military domains through the use of ITSs, which consist of advanced training software that mimics a human tutor by adapting instructional content and feedback to an individual student. An ITS capable of supporting acquisition of the cognitive, perceptual, and psychomotor skills associated with RALS could greatly reduce the associated learning curve and improve patient safety.

Beyond a one-day individualized training, RALS surgeons typically overcome the learning curve in an experiential way. Surgical trainees may encounter their first surgical experience on an inanimate training model, excised tissue, or an actual procedure with a mentor. While this method helps to improve performance with increased experience, these procedures usually take more time to complete and are associated with a greater number or errors, which may be life threatening to the patient. More recently, Virtual Reality (VR) surgical simulators have been introduced to help alleviate this issue. VR simulation was first introduced to surgical education in the late 1980s (Satava, 1993). Since implementation, VR simulators have been established as a valuable training tool for the acquisition of basic surgical skills, allowing a trainee to safely overcome the learning curve associated with new techniques while providing independent and repetitive exposure in a safe and cost-efficient environment (Chou & Handa, 2006). The application of VR simulators in surgery has proven to be essential with the development and implementation of new technology and complex devices. However, these trainers can be expensive and are typically not portable which cause issues for practicing surgeons.

ITS's have been shown to be particularly valuable for teaching complex cognitive tasks such as trouble shooting, problem solving, and resolving critical situations. As a human tutor does, an ITS continually monitors and assesses the individual student's actions, infers the student's state of knowledge, and decides on the next instructional event to maximize the student's learning based on an embedded student model, expert model, and domain model (Perez et al., in press). As highlighted by a recent meta-analysis (Kulik & Fletcher, 2016), research and development within the domain of ITS's has demonstrated the technical feasibility and relative effectiveness of computer-based adaptive instruction as compared to classroom and small group instruction. ITS development has been applied across multiple domains, including within military applications such as ship handling and tactical decision-making. Furthermore, previous development efforts have demonstrated the ability to effectively apply generic ITS components such as authoring tools to specific military domains (Stottler, Fu, Ramachandran, & Jackson, 2001; Sottilare & Holden, 2013). Offering a portable, customized, repeatable tutoring system for RALS would be highly beneficial.

Medical ITS's

Most of the literature on medical ITS use a pedagogically approach to train knowledge-based medicine (Crowley et al., 2007) and more recently aid in imaging recognition. One of the earliest medical ITS's, GUIDON, trained medical students about infectious diseases like meningitis and bacteremia. The objectives were to identify likely causative organisms given a patient's history, medical records, and laboratory results (Clancey, 1988). It used an interactive mixed-initiative method of dialogue where either the student or the system could be in control of how the discussion played out (Clancey, 1988; Crowley et al., 2007). Another tutor, MR Tutor, is a case-based tutoring system. This system focuses on training case similarities across patient instances. This system uses a library of radiologic images where the tutor uses statistical indices to find similarities across the collection (Sharples et al., 2000).

More recently there have been several tutors developed to train on specific diseases, including diabetes's and stomach disease (Almurshidi & Naser, 2017; Almurshidi & Naser, 2017b). Almurshidi and Naser's latest tutor aims to train medical students about multiple stomach diseases. This tutor allows the learner to navigate through the domains of concepts with knowledge checks within. If the student scores a 75% or higher, they may move to the next level of difficulty, if not, they return to repeat the same set of exercises/content review. This method or recall and rehearsal provide repeat exposure to students that have yet to master the knowledge or skills.

The RALS domain represents a complex task environment involving cognitive, perceptual, and psychomotor skill components; which could greatly benefit from real-time assessment and adaptive instruction capabilities. Integration of ITS components into a RALS based course could support a reduction in both self-guided and instructor-led training, as well as a reduction in the initial learning curve observed in the first cases completed by novice surgeons, directly benefiting patient outcomes. In addition, there is a need for effective and standardized curricula and testing devices for training robotic surgeons, providing a more standardized form of guidance to all students and all learning facilities. In addition to initial acquisition training, such a curriculum could be applied to the refresher training learning curve that occurs after periods of nonuse.

ITS DESIGN AND STRUCTURE

The original design of the Robotic Suturing ITS was aimed to train surgeons the cognitive, procedural, and psychomotor skills associated with two basic robotic tasks (i.e., suturing and camera control). However, the development of such a tutor is outside the scope of this project. This tutor is now structure to provide the following:

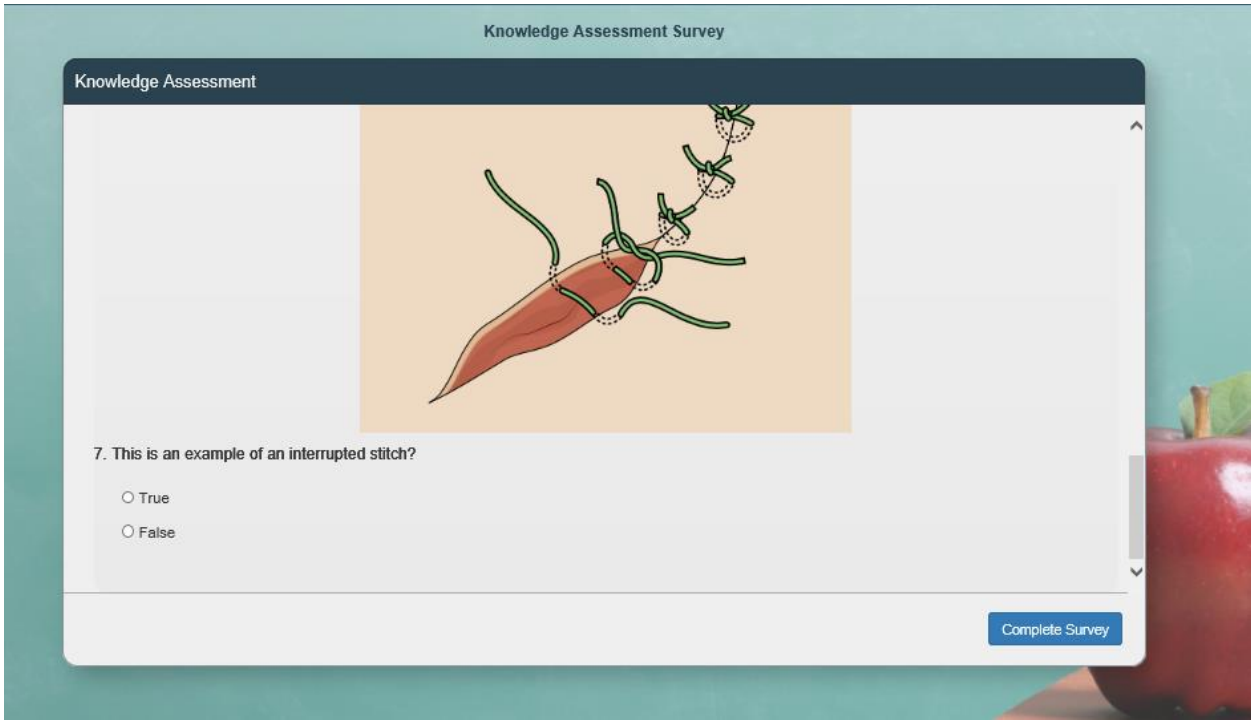
1. Introductory information on surgical robots
2. Technical details on the da Vinci surgical system
3. Basic camera control knowledge
4. Limited basics on suturing with the da Vinci system.

This iteration of the system uses a mastery learning technique to ensure the learner has satisfactory recall and can apply perquisite knowledge before proceeding to the next concept to be covered.

Opening Assessments

Before the course begins, the learners will be asked to complete a demographic survey. This survey is used to collect information about the learners, their specialty, and experience level. This iteration of the system used the demographic survey to collect information only and is a non-actionable questionnaire. As other iterations are developed, the demographic questionnaire may be used as an actionable survey that could in return affect the flow of the course content. For example, if the learner has selected otolaryngologist (i.e., Ear, Nose, and Throat surgeon) as their specialty the course environment can take action and change to provide material for this specialty. In this case, ENT surgery requires little suturing and more energy application, so the suturing content will not be as imperative to this student as others.

Due to the ambiguous training associated with surgical robotic programs, the course will then provide a mandatory actionable knowledge assessment (Figure 1). This assessment is used to measure the learner's prior knowledge on the course objectives. At least one question from each course concept is covered in this assessment.



Knowledge Assessment Survey

Knowledge Assessment

7. This is an example of an interrupted stitch?

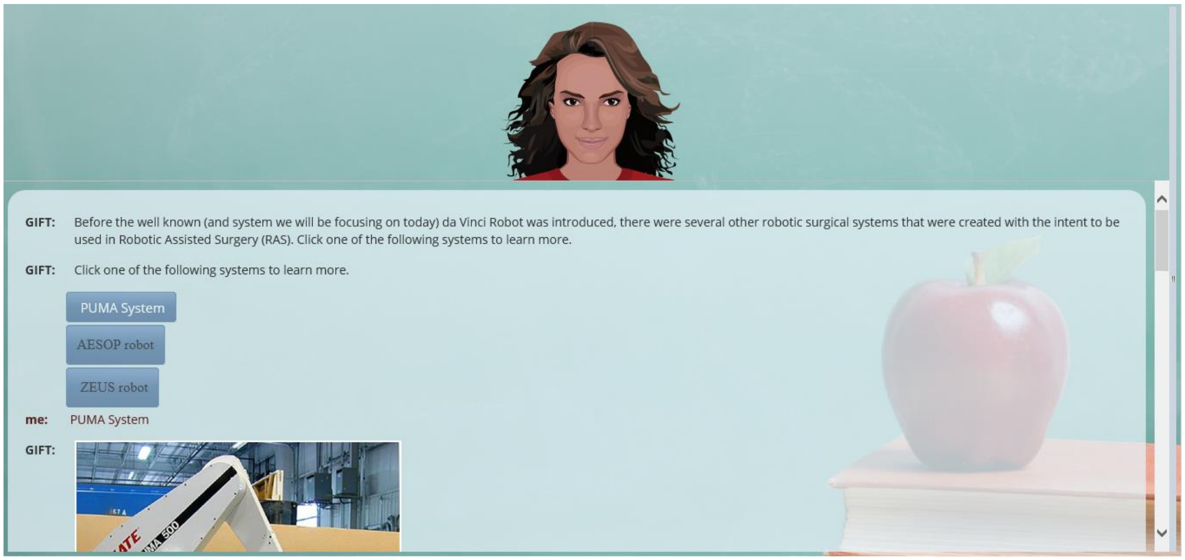
○ True

○ False

Figure 1. Knowledge Assessment Survey Sample

Course Material

As mentioned previously, the da Vinci system now provides a piece of technology that most practicing surgeons will not have any experience with. Before training the skillset to overcome this learning curve, the learners should be familiar with robotic surgery in general. The course starts with basic text explaining the difference between traditional minimally invasive surgery (MIS) and RALS. This portion will be a short mandatory object of the course. The learner will then be presented with a basic overview and history review of the introduction of robotic surgery and how the da Vinci system was brought to fruition. Because the history of surgical robotics is extensive, a conversation tree was selected to help train this material and maintain learner engagement. The conversation tree used looping pathways. The student selects which early robotic system they would like to learn about, then must select another off of the list, eventually moving their way to the end of the tree (Figure 2).



GIFT: Before the well known (and system we will be focusing on today) da Vinci Robot was introduced, there were several other robotic surgical systems that were created with the intent to be used in Robotic Assisted Surgery (RAS). Click one of the following systems to learn more.

GIFT: Click one of the following systems to learn more.

me: PUMA System

GIFT:

Figure 2. Conversation Tree sample.

The main attributes of the learner were based on knowledge checks. If the learner's knowledge were classified as "Novice", the content was more engaging and showed diagrams/pictures. If the learners were classified as "Experts" the course adapted to show more concise textual content. Student's knowledge was the main attribute driving the course flow and content. Figure 3 shows the course layout and flow.

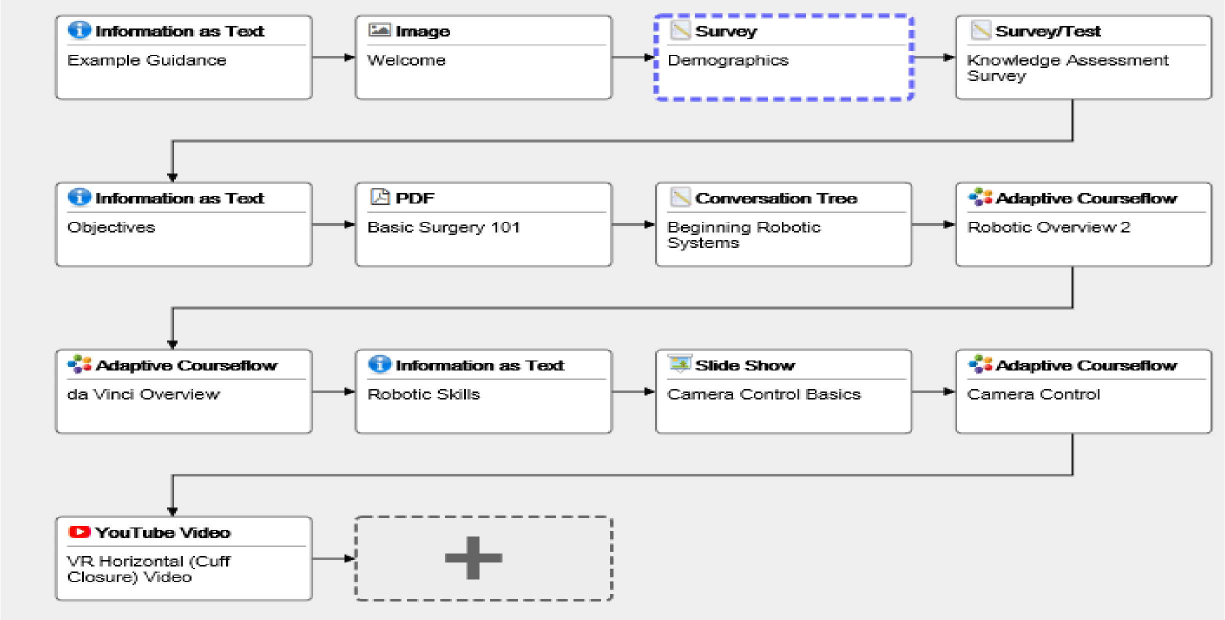


Figure 3. Course Flow

If the student does not do well on knowledge checks via a short questionnaire selected from a larger course question bank, then the student is provided with a more extensive version of the particular concepts content.

The next learning objective is aimed to train basic technical knowledge needed to use the actual da Vinci Surgical System. The answers collected from the knowledge assessment will drive the content for this section of the course as well. If the student scores poorly in the introduction assessment they will be provided with a detailed, but more engaging (e.g., video overview) content delivery. Consequently, if they learner does well on the opening assessment the tutor will provide traditional textual content as review content. After the initial mandatory content, the learners will be provided with a short questionnaire to gage their knowledge, leading back to the original or differing training material if they scored poorly or moving them to the next concept if they scored well. This adaptive course flow helps to provide tailored content specific to the learners existing and acquired knowledge.

To break up some of the textual content a slide show covers the basics of the next objective, Camera Control. The slide show provides images and text to help maintain learner engagement and provides basic technical information regarding the scope for the da Vinci system. The adaptive course flow for this concept mirrors the flow for the basic robotics and da Vinci Surgical System concepts. The camera control adaptive course flow and the suturing content will need further development in order to achieve training psychomotor skills. Figure 4 shows an example of the course content for Camera Control.

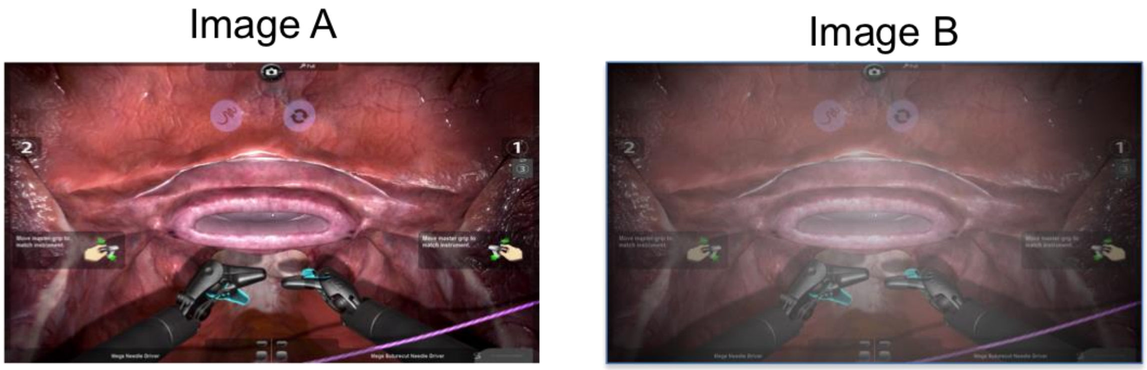


Figure 4. Example of Camera Control content.

Before the tutor moves into attempting to train the psychomotor suturing skill set, a simulated video will play for all students. This video shows a simulated vaginal cuff closure completed robotically (with a robotic surgical simulator). This is a common robotic assisted procedure. This procedure was chosen as course material because it requires camera movement and control and requires the surgeon to complete a cuff closure using an interrupted stitch. This stitch is common, but difficult for novice robotic surgeons. A video was selected to provide the learner with an all-encompassing example of what the tutor is aiming to help train. That is, the video shows why a technical overview of the system is imperative, the importance of camera control, and how an adequate interrupted suture in complete.

DISCUSSION AND RECOMMENDATIONS

GIFT as the authoring tool for developing the cognitive concepts associated with robotic surgery was user friendly. The drag and drop concept was beneficial for course flow planning and content development. However, future iterations of the robotic tutor (including more psychomotor training) will be difficult for a developer with little programming experience or limited GIFT experience. For challenging content the developer must be well versed with this authoring tool. Choosing such a complex topic to train is difficult within this system (and potentially others) because of the psychomotor, procedural, and variance of surgical specialties. For example, suturing for a general robotic surgeon will differ from suturing for a gynecological surgeon.

There are several aspects of GIFT that weren't clearly defined during my experience. For example, the question bank concept. After creating the tutor, I know understand that the question bank in used within the adaptive course flow concept, but why would the question bank be its own concept? As I am sure this has a used within authoring, it was unclear during this development. Creating a more extensive user guide on how to use, when to use, and why to use each concept would be highly beneficial.

Developing a more simplistic, or just one portion of this course, is ideal for novice GIFT users. The authoring tool offered multiple media outlets, supports a substantial amount of content, but a novice user may not know how to integrate any "bells and whistles." The robotic tutoring system could have benefited from including highlighting clues during assessments or interaction within a video. GIFT may be capable of such features but was not easily defined on how to implement these high level capabilities.

While the initial scope of the project was aimed to provide step-by-step instruction for completing an interrupted suture, this was unmet during the scope of this project for two reasons. Reason number one, in order to train a psychomotor and procedural skill set, the developer will need additional time working with GIFT and may require an additional developer with differing credentials. Reason number two, the content for the suturing aspect of the course must be created using surgical imagery or "do's and don'ts" of robotic suturing to provide appropriate content.

REFERENCES

Almurshidi, S. H., & Naser, S. S. A. (2017). Design and Development of Diabetes Intelligent Tutoring System.

Almurshidi, S. H., & Naser, S. S. A. (2017). Stomach disease intelligent tutoring system.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of experimental psychology: learning, memory, and cognition*, *23*(4), 932.

Chou, B., & Handa, V. L. (2006). Simulators and virtual reality in surgical education. *Obstetrics and gynecology clinics of North America*, *33*(2), 283-296.

Clancey, W. J. (1988). Knowledge-based tutoring: the GUIDON program.

Crowley, R. S., Legowski, E., Medvedeva, O., Tseytlin, E., Roh, E., & Jukic, D. (2007). Evaluation of an intelligent tutoring system in pathology: Effects of external representation on performance gains, metacognition, and acceptance. *Journal of the American Medical Informatics Association*, *14*(2), 182-190.

Hopper, A. N., Jamison, M. H., & Lewis, W. G. (2007). Learning curves in surgical practice. *Postgraduate medical journal*, *83*(986), 777-779.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42-78.

Satava, R. M. (1993). Virtual reality surgical simulator. *Surgical endoscopy*, *7*(3), 203-205.

Sharples, M., Jeffery, N. P., du Boulay, B., Teather, B. A., Teather, D., & du Boulay, G. H. (2000). Structured computer-based training in the interpretation of neuroradiological images. *International Journal of Medical Informatics, 60*(3), 263-280.

Sottilare, R. A., & Holden, H. K. (2013, July). Motivations for a generalized intelligent framework for tutoring (GIFT) for authoring, instruction and analysis. In *AIED 2013 Workshops Proceedings* (Vol. 7, p. 1).

Stottler, D., Fu, D., Ramachandran, S., & Jackson, T. (2001). Applying a generic intelligent tutoring system authoring tool to specific military domains tao its, a case study Paper presented at the ITSEC.

THEME VII: ANALYTICS AND EFFECTIVENESS MEASURES

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

J. T. Folsom-Kovarik[1], Michael W. Boyce[2], Robert H. Thomson[3]
[1] Soar Technology, Inc., [2] U. S. Army Research Laboratory, [3] United States Military Academy

INTRODUCTION

Authoring adaptive training can present challenges because instructors, unit leaders, and other non-technical users need to understand and control adaptation in order to accept and make use of a training system such as GIFT. Therefore, adaptation should be presented in a manner that parallels the way these end users think about instruction (Wray, Folsom-Kovarik, Woods, & Jones, 2015). This work enabled future improvements in authoring for adaptation by adding several constructs inside GIFT. First, *patterns* added a new construct for defining learner behaviors and analytics that can drive adaptation. Second, *misconceptions* added information to GIFT concepts in the Learner Module about reasons that individuals might be performing Below Expectation. Third, *mid-lesson reports* tested a specific type of adaptive intervention that prompts learner reflection during training, with reduced authoring via reusable prompts.

A randomized controlled trial was conducted to evaluate the training effectiveness of GIFT when driving adaptive feedback in a newly integrated tool for perceptual and cognitive skills relevant to cross-cultural communication. The combination of GIFT plus the skill training was evaluated by a population of 74 West Point Cadets. A preliminary analysis supported the value of the patterns to identify different classes of learner experience and, in future, to let non-technical personnel define what high-level behaviors and groups of observations would help GIFT respond to these. The analysis also suggested new domain-general misconceptions that might be able to inform adaptation. The evaluation showed an improvement between pre-test and post-test scores across all users. The discovery of new patterns and misconceptions highlights opportunities for instructors or unit leaders to gather evidence about how training is progressing in GIFT and, with future incorporation into the GIFT authoring suite, to quickly add new adaptive interventions that make training more effective.

Cross-cultural Communication and Perceptual-cognitive Skill Training

The proof of concept was demonstrated in a cross-cultural communication training domain. The laptop-based training consisted of four narrative scenarios that challenged learners to make decisions based on perceptual cues such as facial expressions in an environment with simulated characters. The existing training contained delayed feedback in the form of after-action review and the ability to optionally replay each scenario and try different choices. GIFT was used to add adaptive feedback to the existing training system – mid-lesson reports (see below) that were triggered by misconceptions GIFT inferred based on learner interactions. The mid-lesson reports overlaid immediate feedback onto the existing training system via the onscreen GIFT Tutor User Interface (TUI). Combining GIFT with existing training demonstrated how GIFT integration gives the potential to make a system more adaptive to learner needs by adding the tools of patterns, misconceptions, and mid-lesson reports.

We briefly describe the training content for cross-cultural communication. The scenarios and subject-matter tests in this experiment were structured around a simplified version of the Good Stranger approach to cross-cultural communication (Klein, Moon, & Hoffman, 2006). This approach is intended to work independently of a specific culture. It trains learners to perceive, understand, and work within any foreign or unfamiliar culture. The training has previously been used with success in a military setting (Hubal, van Lent, Marinier, Kawatsu, & Bechtel, 2015).

The simplified content for this experiment targeted four learning objectives:

- Initiate and engage in encounters that support the mission and build rapport
- Practice perspective-taking to make sense of encounters
- De-escalate conflicts and repair relationships
- Balance tact and tactics to achieve long-term goals in a safe manner.

Each learning objective offered opportunities during the scenarios and tests for learners to choose either Good Stranger behavior or behavior associated with misconceptions that were selected to be general and possible to express in any of the four learning objectives (see below).

Patterns

Patterns and misconceptions have the potential to add to the authoring experience in GIFT. Patterns enhance the language for expressing constraints on learner performance in the GIFT Domain Knowledge File (DKF). Patterns describe a library of reusable conditions and groups of conditions that non-technical personnel can draw on to relate domain-specific observations without requiring engineering skill to create new conditions in source code. For example, different adaptive feedback could be delivered when domain-specific conditions occur repeatedly, or close together.

A key characteristic of patterns is that they operate not on domain messages, but on conditions. Patterns let end users describe how conditions should relate to each other. So, any conditions that have already been defined by writing and compiling source code can then become part of a pattern that end users control via a future, user-friendly authoring tool. The research has previously defined an initial list of observable patterns that relate multiple performance observations together in ways instructors are likely to use (Folsom-Kovarik & Boyce, 2017). Examples include doing tasks in order or out of order, doing actions too often or too few times, and taking too long or too little time to do an action (a pattern that generalizes the GIFT slide-underdwell condition to check the timing for any pair of conditions).

This experiment led to discovery of several patterns that may be of value to instructors, which are described in the Discussion section. In advance of the experiment, the following patterns were defined:

- Hesitation: change any answer two or more times before submitting
- Hurrying: submit any answer within five seconds of a choice presentation
- Improving: exhibit Good Stranger behavior twice with no intervening mistake. This pattern enabled a positive mid-lesson report, rather than silence, when learners did well.

Misconceptions

Misconceptions enhance the GIFT learner model with additional information about estimates of mastery. Misconceptions can express not just lack of mastery but specific reasons that may underlie any incorrect or unwanted behavior that GIFT observes. By defining patterns that indicate misconceptions and adaptive feedback specific to certain misconceptions, it is possible to control in detail the feedback that GIFT adds to teaching and training.

Four misconceptions were defined for this experiment:

- Cautious: the learner is overly deferential or sacrifices a key goal
- Authoritarian: the learner is overly concerned with being respected or obeyed
- Mission-focused: the learner achieves a near-term mission at a high cost to relationships
- Rules-focused: the learner follows rules too inflexibly

The misconceptions were designed to test the reusability of the misconception idea. The same misconceptions could be expressed in all learning objectives through different learner choices or behaviors for this experiment. As a result, the GIFT pedagogical module only had to understand a single set of rules about misconceptions across scenarios. The default pedagogical module was enhanced to evaluate all misconceptions as they were inferred by conditions and patterns. For generality, the misconceptions were assigned two-dimensional values defining their importance and urgency. As a result, the enhanced pedagogical module could use domain-general rules to address the most important or urgent misconceptions first. It did not need to know about the domain-specific contents of the four misconceptions.

The reusability of these specific misconception definitions was initially limited to the cultural communication domain. In future work, it may also be possible to create misconceptions that are domain-general. Two methods might allow domain-general misconceptions. First, the domain-general misconceptions might tie to a specific instructional model and thus let GIFT infer undesirable facts about ways of learning. An example might be a misconception that it is better to avoid poor outcomes during training, when instead the specific instructional model benefits from presenting poor outcomes that challenge the learner. Second, domain-general misconceptions could be reclassified as characteristics of the learner rather than any concept. Then the GIFT learner module would update learner states and traits, rather than readiness or concept mastery. This would give additional input to existing constructs which GIFT typically infers through surveys, such as learner grit or mastery orientation versus performance orientation.

Mid-lesson Reports

With knowledge of specific misconceptions as inferred from behavior patterns, GIFT could recommend immediate feedback that enhanced the delayed feedback already in the training. Again, reusability and generality of the approach was key. Immediate feedback was designed in the form of mid-lesson reporting. This is a form of adaptation that does not rely on information about the scenario and could be reused in any scenario during the experiment.

Examples of immediate feedback that mid-lesson reporting added to GIFT feedback (DeFalco, 2017; Goldberg, Sottilare, Brawner, & Holden, 2012) include relating good or poor performance examples to underlying reasons for performance and providing appropriate reflective prompts (Swan, 1983). Mid-lesson reports were hypothesized to improve learning outcomes by combining immediate feedback with student-directed learning and reflection. Through reflection, the learner observes the state resulting from actions and uses information from those observations to guide decisions about which actions to perform and how. To encourage reflection required including feedback on observed actions that linked the actions to target competency (Shute, 2008). This was accomplished with report messages that simply and immediately stated a possible misconception when GIFT inferred it. The report thus could be reused across scenarios, because the learners would fill in specifics from their knowledge of the action they just performed. Similar to an open learner model, the mid-lesson report would help to directly link learner choices to the computer's inference and prompt reflection on whether the inference was correct.

Three mid-lesson reports could appear for each misconception. After the third report, the wording repeated again starting from the first message. The wording of each report message was similar, with the name of the misconception differing and colored red on screen for visual differentiation. Nine positive mid-lesson reports were also available to GIFT. The mid-lesson reports let GIFT choose to challenge specific misconceptions, encourage reflection, and improve training. The design of their wording and delivery made them applicable for GIFT to deliver at any point during any scenario, for this experiment.

METHODOLOGY

The training was evaluated with a population of West Point Cadets. The authors owe a great deal of thanks to West Point faculty, staff, and Cadets for supporting and enabling this study. The population consisted of N = 74 Cadets of all years. N was determined by a power analysis targeting moderate effect size. Cadets were randomly assigned to one of two conditions, experimental or control. Each condition had 37 Cadets assigned, although one Cadet in the experimental condition either ended early or lost data due to technical failure (after the training scenarios and before the post-test and final survey). Cadets in the two conditions did not significantly differ in their performance on a subject-matter pre-test.

Study participation took 55 minutes or less per participant. The study proceeded as follows:

1. Participants read and signed an informed consent.

2. An investigator instructed the participants on the use of the training system.

3. Participants completed a demographic questionnaire and subject matter pre-test in GIFT.

4. Participants interacted with the training scenarios in order. They were allowed to review and replay each scenario if desired, but could not return to a scenario once completed.

5. Participants completed a subject matter post-test (with items identical to the pre-test for simpler balancing) and a technology acceptance survey.

The two experimental conditions differed during step 4 only. In both experimental and control conditions, the scenario content and summary feedback was the same. However, in the experimental condition, GIFT overlaid mid-lesson reporting along with the content. The mid-lesson reporting appeared in the form of tailored text messages in the GIFT TUI on the left side of the screen (presented as text only, with no character or speech). GIFT tailored the mid-lesson report messages based on choices participants made during each scenario and across scenarios. In the control condition, the TUI was always left blank.

Situational Judgment Tests

The study instrument this paper describes in detail is the pre-test and post-test. These tests were created for this study and were identical in presentation before and after the training. Each test consisted of ten written *situational judgment test* (Motowidlo & Beier, 2010) items with four to six options for each item. Participants were asked to apply their knowledge of the subject matter by stating how likely they are to try each option in the situation described. Participants could answer each option with an integer between one and six, similar to a Likert scale with excluded middle.

Each test item related to one or more learning objectives in the cross-cultural communication subject matter. Five test items were written as examples of near transfer, closely reflecting circumstances depicted in the training scenarios. Five test items were written to test far transfer. The far-transfer test items applied the same principles in new circumstances that the participants had not experienced during training. The far transfer test items were hypothesized to be valid because they were based on additional training scenarios developed by the same SMEs that were not shown to participants during the study.

Each option within a test item expressed either Good Stranger behavior or behavior associated with a specific misconception. Participants rated each option separately and were not required to choose one option or to rate one option higher than others. As a result, the pre-test and post-test yielded a large amount of data about participant skill equivalent to approximately 50 Likert items. To the extent that the options loaded on separate factors in the hypothesized five-dimensional skill application model, each dimension of that model (misconception or correct behavior) was represented on the test by between seven and ten Likert items. Seven Likert items are suggested to adequately measure most constructs (Willits, Theodori, & Luloff, 2016).

An example test item was Question 4, which produced an interestingly mixed outcome. The outcome is discussed in detail in the Results section. The text of this example test item was as follows:

4. A group of local contractors are compromising security to save time. You talked to them once about it but they have not changed their ways. (**LO = balance tact and tactics / near transfer)**

a. ______ threaten to accuse them of helping the enemy **(Rules-focused)**

b. ______ increase patrols to backfill the compromised security **(Mission-focused)**

c. ______ fake an attack demonstrating how compromised the security is **(Authoritarian)**

d. ______ wait and follow up if something happens **(Cautious)**

e. ______ call higher command **(Good Stranger)**

Additional measures captured included demographic information, cognitive load after each test and training scenario, and a final questionnaire about the training system as a whole. This paper covers initial analysis of the situational judgment tests only, while a more comprehensive analysis is still ongoing. The additional analysis will include detectable differences based on demographics if any, self-efficacy, user acceptance of the technology and training, relationship between patterns and ability to predict post-test performance, inferred misconceptions in relationship to test performance, choices made and content seen in relationship to differences in learning, and interventions presented in relationship to learning.

RESULTS AND DISCUSSION

The study found the following initial results. First, the system supported learning. Second, patterns differentiated learners. Third, inferred misconceptions aligned with ground truth as determined by the pretest and post-test. Fourth, the research suggested improvements to make this approach more general and more effective.

Training Effectiveness

Pre-test and post-test responses on the subject matter test were compared and evaluated for evidence of improvement. Participants learned and improved significantly on three of the five near-transfer test items. One test item was subject to a severe ceiling effect. The remaining test item, question 4 above, was chosen for an in-depth analysis of patterns in learner experience as described below. Figure 1 shows the outcomes of the near-transfer test items.

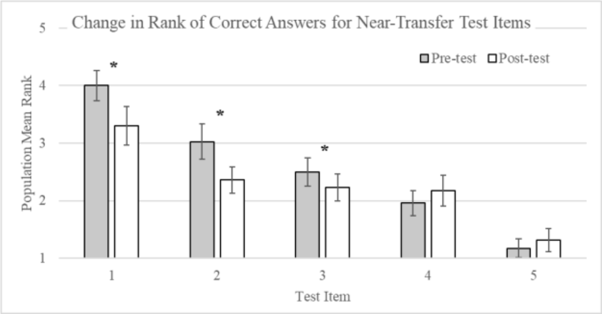


Figure 1: After training, the correct answer was ranked higher (closer to one) in 3 of 5 items.

Learning outcomes in Figure 1 were measured by the differences in scores earned on pre-test and post-test administrations of a subject matter test. The initial analysis assigns scores to test answers based on the *rank* of the Good Stranger choice as compared to other choices within the same test item. Ratings in a form such as Likert data and the answers to these test options are not scalar and therefore cannot be averaged or subjected to t-tests (Jamieson, 2004). They also cannot be compared across participants. Therefore, comparing answer rank is a more valid way to make test scores comparable and subject them to statistical tests for significance. In a test item with five options, each participant might rank the Good Stranger option first (more likely than any other) through fifth (less likely than any other). As a result, the possible ranks are comparable as real numbers between 1 and 5 inclusive, with lower numbers indicating a better score. A significant decrease in the rank score between pre-test and post-test indicates that learning occurred.

Wilcoxon rank tests were applied to the test items in Figure 1. The score improvement between pre-test and post-test was statistically significant for three of the five items (marked with asterisks). The error bars in Figure 2 indicate 95% confidence intervals. In no case do the error bars contain the extreme values (5 and 1). However, test item 5 suffered from a ceiling effect. Most participants (93%) were already rating the correct answer first in the pre-test, before any training. Fortunately, none of the other nine test items placed a majority of learners at ceiling on the pre-test.

Unlike the near-transfer questions, with the five far-transfer questions learners did not show significant improvement from pre-test to post-test. The mid-lesson reporting intervention was expected to support far transfer by making the underlying structure of the material explicit (defining the Good Stranger and the specific expected misconceptions). Further analysis is needed to determine what portion of participants received or viewed which mid-lesson reporting interventions. It is hypothesized that even in the experimental condition, participants may not have received all or enough mid-lesson reports to support effective far transfer. This can be determined with further analysis by correlating test outcomes with the specific interventions that were delivered.

Finally, test item 4 provided an interesting result which led us to examine that item further. Test item 4 may have revealed an instance of ineffective training because learners did worse on the post-test more often than they did better. This trend did not reach statistical significance but it did maintain that direction even after excluding the large number of learners (35%) who were already at ceiling on the pre-test.

A benefit of the patterns being added to GIFT is that instructors or researchers can use them to relate observations into meaningful groups and draw inferences from them. An example is finding and dealing with ineffective training after a system has been published.

Using Patterns to Detect Ineffective Training

Training can be ineffective, even if initially validated, for a number of reasons. Real-world reasons might include (1) changes in learning context that make the training less impactful or (2) changes in tactics, techniques, and procedures that make the training obsolete. Here we discuss an application of patterns to analyze learner experience in Scenario 1, which is directly aligned with the pre-test and post-test item 4.

All training scenarios used in the experiment were created by Army subject matter experts and were previously used in training experiments with Soldiers. However, an unexpected difference in training arose because of these scenarios' presentation outside the context of other training material. Specifically, Scenario 1 offers no action choices that do not result in some negative message. Either the learner's action will compromise unit safety or reduce trust with the local counterparts. The doctrinal answer is choosing to sacrifice trust to maintain security – call higher command to intervene in this case. The learner should "balance tact and tactics," and should choose to accept sacrificing tact under the circumstances.

The design of the first scenario is intended to introduce the idea that in a situational judgement test there may be a tradeoff of competing values. The choice of where to strike the balance measures which value is a higher priority, and there may be no way to simultaneously achieve all good outcomes. This idea was alluded to in the test instructions which stated: "Choices you make in this study may include limited options where there is no wholly right answer. Try to choose the best available option." However, the fact that no answer would avoid all negative outcomes was not made explicit during Scenario 1. As a result, many participants may have drawn an incorrect conclusion when they received negative feedback during the scenario. They assumed that if their first choice resulted in negative feedback, then their second guess must have been the correct one. If learners did not replay the scenario and attempt another choice, then their assumption was not challenged and created a misconception that was reflected in the post-test.

GIFT can detect patterns of behavior that suggest how different learners confronted this scenario. Examples of patterns that appeared in a post hoc analysis included completing the scenario and simply moving on without questioning. However, a relatively high percentage of learners instead changed their answer before submitting it, paged backward to review content both before and after submitting, read more of the after-action review content, and replayed the entire scenario. These patterns suggest different ways of confronting the challenging content that can be detected with patterns that pick out specific relations between multiple domain messages.

GIFT can also help to address training that becomes ineffective and help instructors make it effective again using the ability to easily author tutoring overlays in response to domain patterns. With patterns, an instructor or other non-technical author could author a new pattern that describes reading this scenario without replaying or without sufficient replaying. The pattern could trigger an overlay message that makes the tradeoff more explicit to learners. The overlay might help explain a subjective decision or might simply encourage trying other options to see what happens. In this way, a need for change in the existing content is addressed by the future pattern and overlay authoring in GIFT.

Table 1: Example paths through Scenario 1 can be expressed with patterns of order, timing, and repetition.

Example	Path
1	Cautious option, replay, mission-focused option, replay, Good Stranger option, stop
2	Cautious option, review, replay, same cautious option, review, replay, same cautious option

3	Mission-focused option, review, review, stop
4	Mission-focused option, stop

Table 1 describes a sampling of learner experiences during the choice point of Scenario 1. The number of unique paths through the single choice point, out of 74 participants, was 57. This is higher than anticipated because there was more diversity than expected in how learners engaged with the training system. New patterns should be added and tested to group these paths into meaningful categories. Table 1 suggests some possible patterns that appeared in the data and could be added to those defined in the introduction.

Example 1 describes a learner who is almost ideal in making the most of training. Whether purposefully exploring all options or simply persisting in order to get a better result, this learner plays through the entire first scenario three times, trying three different options. On the post-test, this learner was not confused by tradeoffs in the scenario about the correct answer, and ranked the correct answer highest.

Example 2 seems to describe a learner in a state of non-productive frustration or wheel spinning (Beck & Gong, 2013). When faced with negative feedback, this learner replays the scenario but tries the same option again. As if to amplify the frustration with the available options and the unwillingness to explore another path, the learner tries the same sequence a third time before quitting. Unlike the participants who tried three different options, this learner appears to need a hint to try something new.

Example 3 is the most common pattern observed in the first scenario. The learner attempts an answer, receives negative feedback, and then takes some action to learn more – but not effective action. The act of reviewing the material in this example may indicate an attempt to learn from the mistake, but the learning is not completely effective if no alternative is tried. Subject to further analysis, this pattern may be associated with worse performance on the post-test because the learner may form a belief about what the right choice was, but does not test it.

Example 4 seems quite abrupt in ending the scenario. Further analysis will determine if this pattern also includes shorter time spent reading. The failure to take advantage of the after-action review may indicate disengagement. Perhaps a more extreme example is another learner who changed the answer several times before submitting it, suggesting uncertainty, and even so ended the scenario immediately.

Since there were more paths through a single choice point than expected, and since they seem to be grouped into patterns such as those cited and others, future research should be conducted to create further patterns and test how they help GIFT infer facts about learners.

Inferring Misconceptions and Learner Characteristics

GIFT inferred misconceptions with an updated learner model structure that matched each concept with multiple misconceptions that list specific reasons a learner might be expected to be below expectations on that concept. Standard domain conditions were created to translate performance in the training scenarios to evidence of misconceptions. Presence of misconceptions enabled the pedagogical module to present targeted mid-lesson review interventions.

The accuracy of misconceptions in the learner model was measured by comparing misconception estimates against pre-test and post-test scores. In this initial report, only Scenario 1 and the matched test item 4 were analyzed. A full analysis of all scenarios and test items will be reported separately and will be able to include findings of change over time and statistical significance in the results. Note that experimental and control conditions did not differ in the method of inferring misconceptions, so the analysis includes the full population of learners.

Logs for the experiment were reviewed to determine the first misconception, if any, that GIFT detected for each learner during Scenario 1. This misconception was compared against the learner's pre-test answer to test item 4 only. If the option corresponding to this misconception was the highest ranked option on the pre-test, then GIFT was considered to have correctly identified a misconception the learner did have in mind before starting Scenario 1. The overall accuracy of the first misconception GIFT detected was 52.9%. This accuracy is higher than random chance for a five-way classification task.

Next, the experiment logs were again examined to determine the inferred misconception that GIFT estimated for each learner at the end of Scenario 1. The inferred value was compared to the learner's post-test answer for test item 4 only. If the learner's answer gave top rank to an option reflecting the same misconception as GIFT predicted, then GIFT was considered to have correctly predicted the learner's state after Scenario 1. The overall accuracy on this test was 56.5%, which is again higher than random chance for a five-way classification task.

There were several limitations in this analysis which should be addressed by further analysis of the same data. First, only one pre-test item and one post-test item were considered. A method should be created to combine the results of all test items for each learner, in order to create a better picture of ground truth about their misconceptions. Second, the analysis does not account for learning that takes place after Scenario 1. Presumably GIFT had formed different, possibly improved estimates of learner misconceptions by the end of all the training scenarios, which should be collected and compared to find change over time. Third, the analysis only considered misconceptions detected during Scenario 1 and Scenario 1 actually did not contain an opportunity for the learners to display a rules-focused misconception. As a result, prediction accuracy was zero for all learners who really did have a rules-focused misconception in mind. Therefore a full-scale analysis may find increased accuracy by removing this handicap.

In summary, GIFT could detect misconceptions related to a specific concept. As was discovered in analysis after the experiment, it may also be possible to detect learner characteristics. Some of these characteristics already appear in the default GIFT learner model. They represent an opportunity for GIFT to update the modeling of these characteristics. These include (1) persistence or grit, (2) performance orientation versus mastery orientation, and (3) external locus of control versus internal locus of control. In addition, some characteristics suggested by the data are new and could be added to an enhanced learner model. These include (1) spinning wheels, (2) frustration or disengagement, and (3) curiosity or willingness to explore. Further research is needed to confirm that these possible traits and states are actually present in learners and are detectable by behavior patterns as they seem to be. If they are, then GIFT could have a new capability to detect domain-general "misconceptions" or facts about how people learn and use the facts to change an adaptive training experience.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

In conclusion, the present experiment demonstrated an integration of research into GIFT authoring with an actual training system and typical users. The research capabilities demonstrated were patterns that help tell the difference between individual learners based on fine-grained behavior, misconceptions that make the GIFT learner model more precise about why any learner might be below expectation, and mid-lesson reports that give a reusable tool for addressing misconceptions with reflective prompts.

The research suggests several directions for future research. First, additional research is needed to *close the loop* on mid-lesson reporting and improve the interventions to make them more effective. With future work to incorporate patterns and misconceptions into the GIFT authoring suite, it would then be possible for an instructor to define mid-lesson reports in response immediately on recognizing that training is not proceeding as expected. This quick response to fix ineffective training is made possible by authoring patterns and misconceptions that let GIFT capture and interpret observations in a manner similar to how instructors and end users think about training.

In addition, valuable future research could build on the success of patterns and misconceptions here in order to create an instructor-facing learning analytics capability. Analytics are a burgeoning field in adaptive training and can use the fine-grained data in GIFT to give insight into how learners are using training and what training is more or less effective. Incorporation into GIFT would also help provide instructors and unit leaders with tools to focus their training effort in real time.

Finally, in the immediate term the research will be extended to additional training domains in order to ensure the generality of patterns and misconceptions for describing training in ways that instructors want to understand. A candidate domain might be a VBS room-clearing scenario which is already created and has collected some data pertaining to learner movement in a simulated space. This kind of extension will provide an excellent opportunity to show how patterns and misconceptions let human instructors or SMEs intuitively label complex behavior in a manner that is very challenging to machine learning algorithms.

With these additional advances, the present research will add to the end-user author-ability of increasingly sophisticated and effective adaptive training in GIFT.

REFERENCES

Beck, J.E., & Gong, Y. (2013). *Wheel-spinning: Students who fail to master a skill.* Paper presented at the International Conference on Artificial Intelligence in Education.

DeFalco, J.A. (2017). Examining motivational feedback for sensor-free detected frustration within game-based learning. Columbia University.

Folsom-Kovarik, J.T., & Boyce, M.W. (2017). *Developing a Pattern Recognition Structure to Tailor Mid-Lesson Feedback*. Paper presented at the 2017 GIFT Symposium, Orlando, FL.

Goldberg, B., Sottilare, R., Brawner, K., & Holden, H. (2012). *Adaptive game-based tutoring: Mechanisms for real-time feedback and adaptation.* Paper presented at the Proceedings of the I3M Conference on International Defense & Homeland Security Simulation Workshop.

Hubal, R., van Lent, M., Marinier, B., Kawatsu, C., & Bechtel, B. (2015). *Enhancing Good Stranger Skills: A Method and Study*. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), Orlando, FL.

Jamieson, S. (2004). Likert scales: how to (ab) use them. *Medical education, 38*(12), 1217-1218.

Klein, G., Moon, B., & Hoffman, R.F. (2006). Making sense of sensemaking I: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73.

Motowidlo, S.J., & Beier, M.E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology, 95*(2), 321.

Shute, V.J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Swan, M. (1983). Teaching decimal place value: A comparative study of "conflict" and "positive only" approaches: Nottingham University Shell Centre for Mathematical Education.

Willits, F.K., Theodori, G.L., & Luloff, A. (2016). Another Look at Likert Scales. *Journal of Rural Social Sciences, 31*(3), 126.

Wray, R.E., Folsom-Kovarik, J.T., Woods, A., & Jones, R.M. (2015). Motivating narrative representation for training cross-cultural interaction. *Procedia Manufacturing, 3*, 4121-4128.

ABOUT THE AUTHORS

J.T. Folsom-Kovarik, Ph.D. *is the lead scientist at Soar Technology, Inc. for adaptation and assessment within intelligent training. He earned a Ph.D. in computer science from the University of Central Florida in 2012. His research combines modern data science and machine learning approaches with SoarTech's long experience in modeling expert knowledge and human experience. When expert knowledge guides machine learning and data analytics algorithms, they become more applicable and useful in real-world training settings. The combination of approaches can remain feasible when available data is small, concepts evolve over time, or nontechnical users need to control the training.*

Michael W. Boyce, Ph.D. *is a research psychologist with ARL's adaptive training research program. For the past 3 years his emphasis has been in using technologies like GIFT to accurately assess learner knowledge and performance. Located at the United States Military Academy at West Point, his goal is to better inform the research progress of GIFT through interactions with a military student population. He received his Ph.D. in Applied/Experimental Human Factors Psychology from the University of Central Florida in 2014.*

Robert Thomson, Ph.D. *serves as the Cyber and Cognitive Science Fellow at the Army Cyber Institute and is an Assistant Professor in the Department of Behavioral Sciences and Leadership at the United States Military Academy. Dr. Thomson has over 7 years of post-graduate experience and over 30 invited and refereed academic publications in the domains of computational modeling, intelligence analysis, cybersecurity, and artificial intelligence. He has received funding from IARPA, DARPA, ONR, and ARL. He is also a faculty representative for the Men's Basketball Program at West Point.*

Integrating Sensors and Exploiting Sensor Data with GIFT for Improved Learning Analytics

Jong W. Kim, Robert A. Sottilare, Keith Brawner
US Army Research Laboratory, Orlando, FL
Timothy Flowers
Dignitas Technologies, Orlando, FL

INTRODUCTION

The Generalized Intelligent Framework for Tutoring (GIFT) was constructed in order to make it easier to create intelligent tutoring systems (ITSs), develop a shared set of authoring tools to do so, and to enable research in ITS (e.g., Sottilare, Goldberg, Brawner, & Holden, 2012). ITS research takes many forms; as an example, some of this research is intended to support existing training simulations (e.g., Brawner, Holden, Goldberg, & Sottilare, 2011). Further, some of this research is in the manners in which to model a learner (Brawner & Goldberg, 2012; Goldberg, Sottilare, Brawner, & Holden, 2011). Some of these models of the learner are not solely based on interactions that they have within the environment, but also upon sensors (e.g., Brawner, 2017; Brawner, Sottilare, & Gonzalez, 2012). Such sensors can either be based in software, analyzing information such as system interactions and clicks, or hardware (e.g., DeFalco et al., 2017), with sensors which sense physical items such as posture or gaze.

With computer-based or simulation-based training, sensors are somewhat optional; the learner can interact with the system, take actions, make progress, learn, and perform other activities without the explicit need for monitoring. Certain domains, such as psychomotor training or medical skill training, however, require the use of a sensor to monitor and identify the learner's performance. Particularly, using sensors can benefit to assessment of the learner (e.g., Goldberg, Amburn, Ragusa, & Chen, 2017), providing a source of information to the rest of the system so that the learner with adaptive instructions and feedback.

Integrating and synchronizing data from heterogeneous sources of sensors can be somewhat complicated and challenging, especially in a psychomotor domain, since sensors can have their own sampling behaviors and data stream formats. For example, the experimenter utilizing sensor data would need to connect with the various data streams of various sensors during the data collection from an experiment with human participants. Synchronizing the different sources into a time series and analyzing them would be complicated. It is, thus, necessary to investigate a general and reliable approach to better exploit the sensor data as a series of data points indexed in a time order by synchronizing all the heterogeneous sources of sensors.

The goal of this paper is to provide what is the general approach to exploit heterogeneous sources of sensor data in various domains including cognitive and psychomotor domains. We choose to use and explore the GIFT capability since it provides a generalized framework for a computer guided adaptive instruction, and there are many pre-existing efforts which integrate sensors with GIFT. These sensor streams were able to provide rich learning analytics (Brawner, 2017; Brawner & Gonzalez, 2016; DeFalco et al., 2017). We examine the current capability and provide directions to extend the capability in order to better assess the learner performance in the diverse domain. We also examine the process to integrate new sensors with GIFT, and provide suggestions for improved systematic process of integration.

To pursue the aforementioned goal, in this paper, we first review and summarize the current process of integrating sensors with GIFT, and identify the technical needs to synchronize multiple sensors for improved learning analytics. In addition, we report a use case from our exploratory study. We have created a study environment in GIFT, where a psychomotor skill can be assessed by sensors by extending an adaptive training on rifle marksmanship (Goldberg, Amburn, Ragusa, & Chen, 2017). A golf putting was selected as a psychomotor training task because it is physical and precision-required performance like rile marksmanship. It is argued that breathing techniques would affect the precision-required performance of rile marksmanship (e.g., Grossman & Christensen, 2008), and it is also suggested that a slow breathing skill can help individuals to improve accuracy on their performance in other tasks (e.g., Goldberg, Amburn, Ragusa, & Chen, 2017; Kim, Dancy, Goldberg, & Sottilare, 2017).

SENSORS INTEGRATED WITH GIFT

Several commercial and custom-built sensors have been integrated with GIFT to support learner assessments that include learner engagement, arousal, motivation, knowledge, anxiety, and engaged concentration. These learner states are defined in Table 1. It is reasonable to think that they can influence learning, as prior research has shown effects.

Table 1. Learner states tracked in GIFT.

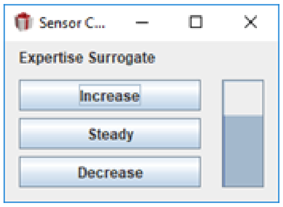
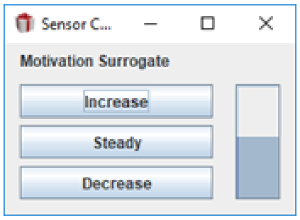
Learner States	Definition	References
Engagement	"refers to the degree of attention, curiosity, interest, optimism, and passion that students show when they are **learning** or being taught, which extends to the level of motivation they have to learn and progress in their **education"**	(The Glossary of Education Reform, 2016)
Arousal	"a major aspect of many learning theories and is closely related to other concepts such as anxiety, attention, agitation, stress, and motivation. One finding with respect to **arousal** is the Yerkes-Dodson law which predicts an inverted U-shaped function between **arousal and performance**"	(Clark, n.d.)
Motivation	"an internal **drive** that activates **behavior** and gives it **direction**"	(Rakus, 2011)
Knowledge	"a familiarity, awareness, or understanding of someone or something, such **as facts, information, descriptions, or skills**, which is acquired through experience or education by perceiving, discovering, or learning"	(Knowledge is a familiarity, n.d.)
Anxiety	"a feeling of **worry, nervousness, or unease**, typically about an imminent event or something with an **uncertain outcome**" (Anxiety [Def. 1], n.d.) "anxiety impacts a student's **working memory**, making it **difficult to learn and retain information**" (Minahan, 2012)	(Minahan, 2012)

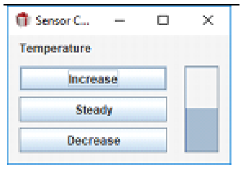
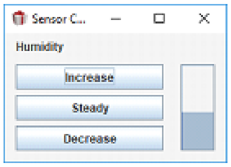
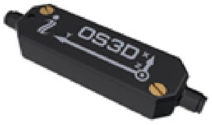
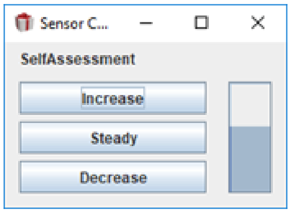
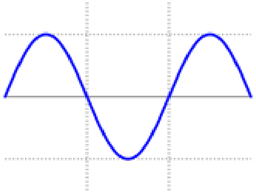
Engaged concentration	“a state of engagement with a task such that concentration is **intense, attention is focused, and involvement is complete**”	(Baker, D'Mello, Rodrigo, & Graesser, 2010)

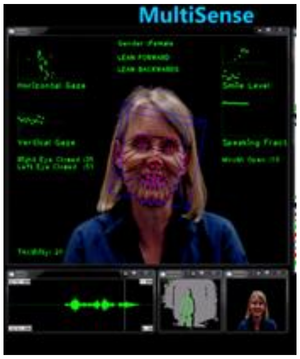
To support the assessment of the learner states listed in Table 1, we have developed interfaces for a series of commercial and customized sensors for use during GIFT instruction and developmental testing. A table of sensors integrated in GIFT are listed in Table 2, along with their descriptions, inputs, derived measures and a picture of the sensor hardware or surrogate.

Table 2. Sensors integrated with GIFT.

Sensor	Description & Inputs	Derived Measures	Picture
Zephyr Bioharness	ECG, respiration, estimated core body temperature, accelerometer, time, and location	Heart rate (HR), breathing rate, heart rate variability, HR confidence, estimated core body temperature, impact, activity, caloric burn, posture, % HR max, % HR at anaerobic threshold (AT), accelerometer, training loads and intensities, jump, bounds, leaps, explosiveness, peak force, peak acceleration, GPS	
Emotive EmoComposer (Alshbatat, Vial, Premaratne, & Tran, 2014)	As part of the Emotiv Software Development Kit (SDK), the EmoComposer is a testing tool for developers building EPOC headset applications	The derived measures are unique to each application developed	
Emotiv Epoc EEG Headset (Lang, 2012)	Brain control interface with 14 channels of EEG data	Instantaneous excitement, long term excitement, engagement/ boredom, frustration, and meditation	

ARL Expertise Surrogate	Allows tester to vary expertise or domain competency This surrogate is used for testing in place of any other measure of expertise (e.g., assessment/test).	Expertise or domain competency	
Microsoft Kinect	Allows users to act as the controller and interact with simulation elements using a combination of body movement and spoken commands IR Depth Sensor measures the distance of each pixel of an object from camera plane	Emotional states (facial markers); engagement (posture); arousal (acceleration measures)	
Microsoft Band 2	Optical heart rate sensor; accelerometer/gyro;; GPS; ambient light sensor; skin temperature sensor; UV sensor; capacitive sensor; galvanic skin response; microphone, barometer	Heart rate, steps, location, galvanic skin response (GSR) Resistance and GSR conductance	
ARL Motivation Surrogate Sensor	Allows tester to vary the motivation level of a user This surrogate is used for testing in place of any other measure of motivation (e.g., survey instrument).	Motivation level	
ARL Mouse Temperature & Humidity Sensor	Temperature and humidity of a user's hand	Arousal, stress	

ARL Mouse Temperature & Humidity Surrogate Sensor	Allows tester to vary temperature and humidity of a user's hand This surrogate is used for testing in place of the actual mouse temperature and humidity sensor	Arousal, stress	 Sensor C... Temperature Increase Steady Decrease 
Inertial Labs 3D Orientation Sensor (OS3D)	Changes to velocity (acceleration) and disturbances to magnetic fields	Real-time heading, pitch and roll orientation information	
Affectiva Q Sensor	Electro-dermal Activity (EDA), Temperature, Acceleration (3D)	Arousal, stress	
ARL Self Assessment Sensor	Allows tester to vary a user's self-assessment This surrogate is used for testing in place of any other self-assessment methods	Self-assessment of performance	 Sensor C... SelfAssessment Increase Steady Decrease
ARL Sine Wave Sensor	Allows tester to vary any user's attributes as sine waves This surrogate is used for testing in place of any other methods to vary learner attributes sinusoidally	Sinusoidal representation of learner attributes (e.g., engagement)	

USC Virtual Human Toolkit Multisense (Scherer et al., 2012)	A perception framework that enables multiple sensing and understanding modules to interoperate simultaneously, broadcasting data through the Perception Markup Language; includes the Generalized Adaptive View-based Appearance Model (GAVAM), Constrained Local Model (CLM) and Flexible Action and Articulated Skeleton Toolkit (FAAST)	GAVAM – head tracking CLM – face tracking FAAST - middleware to facilitate integration of full-body control with games and VR applications	 MultiSense Gender: Female LEAN FORWARD LEAN BACKWARDS Horizontal Gaze Smile Level Vertical Gaze Speaking Fract Right Eye Closed Left Eye Closed Mouth Open

LESSONS LEARNED FROM INTEGRATING SENSORS WITH GIFT

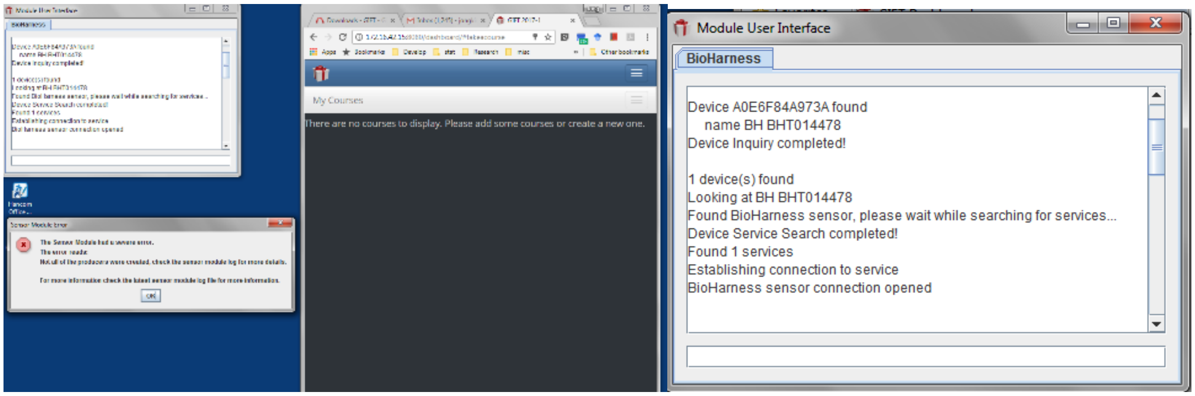
Besides the various sensors integration with GIFT shown in Table 2, it was identified that there is a challenge to expand the instructional domains. One goal for the design of GIFT is to expand the number and type of instructional domains in which it can support tutoring of both individual learners and teams (e.g., Brawner, Sinatra, & Gilbert, 2018; Sottilare et al., 2017), and tutoring of psychomotor tasks beyond the desktop environment (e.g., Sottilare, Hackett, Pike, & LaViola, 2017). We have been extensively involved in developing strategies (Sottilare & LaViola, 2015) and concepts for psychomotor tasks like marksmanship (Goldberg, Amburn, Brawner, & Westphal, 2014), land navigation (LaViola Jr. et al., 2015), and hemorrhage control (Sottilare, Hackett, Pike, & LaViola, 2017). Designing tutoring for the psychomotor domain has also influenced the selection and use of sensors to support assessment during instruction. For example, the land navigation task has necessitated the use of mobile devices (e.g., smartphones) and associated sensors to support assessment. We have also examined pressure sensors and designed how they might be used to assess the use of pressure bandages and tourniquets during combat casualty care to determine blood flow from wounds. As we more fully develop these concepts, we will also develop interfaces for the associated sensors and make them available in the GIFT baseline. In this section, we report the lessons learned from the use of sensors and sensor data analytics in a psychomotor task training.

An Example for Using a Smartphone Sensor with GIFT

Integrating sensors with a system can be somewhat straightforward; it is suggested for the developer to simply follow the template for code, processing, and configuration. Any of the processing which has been authored for any of the sensors may be able to be reused for any of the other sensors with authored configurations. The “sine wave sensor” can be used to test out any individual item (connection, configuration, processing, etc.). Step 1 is to “make a sensor connection using one of the numerous interfaces”, Step 2 is to “configure it with the configuration tool, probably just copy whichever sensor you used previously”. For example, integrating the Android phone’s accelerometer and gyroscope sensor into GIFT consists of a few basic steps. We, first, developed an Android app that can access the phone’s sensor data and that can relay the data stream to the GIFT desktop. The streamed sensor data from the app are formatted as JSON, with timestamped UDP (User Datagram Protocol) packets to an IP address that is configurable from the app. Once the Android application was operational, GIFT could be modified to handle the incoming UDP packets. GIFT defines an abstract class which generically represents communication with a sensor.

An additional implementation of this class (`AbstractSensor`) was created to receive data from the Android phone's sensors. The class is named `AndroidPhoneSensor` and overrides `AbstractSensor`'s methods: `start`, `stop`, and `test`. The internal implementation of `AndroidPhoneSensor` starts a new thread when the `start` method is called. This thread continuously listens for the UDP packets from the Android device. When a packet is received, it parses it and places each of the six data measurements from the packet (three dimensions of accelerometer data and three dimensions of gyroscope data) into a `SensorData` Java object which is then sent to the Sensor Module's existing pipeline for processing by GIFT. Once the `SensorData` object has been sent to GIFT, the thread listens for another packet.

For implementations in the future, it may be beneficial to create an abstract implementation of `AbstractSensor` which receives JSON data via UDP and defers interpreting the JSON to drivers of the class. This would make the majority of the code written within `AndroidPhoneSensor` to be reusable and help separate common boilerplate code from the code which is specialized to a specific sensor.



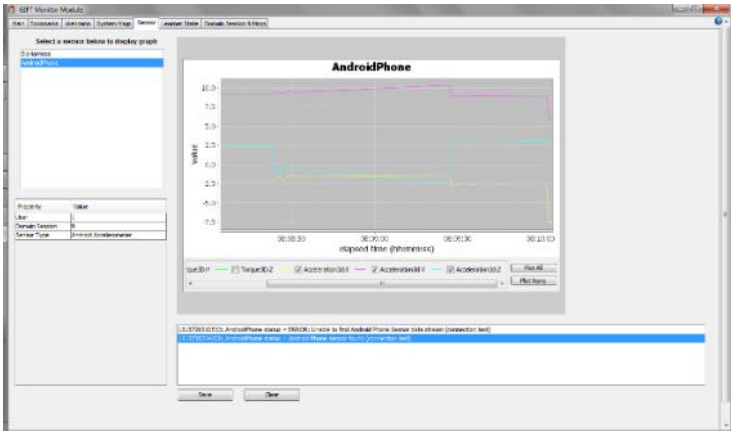
The Sensor Module had a severe error.
The error reads:
Not all of the producers were created, check the sensor module log for more details.

For more information check the latest sensor module log file for more information.

There are no courses to display. Please add some courses or create a new one.

```
Device A0E6F84A973A found
   name BH BHT014478
Device Inquiry completed!

1 device(s) found
Looking at BH BHT014478
Found BioHarness sensor, please wait while searching for services...
Device Service Search completed!
Found 1 services
Establishing connection to service
BioHarness sensor connection opened
```

Fig 1a. The GIFT study environment for a psychomotor task: Initiating the BioHarness sensor through the Bluetooth connection.



Property	Value
User	1
Domain Session	6
Sensor Type	Android Accelerometer

[illegible]: AndroidPhone status = ERROR: Unable to find Android Phone Sensor data stream (connection test)

[illegible]: AndroidPhone status = Android Phone sensor found (connection test)

Fig 1b. The GIFT study environment for a psychomotor task: Visualizing accelerometer sensor data in GIFT, and the three axes in a smartphone accelerometer sensor.

Sensor Data Exploitation

After the sensor integration with GIFT, it is important to consider how to extract the features of the learner performance and behavior from sensor data obtained from sensors. Extracting and processing such data is called analytics. Particularly, when one considers the context of education and learning with a large amount of data, it is called learning analytics (LA) and educational data mining (EDM) (Baker & Siemens, 2014).

Previously, to assess the cognitive and affective states of the learner, researchers have tried to incorporate appropriate sensors into an ITS (e.g., D'Mello et al., 2005). In this line of research, the traditional method to exploit such sensor data for intelligent tutoring is largely dependent on the offline post-processing of the data rather than a real-time model of data analytics (Brawner, 2017) – i.e., taking measurements in a classroom, storing and moving to the offline environment, performing data analytics, and generating a model for the next set of classroom learners. The traditional method is not real-time, which seems to be hard to address varying learning environments. It would be, thus, necessary to advance learning analytics (i.e., an improved real-time assessment model), but it would create another set of problems in the ITS operation since the sensor data could be infinite, outside of control, and strictly constrained by time (Brawner, 2017).

Similar to the affective data exploitation (Brawner, 2017), and the learner logging data processing from a tutor interaction (Baker & Siemens, 2014), sensor data in psychomotor tasks may require us to develop a methodology for efficient data processing and analysis for knowledge discovery, and to compare the sensor data with a theory-based model. A psychomotor task is usually characterized by coordinating cognitive, physical, and physiological variables in executing actions. Thus, sensors are focused on measuring the coordination of the learner features, which can be used to understand the learner according to the features. The physiological data, such as the heart and respiratory rate can be measured using a bioharness strap (e.g., Goldberg, Amburn, Ragusa, & Chen, 2017). Also, acceleration data can be collected and analyzed to identify motions and movements (e.g., Fehlmann et al., 2017; Shamoun-Baranes et al., 2012).

As shown in Fig. 1, we have created a study environment in GIFT where two heterogeneous sensors are combined to measure the learner features. In a pilot testing of the study environment, a participant is to be instructed to learn the breath control skill through a GIFT course, and then to perform a series of golf putting tasks: (a) 5 putting trials under a regular breathing, and (b) 5 putting trials under a tactical breathing condition. Fig. 2 shows plots of the collected sensor data with the time frame from 2:30 to 5:30, which is under a regular breathing with 5 putting trials.

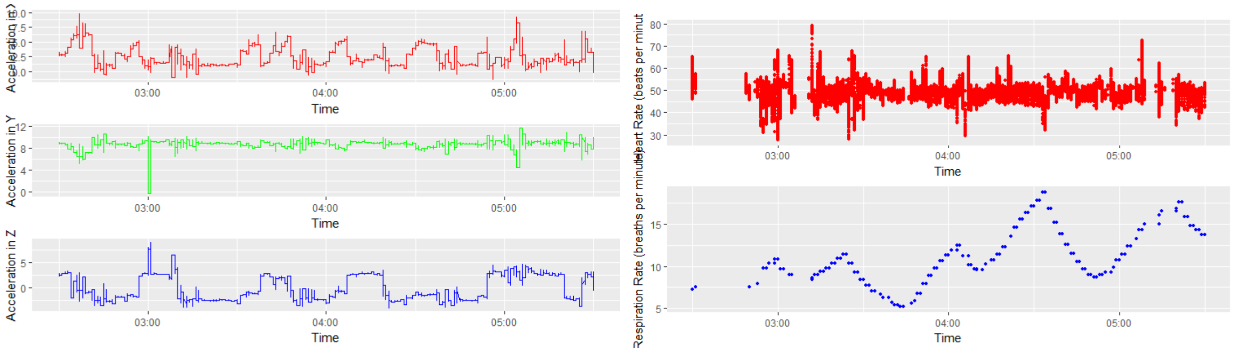


Fig 2. An example of the sensor data.

The Learner Assessment

The current GIFT capability does not fully support the combined sensor data analytics in real time. We report that we have conducted an offline sensor data analytics. We approach learning analytics of the two heterogeneous sensor data from the bioharness strap (e.g., respiratory rate as breath per minute) and the Android phone accelerometer (e.g., the tri-axial values).

We have explored the extended cognitive modeling approach that is based on the ACT-R architecture (Anderson, 2007), and extended to account for a physiological system, called ACT-R/Φ (Dancy, Ritter, & Gunzelmann, 2015). A version of the physio-cognitive model has been implemented (Dancy & Kim, accepted; Kim, Dancy, & Sottilare, submitted), and explored to predict physiological variables (heart and respiratory rates).

Previously, a cognitive model has been used to track knowledge of the learner and to conduct performance assessment in an ITS (Anderson, Boyle, Corbett, & Lewis, 1990; Corbett & Anderson, 1995). This work, however, has been limited to the cognitive task domain in a desktop learning environment. We start to utilize the physio-cognitive model to track the learner knowledge and to predict the learner performance in an attempt to achieve a (near) real-time sensor data analytics in a psychomotor task training.

The leaner models, that can be cognitive (e.g., Anderson, Boyle, Corbett, & Lewis, 1990) or (and) physiological (Dancy & Kim, accepted) based models, can be used for assessing the learner as well. Based on the assessment process, we can provide more reasonable adaptive strategies for training. For example, in a tactical breathing practice, the learner would practice with a 4-4-4-4 cycle of breathing (4 s for breathe-in, 4 s for hold, 4 s for breathe-out, and 4 s for hold). However, the lung capacity or the tidal volume would be different by individuals (e.g., by gender, by age, etc.). A precise and correct learner model can be essential to determine whether a training regimen would be cognitively and physiologically plausible. These aspects of learning can be analyzed through the sensor data exploitation to support improved learning.

Along with the physiological responses, we also explored the sensor data of acceleration. The raw acceleration data is tri-axial, and shows variable changes in values in terms of the xyz axes. The raw data can be processed to recognize movements and motions. That is, acceleration data can be static that is dependent on gravity, and it can be also converted to the dynamic feature of performance as well—e.g., the vectorial dynamic body acceleration can be computed using the dynamic components of the signal to assess the activity level of the individual with three axes all together ($\sqrt{x^2 + y^2 + z^2}$) (Fehlmann et al., 2017).

To further exploit the sensor data of acceleration, it may be helpful to transform the complex signals to another domain. The key idea is to decompose a complex signal in the time domain to the frequency domain through Fourier transform. To identify oscillations in the dynamic body acceleration for each axis, it has been reported that it is possible to compute power spectrum densities (PSDs) and their associated frequencies using Fourier analysis so that we can figure out at which frequency the signal varies the most, indicating a large movement (Fehlmann et al., 2017). Based on this approach, behavior of animals has been investigated to identify six broad states of motions and movements including walking, standing, running, resting (sitting or lying), grooming, and foraging. This technique can be useful to conduct data processing of heterogeneous sensor data collected in a time series manner (e.g., GPS and acceleration sensor data). With regard to the aforementioned psychomotor related sensor data, a machine learning technique (e.g., random forest) can be used to classify a psychomotor task with the leaner data such as sitting, walking, backswing, hitting the golf ball, etc. The sensor data is also worth exploring to predict the learner behavior by validating the model. Brawner (2017) explored machine learning algorithms to address the real-time analytics with cognitive-affective sensor data, highlighting the best real-time model with the learner features is based on offline experimental data validation with a machine learning technique.

DISCUSSION AND CONCLUSIONS

We briefed the use of sensors with GIFT, specifically in the psychomotor task domain. In general, integrating sensors can be relatively considered as a simple process, but interpreting sensor data from multiple sources in a time series manner would be complex and challenging.

Sensor Data in the Psychomotor Domain

In our study environment, the learner’s behavior and performance (i.e., golf putting trials with the breath control skill) can be decomposed to physical, physiological, and cognitive components. On the onset of the tactical breathing course in GIFT, the sensor data collected from a smartphone is sampled and is relayed to the GIFT desktop. The acceleration data shown in Fig. 2 is complex. It shows changes in values within a specific time window and by a series of certain physical motions and movements. A further analysis based on a machine learning technique is needed to reliably cluster and classify changes in the tri-axial values of acceleration, and to identify postures and movements, and to implement a model (e.g., a backstroke, hit, follow-through).

We recorded the participant’s activities (i.e., when the participant starts to perform a slow breathing and a putting trial) by using the Bookmark functionality in GIFT. The Bookmark function affords a record-keeping by an experimenter (or a data collector)—i.e., the timestamped annotations in terms of the participant’s actions. The annotated data can be later matched to the sensor data, and then the data can be labeled by postures and movements.

As an offline analysis of sensor data, we tested a couple of R packages. We found it useful for data analytics and the learner performance assessment. Now the question is how to adopt the approved operational procedures of offline analysis, which attempts to strengthen the GIFT capability. With the pilot testing data, we computed the acceleration raw data to obtain different aspects of the data (e.g., static and dynamic acceleration, vectorial dynamic body acceleration, power spectrum density of acceleration signals). The acceleration data can be mainly categorized into two aspects—static and dynamic. The static acceleration is dependent on gravity, describing postures, and the dynamic acceleration describes dynamic body movements (Fehlmann et al., 2017). Besides the acceleration data from a smartphone, the GPS data can be also explored to investigate movements and motions—e.g., Behavioral Change Point Analysis (Gurarie, Andrews, & Laidre, 2009).

The sensor data regarding the physiological component can be interpreted to identify the breath control skill during the physical performance. In the pilot testing, we collected data from the Bioharness that transmits data through Bluetooth. The participant did wear the Bioharness with the chest strap during the performance. It is observed that the respiration rate (breath per min) looks increasing within the specified time window. We delved into what theory can describe our sensor data. We chose to use a computational model in a cognitive architecture because it can support learning and skill development processes of humans (e.g., Anderson, 2007). Particularly, we implement a physio-cognitive model (Dancy & Kim, accepted; Kim, Dancy, & Sottilare, submitted), which can be used to account for cognitive learning theories (Kim, Ritter, & Koubek, 2013) with physiological features of the learner. The physio-cognitive model supports plausibility of human learning behavior since it is based on a cognitive architecture, ACT-R (Anderson, 2007). That is, the physio-cognitive computational model can support creating a tailored training scenario that can meet cognitive and physiological constraints of humans (e.g., the varying tidal volume of men and women).

The sensor data is usually of a form of oscillations in a time series manner. For efficiency of calculation, the sensor data in the time domain) can be transformed to the frequency domain through Fourier transformation, spectral analysis. This approach has a potential to extend and improve our understanding of the learner behavior (e.g., Fehlmann et al., 2017; Xu & Reitter, 2017). Supposed that an intelligent tutoring system with multiple sensors and with multiple individuals as a team. The aforementioned method, an understanding of power spectrum density of the signals, can be applied to a team performance analytics, e.g., team communication and team collaboration through a dialogue. There is a study about dialogue behavior and effectiveness of conversations, arguing that the spectral analysis can be successful to measure communicative effectiveness (e.g., a successful task collaboration) by considering the alignment of certain linguistic markers, lexical items, or syntactic rules between interlocutors correlates with task success (Xu & Reitter, 2017).

Sensors and Standardization

Recently, we have been involved in developing proposals for standardizing data messaging and interactions between components of adaptive instructional systems (AISs) as part of an IEEE standardization study group. Both sensors for data acquisition and algorithms for state classification may be influenced by the defined functions and information shared between AIS common components. As the types of tasks supported by GIFT expand and as standards take hold in the AIS community, we envision GIFT and its sensor options being updated to optimize models of the learner as a basis for adaptive instruction, which can provide the starting point for standardization (Sottilare & Brawner, in press).

Multiple Sensors and Multiple Learners as a Team

A major design change challenge will also influence the type of sensors and their use in GIFT. With the expansion in GIFT capabilities from individual learner tasks to team tasks, we predict a need for a multi-sensor architecture to track the behaviors of multiple team members in support of team taskwork assessments. Sensors will be needed to disambiguate individual learner data (e.g., position, location, communication) from others on the team to provide individual, subgroup, and group feedback. This is required to provide a model of how individual actions roll up to the attainment of team goals.

Methods of assessment will become more complex as we move from desktop applications to live, augmented, and mixed reality applications. Complexity will also rise as we move from individual to team instructional constructs. The groundwork laid to support individual task domains will largely be reused to support team instruction, but additional team models will be required and team assessments will require logic to understand how individual behaviors and roles influence progress toward team goals. Sensors will continue to play a part in team assessments, but can be provided and extended in the same manner that individual models are extended to team models (Brawner, Sinatra, & Gilbert, 2018).

We are planning spiral development for team model development for adaptive instruction. Initially, we will construct team models that focus only on team measures to simplify the assessment problem. Sensors will be needed to assess whether team objectives have been met. We believe this initial approach will be accomplished with little change to the GIFT architecture as it is today, but more hierarchical modeling of teams in the future spiral phases of development will require methods to link individual learner models and individual roles and responsibilities to team models and objectives. This will require some fundamental additions to the current GIFT architecture. Sensors will be required to disambiguate data from individual learners who may be operating in close proximity in live training environments. Standardizing approaches for different types of team tasks may lead us to more simplified approaches to sensor integration for team tasks.

REFERENCES

Alshbatat, A. I. N., Vial, P. J., Premaratne, P., & Tran, L. C. (2014). EEG-based brain-computer interface for automating home appliances. *Journal of Computers, 9*(9), 2159-2166.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence, 42*, 7-49.

Anxiety [Def. 1]. (n.d.). *Oxford Living Dictionaries*.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223-241.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253). New York, NY: Cambridge University Press.

Brawner, K. (2017). Lessons learned for affective data and intelligent tutoring systems. In *Proceedings of the International Defense and Homeland Security Simulation Workshop (DHSS 2017)* (pp. 9-17). Barcelona, Spain.

Brawner, K., Sinatra, A., & Gilbert, S. (2018). Lessons learned creating a team tutoring architecture: Design reconsiderations. In *Design Recommendations for Intelligent Tutoring Systems: Teams* (Vol. 6). Orlando, FL: U.S. Army Research Laboratory.

Brawner, K. W., & Goldberg, B. S. (2012). Real-time monitoring of ECG and GSR signals during computer-based training. In *International Conference on Intelligent Tutoring Systems* (pp. 72-77). Springer.

Brawner, K. W., & Gonzalez, A. J. (2016). Modelling a learner's affective state in real time to improve intelligent tutoring effectiveness. *Theoretical Issues in Ergonomics Science, 17*(2), 183-210.

Brawner, K. W., Holden, H. K., Goldberg, B. S., & Sottilare, R. (2011). Understanding the Impact of Intelligent Tutoring Agents on Real-Time Training Simulations. In *Proceedings of the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*. Orlando, FL: NTSA.

Brawner, K. W., Sottilare, R., & Gonzalez, A. (2012). Semi-supervised classification of realtime physiological sensor datastreams for student affect assessment in intelligent tutoring. In *International Conference on Intelligent Tutoring Systems* (pp. 582-584). Springer.

Clark, D. (n.d.). Arousal and performance. Retrieved from http://www.nwlink.com/~donclark/performance/arousal.html

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction, 4*, 253-278.

D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R., & Grasser, A. C. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces* (pp. 7-13). New York: AMC Press.

Dancy, C. L., & Kim, J. W. (accepted). Towards a physio-cognitive model of slow-breathing. In *Proceedings of the 40th Annual Conference of Cognitive Science Society*. Wisconsin, USA: Cognitive Science Society.

Dancy, C. L., Ritter, F. E., & Gunzelmann, G. (2015). Two ways to model the effects of sleep fatigue on cognition. In *Proceedings of the 13th International Conference on Cognitive Modeling* (pp. 258-263). Groningen, Netherlands.

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2017). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, 1-42.

Fehlmann, G., O'Riain, M. J., Hopkins, P. W., O'Sullivan, J., Holton, M. D., Shepard, E. L., & King, A. J. (2017). Identification of behaviours from accelerometer data in a wild social primate. *Animal Biotelemetry, 5*(1), 6.

Goldberg, B., Amburn, C., Brawner, K., & Westphal, M. (2014). Developing models of expert performance for support in an adaptive marksmanship trainer. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Goldberg, B., Amburn, C., Ragusa, C., & Chen, D.-W. (2017). Modeling expert behavior in support of an adaptive psychomotor training environment: A marksmanship use case. *International Journal of Artificial Intelligence in Education*, 1-31.

Goldberg, B. S., Sottilare, R. A., Brawner, K. W., & Holden, H. K. (2011). Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 538-547). Memphis, Tennessee: Springer.

Grossman, D., & Christensen, L. W. (2008). *On combat: The psychology and physiology of deadly conflict in war and in peace* (3rd ed.). Belleville, IL: Warrior Science Publications.

Gurarie, E., Andrews, R. D., & Laidre, K. L. (2009). A novel method for identifying behavioural changes in animal movement data. *Ecology Letters, 12*(5), 395-408.

Kim, J. W., Dancy, C., Goldberg, B., & Sottilare, R. (2017). A cognitive modeling approach - Does tactical breathing in a psychomotor task influence skill development during adaptive instruction? In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Augmented Cognition. Enhancing Cognition and Behavior in Complex Human Environments* (pp. 162-174): Springer.

Kim, J. W., Dancy, C., & Sottilare, R. A. (submitted). Towards using a physio-cognitive model in tutoring for psychomotor tasks. In *AIED Workshop on Authoring and Tutoring for Psychomotor, Mobile, and Medical Domains*.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.

Knowledge is a familiarity. (n.d.). Knowledge. Retrieved from https://en.wikipedia.org/wiki/Knowledge

Lang, M. (2012). *Investigating the Emotiv EPOC for cognitive control in limited training time.* Unpublished Doctoral Dissertation, University of Canterbury, Canterbury, New Zealand.

LaViola Jr., J. J., Garrity, P., Sottilare, R. A., Williamson, B. M., Brooks, C., & Veazanchin, S. (2015). Using augmented reality to tutor military tasks in the wild. In *Proceedings of the Interservice/Industry Training Simulation & Education Conference*. Orlando, FL.

Minahan, J. (2012, July 3). Anxiety: The hidden disability that affects one in eight children

Retrieved April, 3, 2018, Retrieved from https://www.huffingtonpost.com/jessica.../anxietythe-hidden-disabil_b_1474089.html

Rakus, E. (2011). Motivation in learning. Retrieved April, 3, 2018, Retrieved from https://www.slideshare.net/ElizabethRakus/motivation-in-learning

Scherer, S., Marsella, S., Stratou, G., Xu, Y., Morbini, F., Egan, A., & Morency, L.-P. (2012). Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents* (pp. 455-463). Santa Cruz, CA: Springer.

Shamoun-Baranes, J., Bom, R., van Loon, E. E., Ens, B. J., Oosterbeek, K., & Bouten, W. (2012). From sensor data to animal behaviour: an oystercatcher example. *PloS One, 7*(5), e37997.

Sottilare, R., Hackett, M., Pike, W., & LaViola, J. (2017). Adaptive instruction for medical training in the psychomotor domain. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 14*(4), 331-343.

Sottilare, R. A., & Brawner, K. W. (2018). Exploring standardization opportunities by examining interaction between common adaptive instructional system components. In *Proceedings of the Adaptive Instructional Standards Workshop*. Orlando, FL: U.S. Army Research Laboratory.

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, 1-40.

Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (pp. 1-13).

Sottilare, R. A., & LaViola, J. (2015). Extending intelligent tutoring beyond the desktop to the psychomotor domain. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2015*. Orlando, FL.

The Glossary of Education Reform. (2016, February 18). Student engament. Retrieved April 3, 2018, Retrieved from https://www.edglossary.org/student-engagement/

Xu, Y., & Reitter, D. (2017). Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 623-633). Vancouver, CA.

ABOUT THE AUTHORS

***Dr. Jong W. Kim** is an adaptive tutoring scientist and ORAU post-doctoral researcher at U.S. Army Research Laboratory. Kim received his PhD in Industrial Engineering at Pennsylvania State University, and MS in Industrial Engineering at University of Central Florida. Kim has developed a theory of learning and retention within ACT-R (D2P: Declarative to Procedural) that is being applied to the development of ITSs. He currently investigates a precision-required psychomotor task training in GIFT.*

***Dr. Robert A. Sottilare** leads adaptive training research at the US Army Research Laboratory where the focus of his research is automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems (ITSs). He is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source, AI-based adaptive instructional architecture. He is the lead editor for the Design Recommendations for Intelligent*

Tutoring Systems book series and the founding chair of the GIFT Users Symposia. Dr. Sottilare has authored over 165 technical publications. He is a program committee member and frequent speaker at the Defense & Homeland Security Simulation, Augmented Cognition, and AI in Education conferences. Dr. Sottilare is a member of the AI in Education Society, the Florida AI Research Society, the IEEE Standards Association, and the American Education Research Association. He is a faculty scholar and adjunct professor at the University of Central Florida where he teaches a graduate level course in ITS design. Dr. Sottilare is also a frequent lecturer at the United States Military Academy (USMA) where he teaches a senior level colloquium on adaptive training and ITS design. He has a long history of participation in international scientific fora including NATO and the Technical Cooperation Program. Dr. Sottilare is the recipient of the Army Achievement Medal for Civilian Service (2008), and two lifetime achievement awards in Modeling & Simulation: US Army RDECOM (2012) and National Training & Simulation Association (2015).

Dr. Keith Brawner *is an adaptive tutoring scientist within the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED). He has 12 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current research is in machine learning, cognitive architectures, learning technologies, and ITS architecture. He manages and advises research in adaptive training and architectural programs towards next-generation training.*

Timothy Flowers *is a software engineer at Dignitas Technologies and is one of the programmers responsible for the development of GIFT. He holds a Bachelor's Degree in Computer Science from the University of Central Florida. His work primarily consists of integrating existing systems into GIFT such as sensors and training applications. He acted as the lead GIFT developer for incorporating the Android phone as a sensor in GIFT.*

ACKNOWLEDGEMENTS

The work was supported by the fellowship from Oak Ridge Associated Universities and US Army Research Laboratory to the first author. Ben Goldberg provided numerous comments about the design of the study environment in GIFT.

THEME VIII: AIS STANDARDS

Developing Standards for Adaptive Instructional Systems: 2018 Update

Robert A. Sottilare, Ph.D.
US Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED)
Learning in Intelligent Tutoring Environments (LITE) Lab
Center for Adaptive Instructional Sciences (CAIS)

INTRODUCTION

At a learner modeling expert workshop held at the University of Memphis in 2012, Robson and Barr discussed the potential of lowering the barriers to adopting intelligent tutoring systems (ITSs) through standardization and subsequently wrote a chapter about market needs and standards for learner modeling (Robson & Barr, 2013). Fast forward five years and the University of Memphis and US Army Research Laboratory brought together a group of ITS stakeholders from the IEEE standards association, industry, government, and academia in November 2017 to discuss potential standards across ITSs and other intelligent media that we labeled *adaptive instructional systems (AISs)*. Sottilare & Brawner (2018) define AISs as: *computer-based systems that guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of domain learning objectives*.

Based on the stakeholder meeting in November 2017, the IEEE Learning Technologies Standards Committee established a 6 month AIS standards study group. An essential role of this study group is to engage AIS stakeholders to understand the marketplace needs and identify opportunities to reduce barriers to adoption through standardization. As part of their activity, the AIS study group established four workshops to engage stakeholders:

- First AIS Standards Workshop – 7-8 March 2018, Orlando, Florida
- AIS Standards Workshop at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium – 11 May 2018, Orlando, Florida
- AIS Standards Workshop at the Intelligent Tutoring Systems Conference – 12 June 2018, Montreal, Canada
- AIS Standards Workshop at the Artificial Intelligence in Education Conference – 30 June 2018

The purpose of this paper is to highlight some of the ideas and opportunities for standards identified through the development and conduct of these workshops.

POTENTIAL AIS STANDARDS

This section identifies some of the ideas put forth as opportunities for standards and discusses their merit with respect to the following criteria:

- the idea solves a specific problem identified by AIS developers and/or users

- the idea reduces the time and skill required to develop AISs
- the idea promotes opportunities for interoperability and reuse without negative impact on intellectual property
- the ideas promotes opportunities for new AIS markets or collaboration opportunities

Common AIS Conceptual Model

Robson, Sottilare & Barr (2018) identified the need for an AIS conceptual model including definitions, common components, and functions. They states that "a hierarchical common understanding of the composition of AISs would be useful in communicating ideas in lectures, presentations, and technical papers as well as system specifications". While this proposal will stir some debate, it seems that a common conceptual model of AISs is low hanging fruit that could be implemented quickly as a standard. The likely results of this proposal being a reduction in development time and expanded opportunities for collaboration based on a greater common understanding of AIS design.

AIS Component Interoperability and Reuse

Three workshop papers have suggested standards opportunities based on component interoperability (Sottilare & Brawner, 2018a; Brawner & Sottilare, 2018; Sottilare & Brawner, 2018b). The basis for this proposal is the fact that the literature is fairly consistent in identifying four common components or models within ITSs: domain, learner, instructional (or pedagogical), and user interface. While the functions contained within these components can vary widely among ITSs, the data they exchange and act upon are fairly consistent. Domain models generally provide assessments of the learner's progress toward learning objectives to the learner model. In addition to learner performance, the learner model contains a large number of learner states (e.g., affect, engagement, interest, and preferences). The instructional model receives information about the learner's states and uses this to recommend next steps in the instruction. A user interface model collects information about the learner that can be used to ascertain their current and future states. A standard set of messages could be easily implemented and allow for the swap of one component for another more appropriate or effective component without redesigning the AIS.

Another aspect of component interoperability is AIS compliance with external "standards" like the experience application program interface (xAPI) which generates statements of achievement based on formal and informal education and training experiences (Sottilare, Long & Goldberg, 2017) or the learning tools interoperability (LTI) standard which enables data exchanges with courses in learning management systems (LMSs) like edX, Canvas, and Blackboard to support adaptive massive online open courses (MOOCs; Aleven etal, 2018). Durlach (2018) has suggested that adoption of standards like National Information Exchange Model (NIEM) could also facilitate interoperability. More work is needed to identify potential use cases and existing standards in which we wish AISs to interoperate with.

Learner Modeling Standards

Robson & Barr (2013) mention a previous effort to develop IEEE standards to enable learners to build a personal learner model, to enable developers to provide more personalized instruction, to provide standard sources of data to researchers, to enhance the learner-centric design of instructional systems, and to provide architectural guidance for instruction system designers. However, they also note that this noble effort never resulted in a standard. For this reason, Robson & Barr (2013) have suggested that learner information be standardized, but not the learner model itself or any other model within ITSs.

Several workshop papers have proposed learner modeling standards. Baker & Coleman (2018) have recommended that a yet-to-be-specified set of behavioral models be standardized to represent learner engaged and disengaged behaviors. Biswas & Rajendran (2018) have suggested a three-tiered learner model to represent metacognitive process, cognitive strategies, and cognitive skills. Rus (2018) and Tackett et al (2018) note the need to standardize the representation of a learner's knowledge (prior, current, and predicted).

Another idea discussed during recent AIS standards workshops and meetings has centered on learner records which contain a set of common features that could form the basis of a default learner model. Learner record features could include demographic data, historical records of experience and achievement, and a longer term model of domain competency along with associated models of skill decay. Standard learner record fields would allow systems other than the originating system to read in and interpret learner data in support of new instructional experiences (Robson, Sottilare, & Barr, 2018).

Domain Modeling Standards

Much fewer stakeholders have put forward ideas for standardizing domain models and their associated content, but "it is content and domain modeling that most subject matter experts think about when they create curricula and learning environments" (Hu, Graesser, & Cai, 2018). We believe this points to the need for a methodology to structure domains models and content as a framework in which old domain knowledge and content can be swapped out to the system for old, less effective content. They also recommend that domain models and content be aligned with the developers mental model of the process and be sufficiently specified so as to be functional and effective. McCoy (2018) also suggests a structured domain model based upon hierarchical relationships.

Validation Standards

The idea of validation standards was extracted from Robson, Sottilare, & Barr (2018). "Once standards have been adopted for common conceptual models, component interoperability, and learner record features, we will not only want to validate AIS compliance to those standards but will also want to test their effectiveness, their fit for purpose, and their compatibility with other learning systems. Support for other standards, such as the experience API (xAPI) must be considered, and authors of AISs may desire to evaluate the effectiveness of their systems as a whole or in part to understand how their product stacks up against marketplace expectations for performance and learning effectiveness".

Examples of this type of *testbed* or *quality function* can be found widely. In the 1990's compliance testbeds were established to support interoperability testing for both IEEE 1278 Distributed Interactive Simulation (DIS) standard and IEEE 1516 High Level Architecture (HLA) standard to allow participants in large scale distributed simulation training exercises and experiments to gauge their readiness to be compliant with the standard, interoperable with other federates, and compatible with the simulation information required to be exchanged between federates.

Early in its development, US ARL adopted a testbed function for GIFT to support experimental evaluation of its components to determine whether they met validation criteria. Our experience with GIFT may serve as a model for how we might approach validation, and therefore serve as a guide to normative language in a broader series of standards that address the quality of AISs and their compatibility with other learning systems.

NEXT STEPS

The next steps are to complete the approved workshops and begin to share a work program for our potential IEEE working group. A project authorization request (PAR) has been formulated and submitted to the IEEE Learning Technologies Standards Committee. According to IEEE, “a PAR is a legal document and the means by which a working group assigns copyright to and indemnification from IEEE. Every PAR that is submitted must have a Sponsor to oversee the project. A PAR is a document that states the reason for the project and what it intends to do”. The specific PAR for AIS standards is P2247.1 and the PAR and our request to establish an AIS Working Group are up for approval with an expected decision in July 2018.

ACKNOWLEDGMENTS

A portion of the research described herein has been sponsored by the U.S. Army Research Laboratory. Statements and opinions expressed in this paper do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

Aleven, V., Sewell, J., Popescu, O., Sottilare, R., Long, R., & Baker, R. (2018, June, in press). Towards Adapting to Learners at Scale: Integrating MOOC and Intelligent Tutoring Frameworks. In Proceedings of the Learning @ Scale Conference, London, England, June 26-28, 2018.

Baker, R. & Coleman, C. (2018, March, *in press*). Standardizing Modeling of User Behaviors: Which Behaviors Matter? In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Biswas, G. & Ranjendran, R. (2018, March, *in press*). Standardizing Learner Modeling for Complex Domains Using Multi-Level Learner Modeling Schemes. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Brawner, K. & Sottilare, R. (2018, June, *in press*). Proposing Module-level Interoperability for Intelligent Tutoring Systems (ITSs). In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AISs) Workshop of the 19th International Conference of the Artificial Intelligence in Education (AIED) Conference, London, England, United Kingdom, June 2018.

Durlach, P. (2018, March, *in press*). The Potential of the National Information Exchange Model (NIEM) for Standardizing Adaptive Instructional System Information Exchange and Domain Knowledge. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Hu, X., Graesser, A., & Cai, Z. (2018, March, *in press*). Standardizing Content and Domain Models in Intelligent Tutoring Systems: A Few Ideas. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

McCoy, D. (2018, *in press*). Domain models, student models, and assessment methods: three areas in need of standards for adaptive instruction. In the Adaptive Instructional System (AIS) Standards Workshop of the 14th International Conference of the Intelligent Tutoring Systems (ITS) Conference, Montreal, Quebec, Canada, June 2018.

Robson, R. & Barr, A. (2013). Chapter 2 – Lowering the Barrier to Adoption of Intelligent Tutoring Systems through Standardization. In R. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.) Design Recommendations for Intelligent Tutoring Systems: Volume 1- Learner Modeling. *US Army Research Laboratory*, Orlando, Florida. ISBN 978-0-9893923-0-3.

Robson, R., Sottilare, R. & Barr, A. (2018, June, *in press*). Examining Barriers to the Adoption of IEEE Standards for Adaptive Instructional Systems (AISs). In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AISs) Workshop of the 19th International Conference of the Artificial Intelligence in Education (AIED) Conference, London, England, United Kingdom, June 2018.

Rus, V., Graesser, A., Hu, X., & Cockroft, J. (2018, June, *in press*). A Computational Perspective of Adaptive Instructional Systems for Standards Design. In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AISs) Workshop of the 19th International Conference of the Artificial Intelligence in Education (AIED) Conference, London, England, United Kingdom, June 2018.

Sottilare, R. & Brawner, K. (2018a, March). Exploring Standardization Opportunities by Examining Interaction between Common Adaptive Instructional System Components. In Proceedings of the *First Adaptive Instructional Systems (AIS) Standards Workshop*, Orlando, Florida.

Sottilare, R. & Brawner, K. (2018b, June, *in press*). Component Interaction within the Generalized Intelligent Framework for Tutoring (GIFT) as a Model for Adaptive Instructional System Standards. In the Adaptive Instructional System (AIS) Standards Workshop of the 14th International Conference of the Intelligent Tutoring Systems (ITS) Conference, Montreal, Quebec, Canada, June 2018.

Sottilare, R. A., Long, R. A., & Goldberg, B. S. (2017, April). Enhancing the Experience Application Program Interface (xAPI) to Improve Domain Competency Modeling for Adaptive Instruction. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale* (pp. 265-268), Cambridge, Massachusetts, April 20-21, 2017. ACM. DOI: http://dx.doi.org/10.1145/3051457.3054001.

Tackett, A.C., Cai, Z., Hampton, A.J., Graesser, A., Hu, X., Ramirez-Padron, R., Folsom-Kovarik, J.T., & Copland, C. (2018, June, *in press*). Knowledge Components as a Unifying Standard of Intelligent Tutoring Systems. In the Exploring Opportunities to Standardize Adaptive Instructional Systems (AISs) Workshop of the 19th International Conference of the Artificial Intelligence in Education (AIED) Conference, London, England, United Kingdom, June 2018.

ABOUT THE AUTHOR

***Dr. Robert Sottilare** leads adaptive training research within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). He is ARL's technical lead for the Center for Adaptive Instructional Sciences (CAIS).*

Learning Technology Standards - the New Awakening

Dr. Robby Robson, Eduworks Corporation
Avron Barr, IEEE Learning Technology Standards Committee

INTRODUCTION

Standardization has been a lynchpin of industrialization. It is hard to imagine railroads, automobiles, electric grids, or the Internet without standards for rail gauges, oil viscosity, voltages, internet protocols, and thousands of other things. Organizations such as the International Organization for Standardization (ISO) and IEEE Standards Association (IEEE-SA) have published thousands of standards ranging from highly technical standards that define the inputs and outputs of systems to process and quality standards designed to ensure that goods are produced in a repeatable, auditable fashion.

Yet, within any given discipline or area, the rate and quantity of standardization is often cyclic. The field of learning technology is no different and is currently in the middle of a surge. The years 2017 and 2018 have seen the establishment of new standardization efforts in the areas of competencies, credentials, virtual reality, eBooks, data privacy, learning pathways, and adaptive instructional systems (AIS). This paper examines the forces that have led to this surge, suggests that *learning portability* is the new problem that is the driving force behind this new awakening, and discusses this in the context of efforts launched by the US Army Research Laboratory GIFT project (Sottilare et. al., 2012; Sottilare et al, 2017) to develop standards for AIS.

Disclaimer: ***The views expressed in this paper, including characterizations of standards and standards development organizations, are those of the authors and should not be interpreted as representing the views of any organization.***

INTEROPERABILITY STANDARDS

The standards addressed by this paper are *interoperability standards.* Interoperability standards permit multiple systems or services to work together even if they were designed and manufactured by different vendors – potentially from in different countries and who speak and code in different languages.

As measured by adoption and incorporation into products, *interoperability standards typically succeed because they solve a market-relevant problem,* usually a problem related to supply chains, the cost of production, or market expansion. It is an instructive exercise to think of familiar standards and then to identify the problems they solved and the impact they have had. Standards for weights and measures, for example, solved the problem of parts being produced separately and still fitting the devices that used them. This capability was a pre-requisite for the industrial revolution. The Domain Name System (DNS), as another example, enabled users to refer to servers by names they could remember rather than by meaningless strings of numbers. This convenience was vital to the early growth of the Internet.

Disruptive innovations and technical breakthroughs often lead to new interoperability problems and hence to surges in standardization. Disruptions include new inventions (the steam engine or personal computer), new processes (the assembly line), or new business conditions (the Software-as-a-Service business model). Once these have played out in the market, however, there is often a reduction in the number of high value interoperability problems that must be addressed, so one expects a corresponding reduction in standards activity. This explains the cyclic nature of standards activities mentioned in the introduction and suggests that increases in standardization are often triggered by innovations. Identifying and understanding these innovations can help standards development organizations (SDOs) determine which proposed standards are most likely to have an impact.

Conceptual Interoperability Standards

In software, most interoperability standards address data formats, communications protocols, and system requirements, but there is another type of interoperability that might be termed *conceptual interoperability*. Conceptual interoperability includes shared vocabularies, system architectures, frameworks, and reference models that help producers and consumers communicate effectively. A well-known example is the Open Systems Interconnection (OSI) model developed as the ISO 35.100 series of standards in the 1970s (Wikipedia, 2018). OSI defines the network layers (physical, data link, network, transport, session, presentation, and application) that are used in product manuals, purchasing requirements, engineering courses, and many other places.

As illustrated by OSI, conceptual interoperability standards can stimulate markets. However, for this to occur, there must be a market in place and that market must be experiencing a communication problem. The learning technology field, in contrast, has tended to create conceptual interoperability standards well in advance of the development of markets for the systems they conceptualize. Such standards, while useful for research and academic purposes, rarely have a significant effect on products, consumers, or end-users.

LEARNING TECHNOLOGY STANDARDS (1997 – 2009)

In the view of the authors, the innovation that spurred the development of the first wave of learning technology standards was the web and publicly available web browsers. This led to online and web-based courses and to what is now called “eLearning.” The fundamental interoperability problem that plagued the early eLearning market was **content portability.** At first, the functionality of content depended on the features and functionality of the specific system that delivered it, usually some type of learning management systems (LMS). This market issue had significant economic impact: Organizational consumers were locked into their LMS, and there could be no general eLearning authoring tools or mass-market content distribution, since instructional designers and developers had to develop content for a one LMS at a time.

A wave of standards emerged from c. 1997 – 2009 that separated content from its delivery mechanism. These included:

- Aviation Industry CBT Committee (AICC) set of Computer Managed Instruction (CMI) standards;

- IMS Global Learning Consortium (IMS) Content Packaging and Question and Test Interoperability (QTI) – which was derived from a specification called Question Markup Language (QML) contributed by Questionmark (Questionmark, 1997) – and Common Cartridge standards;

- The IEEE Learning Technology Standards Committee (LTSC) standards for Learning Object Metadata and “Content Object Communication” (derived from the AICC CMI standard) and

- The Shareable Content Object Reference Model (SCORM) published by the Advanced Distributed Learning (ADL) Initiative, which adopted standards from both the IMS and the IEEE LTSC to define a procurement requirement for LMSs and content that ensured interoperability.[3]

These standards catalyzed the growth of a multi-billion-dollar industry (Reuters, 2017) that includes LMS vendors; a variety of eLearning authoring tools (ranging from “rapid eLearning tools” to products such as Authorware™ and Learning Content Management Systems (LCMS)); and both mass-market publishers and bespoke eLearning development companies. The SDOs listed above created many other eLearning standards as well – including standards for competency definitions, learner profiles, architectures, learning systems design, digital rights management – but none of these achieved the same level of adoption as basic content portability, presumably because none of these addressed a pressing market need.

THE NEXT WAVE: ENABLING A SUPPLY CHAIN (2010 – 2016)

The LMS was invented at the time when the prevailing model of the web was as a content delivery system. As the web evolved to a networking tool characterized by social media and eCommerce, learning technology standardization activities related to content portability predictably subsided and work started to focus on specific eLearning supply chain problems. The most active development during this period came from the IMS Global Learning Consortium, which describes itself as serving a “community of educational institutions, suppliers, and government organizations” (IMS Global Learning Consortium, 2018a). From a standards perspective, this constitutes an eLearning supply chain that runs from academic publishers to educational institutions to students via LMSs and associated technologies. IMS Global standards released between 2010 and 2016 included (IMS Global Learning Consortium, 2018b):

- Standards for the accessibility of content by learners with have special needs or are in problematic environments (such as in poorly lit environments or on noisy airplanes), many of which were based on prior work of the accessibility community (IMS Global Learning Consortium, 2012);

- Updates to its Common Cartridge and Question and Test Interoperability standard; and

- A Learning Tools Interoperability standard that originally focused on plugging tools into an LMS but via several updates has been generalized to enable sharing content, user management services, and “launch” messages among tools;

These standards address important problems in the learning delivery ecosystems maintained by institutions and the supply chain that supports them. They improve interoperability and make the sharing of data, services, and tools within each ecosystem more efficient. However, they do not address any problems that are fundamentally different from those addressed by the first wave of content portability standards.

Another significant pre-2017 standards release was the ADL’s Experience API (xAPI) (ADL Initiative, 2018). First released in 2013, xAPI focuses on reporting and sharing the outcomes of learner activities. Developed and branded as “tin can” by Rustici Software, xAPI was incorporated into Rustici’s SCORM engine and SCORM Cloud. This led to immediate adoption, but largely as a substitute for the reporting mechanisms in SCORM. The ADL’s initial vision of xAPI may have included innovations such as

[3] SCORM achieved an unprecedented level of global adoption, in part because one vendor, Rustici Software, provided services that allowed all LMS vendors to implement SCORM consistently and cost effectively.

integration into the Internet of Things (IoT) and supporting a new wave of learning analytics, but as applied, it primarily overcame constraints imposed by SCORM in the corporate (and military) eLearning supply chain.

Other Standards Activities

IMS Global and the ADL were not the only organizations developing learning technology standards and specifications during this period. The IEEE LTSC lay largely dormant but released a Resource Aggregation Model for Learning Education and Training (RAMLET) standard that defines a conceptual model for expressing aggregations of digital learning resources as ontologies and applies this to a variety of other standards, ranging from IMS Content Packaging to the Metadata Encoding and Transmission Standard (METS) (Library of Congress, 2018). ISO/IEC JTC1 SC36, also known as ITLET ("Information Technology for Learning Education and Training"), has published 39 standards documents, including many in the period in question. These range from adoption and adaptations of IEEE LTSC and IMS Global standards to standards that address e-portfolios, learning analytics, competency, virtual experiments, collaborative workplaces, and requirements for e-textbooks.

ITLET standards and RAMLET are primarily conceptual interoperability standards. They consist of abstract frameworks and reference models that identify components of systems architectures and data models but do not define concrete Application Programming Interfaces (APIs) or data formats. They may, of course, prove useful in the future – for example, a new generation of intelligent agents may successfully reason about real-world phenomena using the ontologies defined by RAMLET. Nonetheless, the IEEE or ITLET standards developed prior to 2016 appear to have had little impact on learning technology products. The base RAMLET conceptual model, for example, had a total of 166 fulltext views from IEEE Xplore as of April 2018, whereas the IEEE Learning Object Metadata data model – released the same year – had almost ten times as many (1640) such views, despite being freely and easily available in multiple places on the web (IEEE Xplore, 2018a and 2018b).

CHANGE FACTORS

Until very recently, almost all learning technology standards that achieved adoption in the marketplace focused on concrete interoperability problems associated with the models of eLearning that were established in the early days of the web. These models have undergone refinement and evolution but not revolution, which has often frustrated researchers and technologists (including the authors). From the very start, these innovators envisioned the web as a means to radically transform education and training from a didactic endeavor dominated by classroom lectures into something completely new that is personalized, immersive, intelligent, and far more effective. For example, in 1999, Robson wrote (Robson, 1999):

"Once the content of WWW (sic) pages can interact with other types of records, it is possible to design systems in which content and even the functionality of the interface adapts to the record and preferences of the user and can be easily edited by an instructor or author. Communication could be managed in new and interesting ways not widely used in the university classroom. Authentic scenarios, role playing, virtual realities … all of these could support and could interact with other parts of the on-line learning environment."

Many of the research community's hopes and visions were reflected in conceptual standards or in standards that codified a specialized technical approach. Retrospectively, these had very little chance of being adopted for a variety of reasons, ranging from their lack of real-world market relevance to their overly complex nature that made them impractical to implement. More recently, however, there has been a new awakening, characterized by greatly increased interest and collaboration across multiple relevant industry and government sectors and by a focus on a new set of standards that do not involve content portability. In the view of the authors, the root cause of this awakening lies in the several disruptions that are outlined in this section.

Power to the People!

Over the past decade, social media, online video (YouTube), and increases in possibilities for real-time networking and communication gave rise to MOOCs and other technologies that incorporated connectivism and social constructivist theories of learning (Siemens, 2014), but the ability of learners to network with each other did not fundamentally change the nature of learning content or learning technology. The maturation of eCommerce and the spread of online courses did, however, trigger changes that are having a profound long-term effect on the education and training marketplace. What was once a world of disconnected, institution-bound opportunities from which a learner could select at most one at any given time has gradually become a supermarket that offers multiple concurrent choices. This development has created a consumer market that is shifting buying power and power over the curriculum from institutions to individuals.

Content portability standards benefit institutions by increasing the efficiency of their supply chains, but they do not prevent institutions from locking learners into their LMS, HR, or registrar system. Individuals, on the other hand, have a vested interest in ensuring that they can seamlessly move from one system to another – and receive the outcomes and credentials they seek – regardless of which systems they use or from whom they obtain their education or training. This is creating a demand for sharing data about learner preferences, backgrounds, traits, and achievements across learning systems and across institutional and content provider boundaries – a demand that will only grow with increasing global availability of online opportunities, the “gig economy,” and labor market pressures to facilitate frequent re-training and career changes.

Power to the Technology!

The demand to share new types of data does not only come from market shifts. It is also being amplified by technology drivers, creating a perfect storm of change agents.

- **Cloud-based Web Apps:** The first technology driver is the shift from desktop applications to cloud-delivered web applications. Learning activities are increasingly delivered as web apps running on cloud servers. In this model, content is once again tied to delivery systems, but the web browser has become a standardized and ubiquitous player that works for any cloud-based activity, so content portability is no longer a market problem. At the same time, enterprise cloud applications typically communicate with each other as “services” that exchange relevant data behind the scenes. The relevant data are those that enable learners to transition seamlessly among learning activities and that enable the learning activities to adapt to the learner.

- **The Data Storm:** The second technology driver is the expansion of learning platforms to include mobile phones, tablets, and virtual/augmented reality devices *and the instrumentation of learners with devices that measure their location, movements, biometrics, and instant-by-instant actions.* These platforms *and devices* drastically increase the variety, velocity, and volume of data that is produced by and required by learning systems, a phenomenon that is closely coupled with the third technological driver, which is the emergence of artificial intelligence (AI).

- **Artificial Intelligence:** AI is being applied to education and training in many forms, including machine learning, educational data mining/learning analytics, and natural language processing.

The combination of AI with the increasing availability of location, motion, visual, biometric, and other real-time data – not just from a single learner but from all learners in a learning ecosystem – makes it feasible to develop adaptive instructional systems (AIS) that are active participants in the learning experience. These systems are already posing questions that standards are suited to address, ranging from data and protocol interoperability to questions of privacy, data protection, and the transparency of AI systems' reasoning (Rozenfeld, 2017).

THE STANDARDS RESPONSE

The learning technology standards community is responding to these forces for change:

- The IEEE LTSC is now running several standards activities that have active participation from dozens of companies. The last time that this happened was over 15 years ago, during the early days of the AICC, IEEE LTSC, IMS Global, and ITLET.

- An IEEE Standards Association activity associated with learning technology – the Industry Connections Industry Consortium on Learning Engineering (ICICLE, 2018) – was launched in January of 2018. Its premise is that a cadre of specifically-trained professional engineers will be required to design and support future learning environments. ICICLE had 259 participants, 65 organizations/entities, and nine active special interest groups as of 23 April 2018, less than four months later.

- The IMS Global Learning consortium took over the Open Badge Alliance from the Mozilla Foundation on January 1, 2017 (Badge Alliance, 2017) and is also involved in standards for the exchange of competencies, academic standards, and learning pathways.;

- The Learning Resource Metadata Initiative (DCMI, 2018) and the Credential Engine (Credential Engine, 2018) are focusing efforts on Schema.org, which is an effort supported by the Google, Yahoo, and Yandex search engines;

- An ongoing joint effort started in 2017 is focusing on standards for competencies and credentials. This effort involves multiple standards development groups that represent HR, medical education, corporate training, and formal education;

- Standardization efforts in the eBook-for-education arena are taking place within the IEEE LTSC, ISO/IEC JTC1 SC36, and IMS Global in conjunction with the W3C Publishing Business group that was formed by the merger of the International Digital Publishing Forum (IDPF) with the W3C.

- The US Army Research Laboratory, which has been the home to many activities in Intelligent Tutoring Systems, has launched the Center for Adaptive Instructional Science and is working within the IEEE LTSC to explore related standards. (ARL, 2018)

Many of the above initiatives involve the standardized exchange of data about individual learners. These data range from granular activity reporting to data about learner competencies, preferences, traits, goals, and credentials (including formal credentials and micro-credentials represented by badges). These data are essential for AISs and for AI applications and are important for recruiting, staffing, talent management, and many other activities that intersect strongly with learning technology. Data standards in these fields are being embraced by public-private partnerships (U.S. Chamber of Commerce, 2018), lending added impetus and visibility to the standardization efforts listed above.

THE NEW PROBLEM: LEARNING PORTABILITY

The driving force behind many of these standards is the requirements that the learning experienced by an individual be portable across education, training, HR, staffing, talent management, career guidance, college admissions, and similar systems. In addition, AIS need previously generated learning data to make adaptations, and AI algorithms need to be trained on sufficiently large collections of such data from target populations. All these requirements involve what might best be termed *learning portability,* which extends to the portability of data generated by new learning platforms, e.g. eBooks, VR/AR, advanced forms of simulations, serious games, and AIS. The standards being developed for these platforms are less concerned with the ability to move content across them as with the ability for these platforms to exchange learner data using standardized data services.

PLUS ÇA CHANGE, PLUS C'EST LA MÊME CHOSE

Many learning portability standards projects are continuations of projects started long ago, some before 2000. These include standards ranging from competency and credentialing standards to standards for agent interoperability. The difference is that earlier standardization attempts were conceptual or focused on point solutions developed by small groups or single organizations. At their core, they addressed pedagogical problems rather than business and market problems. This relegated learning technology standards to a sleepy corner of the fast-moving world of digital information and communication technology (ICT).

In one respect, things are now different. The forces for change outlined earlier have brought learning technology interoperability into greater alignment with other ICT issues, while changes in business models and in the power relationship between individuals and institutions are disrupting all education and training market segments on multiple fronts. This combination has awoken the industry to the value and need for a new set of learning technology standards, just as the requirement for content portability did during the last wave of standardization. Nonetheless, there is still a propensity for the learning technology standards community to believe that its mission is to radically improve learning. It may turn out that this works out well as pedagogical problems converge with market problems, but without injecting a keen awareness of the need for market, business, and technological relevance into the standards development process, it is also likely that many standards will be produced that experience low adoption rates.

WHAT THIS MEANS FOR AIS

The context of this paper is an effort being led by the Army Research Laboratory and the GIFT project, to explore standards for AIS. In this context it is important to observe that intelligent tutoring systems (ITS) and other AIS have until recently been self-contained systems designed for desktop use. ITS were developed in response to Bloom’s work on the effectiveness of various modes of instruction that concluded that one-on-one tutoring could realize a two-sigma increase in learning effectiveness over classroom instruction (Boom, 1984) Modeling this form of instruction requires computational power and user interactions that until recently could only be realized using specialized software running on dedicated devices, mostly on desktop computers. These monolithic systems, including GIFT and commercial systems such as Knewton and Carnegie Learning’s tutors, have been small players in the learning technology marketplace (Robson & Barr, 2013) and are not part of the eLearning supply chains addressed by standards.

In addition, the general problem of content portability for AIS seems too complex for standards. Attempts to standardize representations for adaptive content, which arguably include the more complex portions of SCORM's simple sequencing and standards such as IMS Learning Design (IMS Global Learning Consortium, 2013), have proved to be problematic and have not been adopted – and these efforts only scratch the surface of what it would truly mean to have plug-and-play representations of the learning experience delivered by a sophisticated AIS. The problem is that AIS are not about content at all but instead are about guiding a learner through a set of experiences in a personalized manner, which differs radically from the traditional "web as a delivery system" model that is reflected in most content portability standards.

The new problem of *learning portability*, in contrast to content portability, is more aligned with AIS. From the perspective of the learner, the knowledge or skills or credentials gained from an AIS are no different than those obtained from any other system. From the perspective of an AIS, domains and data on a learner's state should be independent of the algorithms it uses. Moreover, as ITSs such as GIFT move to the cloud, they become more capable of exchanging and consuming data of the type associated with the new wave of learning technology standards through standardized APIs. It is therefore reasonable to expect that judiciously developed and adapted learning portability standards can bring AIS into the mainstream of learning technology ecosystems. As a benefit, the will reduce the cost of developing and deploying AIS while providing their proprietary AI algorithms with the data they need to perform better and better adaptations and to improve their effectiveness as learning systems.

REFERENCES

ADL Initiative. (2018). The xAPI Overview. Retrieved 28 April 2018 from https://www.adlnet.gov/research/performance-tracking-analysis/experience-api/.

ARL. (2018). Center For Adaptive Instructional Sciences (CAIS). Retrieved 04 May 2018 from https://www.arl.army.mil/opencampus/?q=centers/cais

Badge Alliance. (2017). The Badge Alliance is now part of IMS Global Learning Consortium. Retrieved 04 May 2018 from http://www.badgealliance.org/

Bloom, B. S. (1984). The 2-sigma problem. *Educational Researcher*, 13(6), 4-16.

Credential Engine. (2018). Credential Engine web site. Retrieved 29 April 2018 from http://www.credentialengine.org/

DCMI. (2018). LRMI Terms. (Learning Resource Metadata Initiative). Retrieved 06 May 2018 from http://dublincore.org/dcx/lrmi-terms/

IEEE Xplore. (2018a). 1484.12.1-2002 - IEEE Standard for Learning Object Metadata. Retrieved 29 April 2018, from https://ieeexplore.ieee.org/document/1032843/

IEEE Xplore. (2018b). 1484.13.1-2012 - IEEE Standard for Learning Technology -- Conceptual Model for Resource Aggregation for Learning, Education, and Training. Retrieved 29 April 2018 from https://ieeexplore.ieee.org/document/6228480/

IMS Global Learning Consortium. (2012). MS Global Access for All (AfA) Primer. Version 3.0 Specification, Public Draft 1.0. Retrieved 06 May 2018 from https://www.imsglobal.org/accessibility/afav3p0pd/AfAv3p0_SpecPrimer_v1p0pd.html

IMS Global Learning Consortium. (2013). Learning Design Specification. Retrieved 06 May 2018 from http://www.imsglobal.org/learningdesign/index.html

IMS Global Learning Consortium. (2018a). About the IMS Global Learning Consortium. Retrieved 28 April 2018, from https://www.imsglobal.org/aboutims.html.

IMS Global Learning Consortium. (2018b). Download IMS Interoperability Standards. Retrieved 28 April 2018, from http://www.imsglobal.org/specifications.html.

ICICLE. (2018). ICICLE web site. Retrieved 29 April 2018, from https://www.ieeeicicle.org/.

ISO. (2018). Standards Catalog for ISO/IEC JTC1 SC36. Retrieved 29 April 2018 from https://www.iso.org/committee/45392/x/catalogue/p/1/u/0/w/0/d/0.

Library of Congress. (2018). METS Website. Retrieved on 29 April 2018 from http://www.loc.gov/standards/mets/.

Questionmark. (1997). QML (Question Mark-up Language) Overview 0.9.5. Retrieved 06 May 2018 from https://www.questionmark.com/sites/default/files/PDF/overview.doc.

Reuters. (2017). Global E-Learning Market 2017 to Boom $275.10 Billion Value by 2022 at a CAGR of 7.5% – Orbis Research. Retrieved 29 April 2018 from https://www.reuters.com/brandfeatures/venture-capital/article?id=11353.

Robson, R. (1999). WWW-based course-support systems: The first generation. *International Journal of Educational Telecommunications*, 5(4), 271-282.

Robson, R., & Barr, A. (2013). Lowering the Barrier to Adoption of Intelligent Tutoring Systems through Standardization. Design Recommendations for Intelligent Tutoring Systems: Volume 1-Learner Modeling, 7 - 13.

Rozenfeld, M. (2017). Seven IEEE Standards Projects Provide Ethical Guidance for New Technologies. *The Institute*, 5 May 2017. Retrieved 28 April, 2018 from http://theinstitute.ieee.org/resources/standards/seven-ieee-standards-projects-provide-ethical-guidance-for-new-technologies.

Siemens, G. Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learni*ng, 2(1) (2005), 3-10.

Sottilare, R., Brawner, K., Goldberg, B. & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.

Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: *US Army Research Laboratory*. May 2017. DOI: 10.13140/RG.2.2.12941.54244.

U.S. Chamber of Commerce. (2018). U.S. Chamber Foundation to Develop Job Registry Project with Support from Google.org and JPMorgan Chase & Co. Retrieved 29 April, 2018 from https://www.uschamberfoundation.org/press-release/us-chamber-foundation-develop-job-registry-project-support-googleorg-and-jpmorgan

Wikipedia. (2018). OSI model. Retrieved April 29, 2018 from https://en.wikipedia.org/wiki/OSI_model.

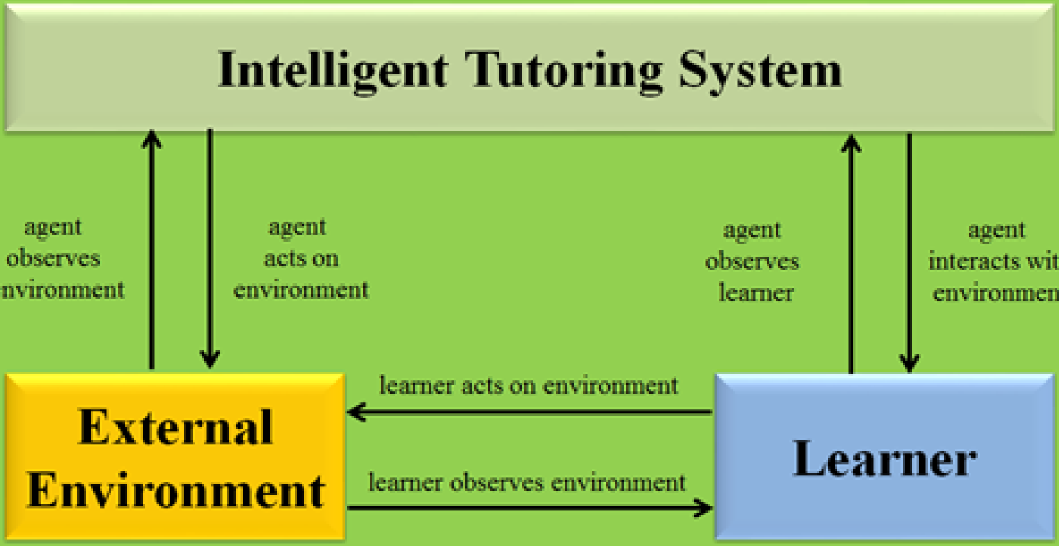
ABOUT THE AUTHORS

__Dr. Robby Robson__ is co-founder and CEO of Eduworks. He is a long-time contributor to learning technology research, products, and standards who has held leadership positions in multiple standards development organizations including the IEEE Learning Technology Standards Committee, the IMS Global Learning Consortium, ISO/IEC JTC1 SC36 (Information Technology for Learning, Education, and Training), and the Learning Resource Metadata Initiative. His current focus outside of standards is the development of global infrastructure for competency-based education and training – see www.cassproject.org.

__Avron Barr__ started his career as an editor of The Handbook of AI (a seminal 3-volume book on Artificial Intelligence) and a founder of Teknowledge, an early AI startup. Since Teknowledge was sold in 1986, he has been an independent consultant in Silicon Valley, helping innovators understand, explain, and market technologically-advanced software. He currently consults for the Institute for Defense Analyses, helping evaluate the US Advanced Distributed Learning (ADL) Initiative's Total Learning Architecture, and is serving as chair of the IEEE Learning Technology Standards Committee.

Proceedings of the Sixth Annual GIFT Users Symposium

GIFT, the Generalized Intelligent Framework for Tutoring, is a modular, service-oriented architecture developed to lower the skills and time needed to author effective adaptive instruction. Design goals for GIFT also include capturing best instructional practices, promoting standardization and reuse for adaptive instructional content and methods, and methods for evaluating the effectiveness of tutoring technologies. Truly adaptive systems make intelligent (optimal) decisions about tailoring instruction in real-time and make these decisions based on information about the learner and conditions in the instructional environment.



Intelligent Tutoring System

agent observes environment

agent acts on environment

agent observes learner

agent interacts with environment

External Environment

learner acts on environment

learner observes environment

Learner

The GIFT Users Symposia began in 2013 to capture successful implementations of GIFT from the user community and to share recommendations leading to more useful capabilities for GIFT authors, researchers, and learners.

About the Editor:

Robert Sottilare, Ph.D. *is the Adaptive Training Research Lead at the US Army Research Laboratory where the focus of his research is automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems (ITSs). He is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source, AI-based adaptive instructional architecture. He is a faculty scholar and part-time professor at the University of Central Florida where he teaches a graduate level course in ITS design. Dr. Sottilare is also a frequent lecturer at the United States Military Academy (USMA) where he teaches a senior level colloquium on ITS design.*

Part of the Adaptive Tutoring Series

9 780997 725759