

# Proceedings of the Eleventh Annual GIFT Users Symposium

May 2023  
Orlando, Florida  
and Hybrid



**GIFT**

*Edited by:*  
*Anne M. Sinatra*

**Part of the Adaptive Tutoring Series**

**Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)**

**Proceedings of the 11<sup>th</sup> Annual  
Generalized Intelligent Framework  
for Tutoring (GIFT)  
Users Symposium  
(GIFTSym11)**

*Edited by:  
Anne M. Sinatra*

**Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)**

Copyright © 2023 by the U.S. Army Combat Capabilities Development Command – Soldier Center.

**Copyright not claimed on material written by an employee of the U.S. Government.  
All rights reserved.**

**No part of this book may be reproduced in any manner, print or electronic, without written  
permission of the copyright holder.**

*The views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Army  
Combat Capabilities Development Command – Soldier Center*

Use of trade names or names of commercial sources is for information only and does not imply endorsement  
by the U.S. Army Combat Capabilities Development Command – Soldier Center.

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 Combat  
Capabilities Development Command – Soldier Center, 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, June 2023

*U.S. Army Combat Capabilities Development Command  
Soldier Center  
Orlando, Florida*

International Standard Book Number: 978-0-9977258-4-1

*We wish to acknowledge Alexandra Lutz for her effort in collecting the book chapters, confirming that  
chapters were in format, and checking/correcting references.*

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

# CONTENTS

<b>From the Editor</b> .....	<b>V</b>
<b>Theme I: New GIFT Features and Future Learning Concepts</b> .....	<b>1</b>
The GIFT Architectural and Features Update: 2023 Edition .....	3
<i>Nicholas Roberts, Thomas Lenz, and Benjamin Goldberg</i>	
Supporting Future Learning Concepts: GIFT in the Year 2040.....	19
<i>Benjamin Goldberg and Randall Spain</i>	
<b>Theme II: Measurement and Assessment</b> .....	<b>29</b>
Investigating the Effect of Realistic Agents on Team Learning in Adaptive Simulation-based Training Environments using GIFT .....	31
<i>Robert Sottolare, Christopher McGroarty, Christopher Ballinger, and Timothy Aris</i>	
Using GIFT to Develop an Adaptive Distributed Learning Environment Supporting Data Science Competencies.....	41
<i>Fadjimata Anaroua, Hong Liu, and Naomi Malone</i>	
<b>Theme III: Adaptive Instructional System (AIS) Architecture and Ontology</b> .....	<b>51</b>
Cloud Technology Evolutions and Their Influences to GIFT and the STEEL-R Cloud.....	53
<i>Ruth Feliz, Michael Kalaf, Christopher Meyer, Nicholas Roberts, and Benjamin Goldberg</i>	
Defining and Measuring Competencies under Varying Conditions.....	65
<i>Robby Robson, Fritz Ray, Benjamin Goldberg, Shelly Blake-Plock, Cliff Casey, Kevin Owens, and         Sazzad Nasir</i>	
The Evolution of an Experience Design Tool: From Live Exercises to Synthetic Experiential Learning with the Experience Design Tool and Integration with GIFT.....	73
<i>Sanjay Mishra, Kevin Owen, and Benjamin Goldberg</i>	

The Value Proposition of GIFT for the Field of Learning Engineering .....87  
*Shelly Blake-Plock, Kevin Owens, and Jim Goodell*

**Theme VI: Authoring Tools and Interfaces.....99**

A Comparative Analysis Interface to Streamline After-Action Review in  
Experiential Learning Environments .....101  
*Caleb Vatrak, Naveeduddin Mohammed, Gautam Biswas, Nicholas Roberts, and Benjamin Goldberg*

ChatGPT in the Generalized Intelligent Framework for Tutoring .....109  
*Faruk Ahmed, Keith Shubeck, and Xiangen Hu*

Flexible and Live Adaptive Training Tools (FLATT) – Updates and  
Improvements .....115  
*Michael Cambata, Mitchell Gorsd, Thomas Lenz, and Nicholas Roberts*

**Theme V: Collective/Team Based Methods .....125**

Improving Dialogue Classification Models to Support Generalizable Team  
Communication Analytics in GIFT .....127  
*Jay Pande, Stephen Paul, Wookhee Min, Randall Spain, and James Lester*

Measurement Error and the Generalizability of Competency Assessments .....137  
*Grace Teo, Eric Sikorski, Michael King, and Jennifer Solberg*

Supporting Data-Driven Team Feedback and Scenario Adaptations in GIFT...145  
*Andy Smith, Randall Spain, Nicholas Roberts, Benjamin Goldberg, Jonathan Rowe, Bradford Mott, and  
James Lester*

Refining a Team Training Approach to Prepare Command Staff for Wargaming  
.....155  
*Randy Jensen, Evan Finnigan, Grace Teo, and Lisa N. Townsend*



# **FROM THE EDITOR**

**Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)**



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

Welcome to the Proceedings of the 11<sup>th</sup> Annual GIFT User Symposium! This year we are celebrating 11 years of GIFT Symposiums and have accepted 15 papers for publication. All of the presentations that occurred at GIFTSym11, and the papers in this volume show the versatility of the Generalized Intelligent Framework for Tutoring (GIFT), and the work that is being done with GIFT.

GIFT is an open-source intelligent tutoring system (ITS) architecture that is freely available online at [GIFTtutoring.org](http://GIFTtutoring.org). There are both Cloud and Downloadable version of GIFT. GIFT has been developed with multiple goals in mind including supporting ITS research, and simplified creation of ITSs and Adaptive Instructional Systems (AISs).

Our fantastic team, and our program committee did a great job supporting the development of GIFTSym11, reviewing papers, and assisting with the facilitation of the event this year. We want to recognize them for their efforts:

- **Benjamin Goldberg**
- **Gregory Goodwin**
- **Michele Myers**
- **Alexandra Lutz**
- **Randall Spain**
- **Lisa N. Townsend**

We were very pleased to have GIFTSym11 return to being an in-person event this year! Additionally, this was our first year offering a hybrid option for attendees. We are very happy that both modalities were well attended!

The themes for this year's GIFTSym include:

- New GIFT Features and Future Learning Concepts
- Measurement and Assessment
- Adaptive Instructional System (AIS) Architecture and Ontology
- Authoring Tools and Interfaces
- Collective/Team Based Methods

The editor and program committee would like to thank all of the contributions and authors on the papers in this proceedings. We also would like to thank all those who contributed to GIFTSym in previous years. The feedback, lessons learned, suggestions and research that have been provided from GIFTSym through the years have been important in the development of GIFT.

We would also like to encourage readers to visit the documents tab on [www.GIFTtutoring.org](http://www.GIFTtutoring.org). Proceedings from each year of GIFTSym as well as the Design Recommendations for Intelligent Tutoring Systems book series are available for free on the documents tab. We also encourage users to sign up for a free GIFTtutoring.org account so that they can receive GIFT news, access to the user forum, and access the GIFT software.

Thank you for 11 great years of GIFTSym!

Anne M. Sinatra, Ph.D.  
GIFTSym11 Chair and Proceedings Editor

**Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)**



**THEME I: NEW GIFT  
FEATURES AND FUTURE  
LEARNING CONCEPTS**

**Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)**

# The GIFT Architectural and Features Update: 2023 Edition

Nicholas Roberts<sup>1</sup>, Thomas Lenz<sup>1</sup>, and Benjamin Goldberg<sup>2</sup>

<sup>1</sup>Dignitas Technologies; <sup>2</sup>US Army Combat Capabilities Development Command (DEVCOM) – Soldier Center

## 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 regarding the architectural “behind the scenes” work and how the GIFT project is addressing user requirements suggested in the previous GIFTSym10 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 previous GIFT Symposium architecture updates (Brawner & Ososky, 2015; Ososky & Brawner, 2016; Brawner et al., 2017; Brawner & Hoffman, 2018; Brawner et al., 2019; Goldberg et al., 2020; Hoffman et al., 2021; Hoffman et al., 2022) this version highlights new tools and feature requests accomplished over the latest development cycle. The feature requests and derived architectural improvements are derived from two primary sources: (1) symposium paper recommendations collected across the GIFT user base, and (2) stakeholder interactions linked to capability and project needs. The features are organized into logical sections within this update and cover modifications across all core modules operating within GIFT.

## WELCOME

---

Welcome back for our yearly GIFT update! For those new to the community, there are a number of recommended resources that will help to orient you to this project and ecosystem of capability. GIFT has come a long way since its original goals were defined in its description paper (Sottolare et al., 2012). First, we would encourage you to simply start by exploring the available resources, as the tools and example courses have been designed to assist users in learning about GIFT’s tools and methods for the purpose of creating Adaptive Instructional Systems. It also highlights various domains and use cases that were used to guide GIFT’s development (Goldberg & Sinatra, 2023). Outside of the introductory materials and tutorials available in GIFT, there is also developer support through detailed documentation and active help forums. 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 1700 postings and responses.

Documentation has been made freely available online at <https://gifttutoring.org/projects/gift/wiki/Documentation>, with interface control documentation available at [https://gifttutoring.org/projects/gift/wiki/Interface\\_Control\\_Document\\_2022-1](https://gifttutoring.org/projects/gift/wiki/Interface_Control_Document_2022-1), and a developer guide available at [https://gifttutoring.org/projects/gift/wiki/Developer\\_Guide\\_2022-1](https://gifttutoring.org/projects/gift/wiki/Developer_Guide_2022-1). These documents are updated each software release. In this release, we would also like to highlight the available instructions for hosting your own Amazon Web Services (AWS) instance ([https://gifttutoring.org/projects/gift/wiki/Amazon\\_Web\\_Service\\_Install\\_Instructions](https://gifttutoring.org/projects/gift/wiki/Amazon_Web_Service_Install_Instructions)).

## GIFT Development and Release Strategy

Our goal from the start was to develop GIFT through a community-driven approach by making functions, components, and source code available. There are two GIFT instances available to everyday users, GIFT Cloud and GIFT Desktop. GIFT Cloud follows an every-Friday system update schedule when relevant updates are ready from the engineering team. For the desktop version, we have maintained a near 12-month release cycle with a recent regression tested version in November 2022. To support experimentation, intermittent extensions of the core GIFT baseline are performed to facilitate data and interaction requirements based on specific research questions of interest. These are performed on an “as needed” basis, and often serve as the feature extensions included in the next public release. Adjustments to the release strategy will be considered as more agile software development approaches are being applied at the organizational and enterprise level. As a member of the community, if you see a feature in the cloud release which you would like to use locally, simply ask over a forum post or email to the authors.

## Cloud General Reporting

GIFT Cloud (see Figure 1) has been running continuously for the last seven years over AWS. The cloud instance is kept online and updated in advance of the downloadable version, meaning that cloud content must be backwards-ported to be compatible with the perpetually out of date offline version. We do our best to keep the downloadable version to regularly scheduled improvements, but, for ordinary users, we would encourage you to use the Cloud version. It supports hundreds of simultaneous users for experiments. We are generally confident in the systems’ ability to stay up and cope with demand. The current limitations are that team training in a virtual environment and sensor-based interactions are not supported on the cloud instance, but that requirement will be addressed.

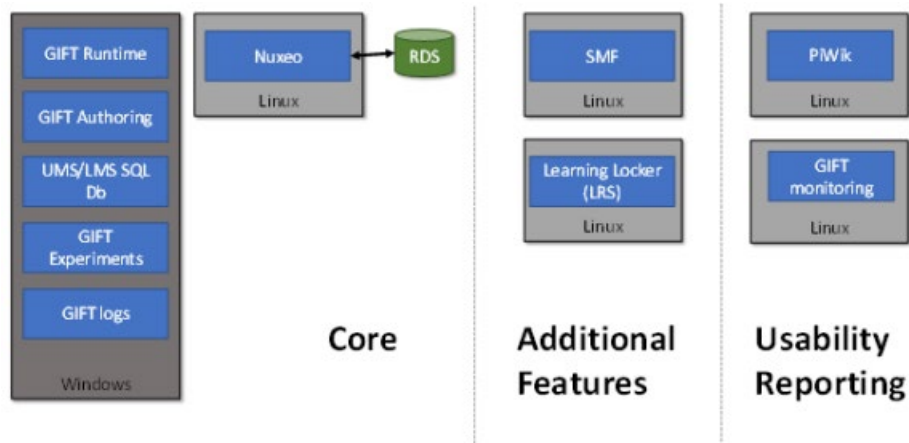


Figure 1. Simple Diagram Overview of GIFT Cloud Items

Behind the scenes, however, the re-tooling to move to a deployment version of development in a desktop instance to a cloud environment in production has been working well. For the remainder of the paper, we will cover the latest improvements added over the last development cycle.

## NEW GIFT FEATURES AND UPDATES

---

Since the last feature update from GIFTSym10 (Hoffman et al., 2022), there have been multiple additions to the GIFT capability set. Each tool or method described in this section is now available in the latest public-

facing open-source version of GIFT or on GIFT Cloud. Each new feature will be presented with information on the functions it supports and the system and data level dependencies to implement.

## xAPI Improvements

As with the previous version, the current 2022-1 release has continued to expand upon GIFT’s support for persistent modeling and competency-based training. This has seen the addition of new xAPI statements to GIFT’s custom xAPI profile for the Total Learning Architecture’s (TLA’s) Competency and Skill System (CASS), as well as accompanying changes to GIFT’s Learning Management System (LMS) Module. For more information about GIFT’s xAPI profile and the xAPI statements that GIFT produces, visit [https://giftutoring.org/projects/gift/wiki/XAPI\\_Statements\\_2023-1](https://giftutoring.org/projects/gift/wiki/XAPI_Statements_2023-1).

## Stress and Difficulty

The last release’s logic to support stress and difficulty values in GIFT’s xAPI profile has been expanded upon to allow authors to configure these values in the Course Creator. When creating or editing a task in the Course Creator, the initial difficulty value associated with a task can be changed by selecting the new “Easy”, “Medium”, and “Hard” buttons that have been added to the task editor. The task’s initial stress can also be adjusted between 0 and 1 by dragging the slider underneath the difficulty level buttons (see Figure 2).

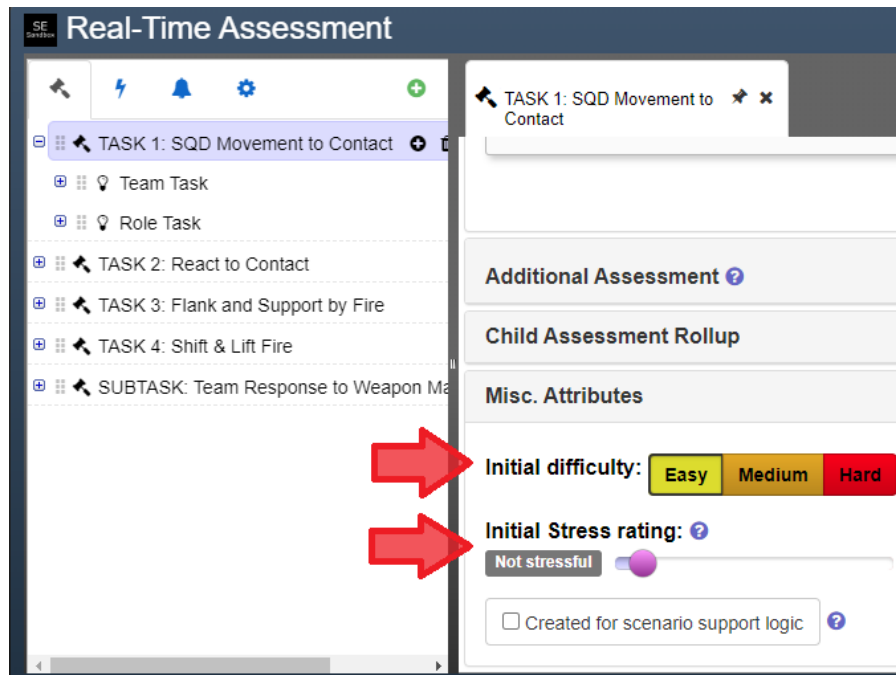


Figure 2. Task Initial Stress and Difficulty

Similarly, when creating or editing a strategy, the strategy editor provides two sliders, one for stress and one for difficulty, that both can be used to select values between -1 and 1 (see Figure 3).

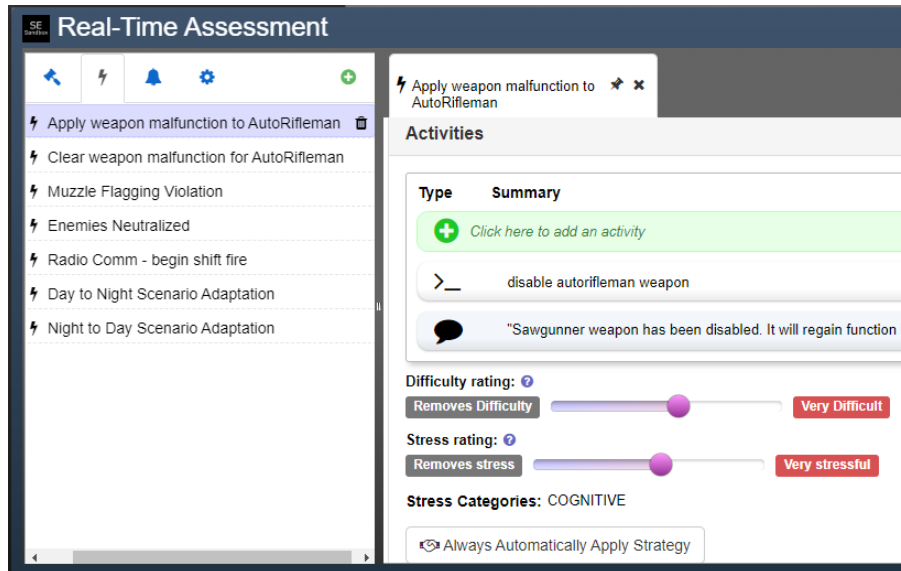


Figure 3. Strategy Stress and Difficulty Rating

If a strategy is given non-zero stress or difficulty values using this editor and then is performed during a task, then GIFT will recalculate the stress and difficulty values of the task based on what values were entered for the strategy. As an example, suppose a task named “Task 1: SQD Movement to Contact” was given an initial difficulty of Medium and an initial stress of 1 and that a strategy named “Apply weapon malfunction to AutoRifleman” was given a difficulty rating of 1 and a stress rating of -1. If the task was run during a course and the strategy was invoked once, the resulting xAPI statements would collectively represent how the stress and difficulty values of this task changed over time, adding the strategy’s stress and difficulty adjustments to the task. This would make the task’s final difficulty value “Hard”, since the rating of 1 from the strategy was added to the initial difficulty of “Medium”, and the final stress value 0, since -1 was added to the initial stress of 1. It is worth noting that current practice in defining stress and difficulty is subjective and dependent on subject matter expert input. Future work aims to operationalize stress and difficulty values based on conditions and standards applied in a scenario or problem set.

### *Cognitive and Affective Learner State*

GIFT now produces additional xAPI statements to share Cognitive and Affective measurements from its internal learner states. Performance measurements from these learner states, such as formative and summative assessments, were added to GIFT’s xAPI profile in a previous release, but the Cognitive and Affective states were left out until now. The addition of these xAPI statements means that measurements of Cognitive learner state attributes, such as Knowledge measurements from Adaptive Courseflow objects and Skill from Real-time assessment activities (e.g., training application course objects and Adaptive Courseflow practice), will now be sent out as xAPI statements to a connected Learner Record Store (LRS). Similarly, Affective learner state attributes like Engagement, Arousal, and Frustration that can be collected using survey course objects and sensors will also be written out to xAPI statements. This establishes a strategy for integrating affective and cognitive state classifiers into a data model schema used to control context tracking and task assessments, and enables auto-generation of xAPI to better report learner experiences within an experiential training event.



## RIDE Training Application

The [Rapid Integration and Development Environment \(RIDE\)](https://ride.ict.usc.edu/) developed by the USC Institute for Creative Technologies (ICT; visit <https://ride.ict.usc.edu/> for more information)) has been integrated into the latest release of GIFT. Like SE Sandbox, which was integrated in GIFT’s 2021-2 release, RIDE is designed to be game engine agnostic. The middleware components of RIDE were leveraged to re-create a Unity-based “React to Fire (Forest)” scenario. This RIDE scenario has been released to the GIFT portal for the community to use with the GIFT 2022-1 release. Following the standard development approach for integrating any new training applications (see [https://giftutoring.org/projects/gift/wiki/Developer\\_Guide\\_2022-1](https://giftutoring.org/projects/gift/wiki/Developer_Guide_2022-1)), a new RIDE course object has been added to the Course Creator to allow RIDE to be used during GIFT courses, and a new Gateway module interop plugin class was introduced to support the networking communication between GIFT and RIDE (see Figure 4).

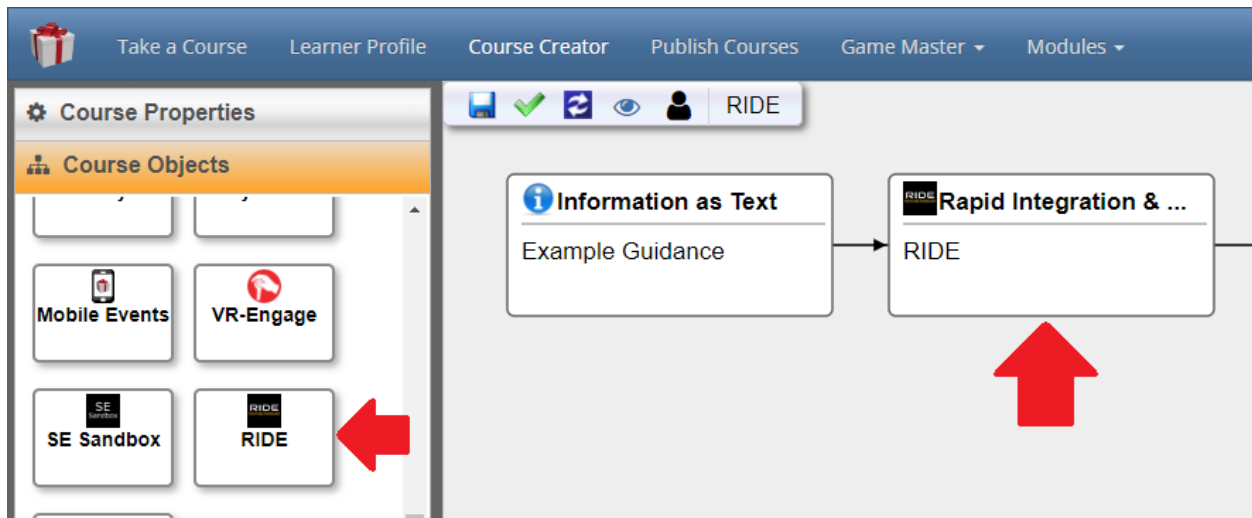


Figure 4. RIDE Course Object in Course Creator

One of the goals of this integration was to preserve and expand upon the existing behavior of the SE Sandbox interop plugin. As a result, the RIDE plugin uses the similar gRPC communication and Google Protocol Buffer messages to support the same automated condition assessment logic within GIFT. The RIDE scenario included with this release is paired with 3 GIFT courses that can be downloaded from the Downloads page on the GIFT Portal. The first of these, “Set Up Your Machine for STEEL-R RIDE”, provides guidance on how to set up a machine to use RIDE with GIFT. The following courses, “STEEL-R RIDE Individual Exemplar” and “STEEL-R RIDE Team Exemplar” showcase GIFT’s interactions with RIDE during a single-player and team-based training experiences. The individual course provides instructions to trigger automated condition assessments within GIFT based on a learner’s actions in RIDE.

These automated assessments take into account the learner’s position, aiming direction, weapon fire, and pace during the training scenario and use these metrics to assess the learner’s performance, which is then used to trigger scenario adaptations to provide feedback and additional directions to the learner. The remaining team exemplar course allows up to 4 players, assessed within GIFT as a single team, to participate in a RIDE training scenario together, with an additional user required to act as an Observer Coach/Trainer (OC/T) to monitor the session in Game Master. If GIFT is configured to connect to an LRS instance, the individual and team exemplar courses will produce xAPI statements based on the automated assessments that occur during the course. Some of the possible statements that can be produced

require the OC/T in Game Master to provide a manual assessment, which is shown in Game Master using yellow flashing indicators around the concepts that are being assessed.

## Experience Training Support Package (xTSP) Import

GIFT's support for importing Experience Training Support Package (xTSP; see Hernandez et al., 2022)) files has been expanded upon from what was announced at last year's GIFTSym. As a reminder, an xTSP is a JSON file that defines a structure of tasks, standards, and conditions that can be imported by GIFT's Course Creator to create a Domain Knowledge File (DKF) for a training application course object (see Figure 5).

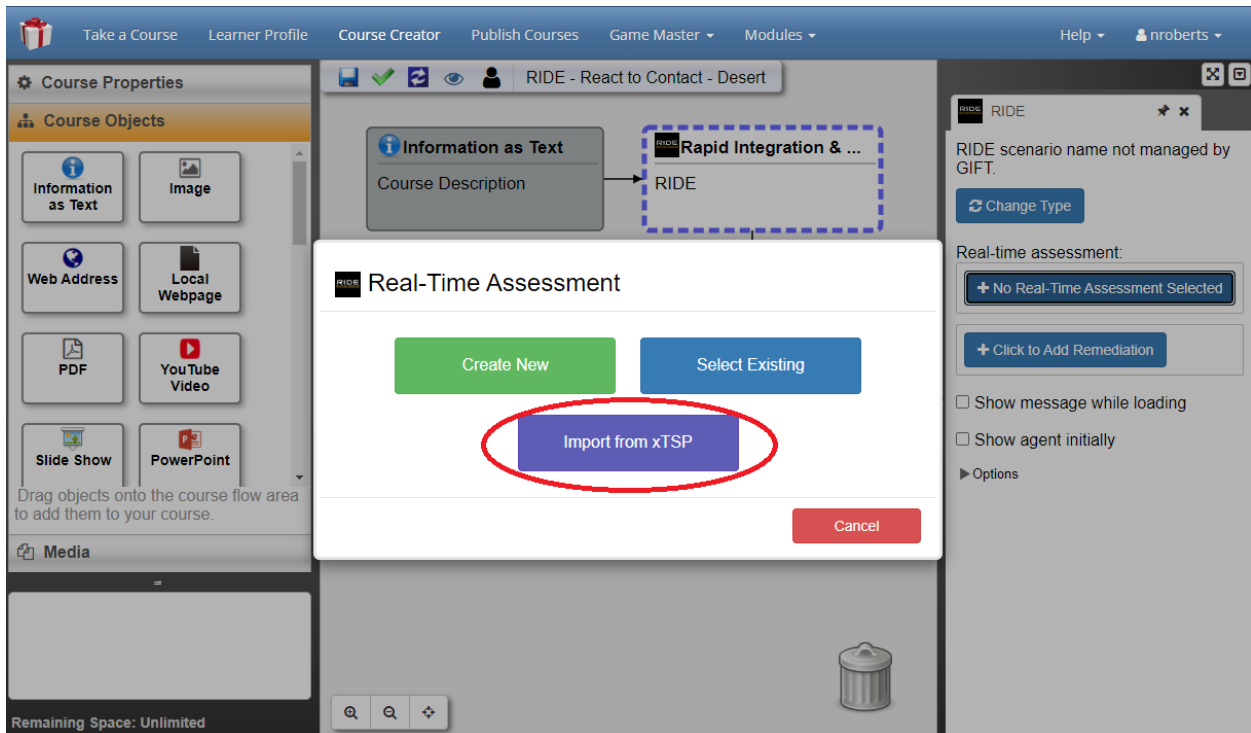


Figure 5. Importing Real Time Assessment from xTSP in GIFT Authoring Tool

In the current release, GIFT now supports version 0.9.7.2 of the xTSP schema and can import xTSP files using that version. Importing an xTSP to a course will also now update the course concepts to match the task structure within the xTSP, ensuring that DKFs created from xTSP files can produce summative assessments and xAPI statements. Together, these changes can be used to more easily create DKFs from scratch and allow those DKFs to generate xAPI statements.

## Game Master Updates

Game Master's interfaces for viewing and editing a past knowledge session's summative assessments have been improved upon. First, the Calculate Score dialog that is used to change the summative assessment has seen a design overhaul. Previously, if an OC/T inspecting a past session wanted to change a concept's summative assessment, they would need to manually enter scores for every condition underneath that concept, which would combine those scores to calculate the summative score of the concept. Now, concepts that only have conditions underneath them can have a summative assessment applied to them by clicking

the stars next to them, removing the guesswork on how a concept's child condition scores will adjust the summative assessment (see Figure 6).

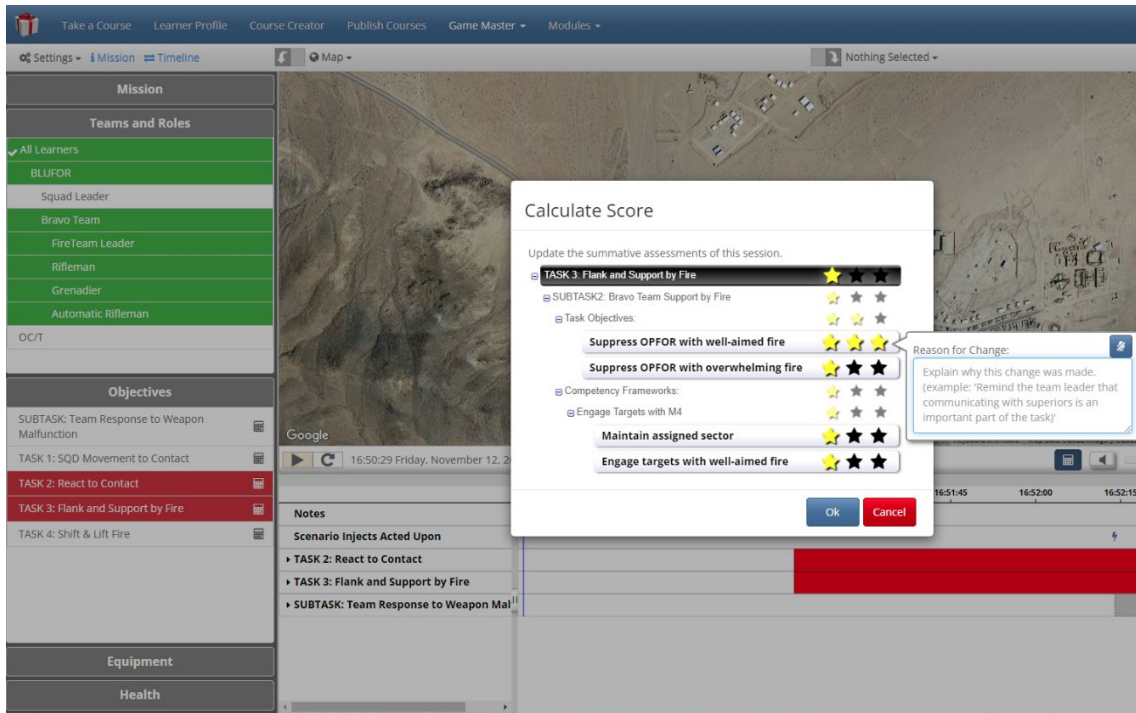
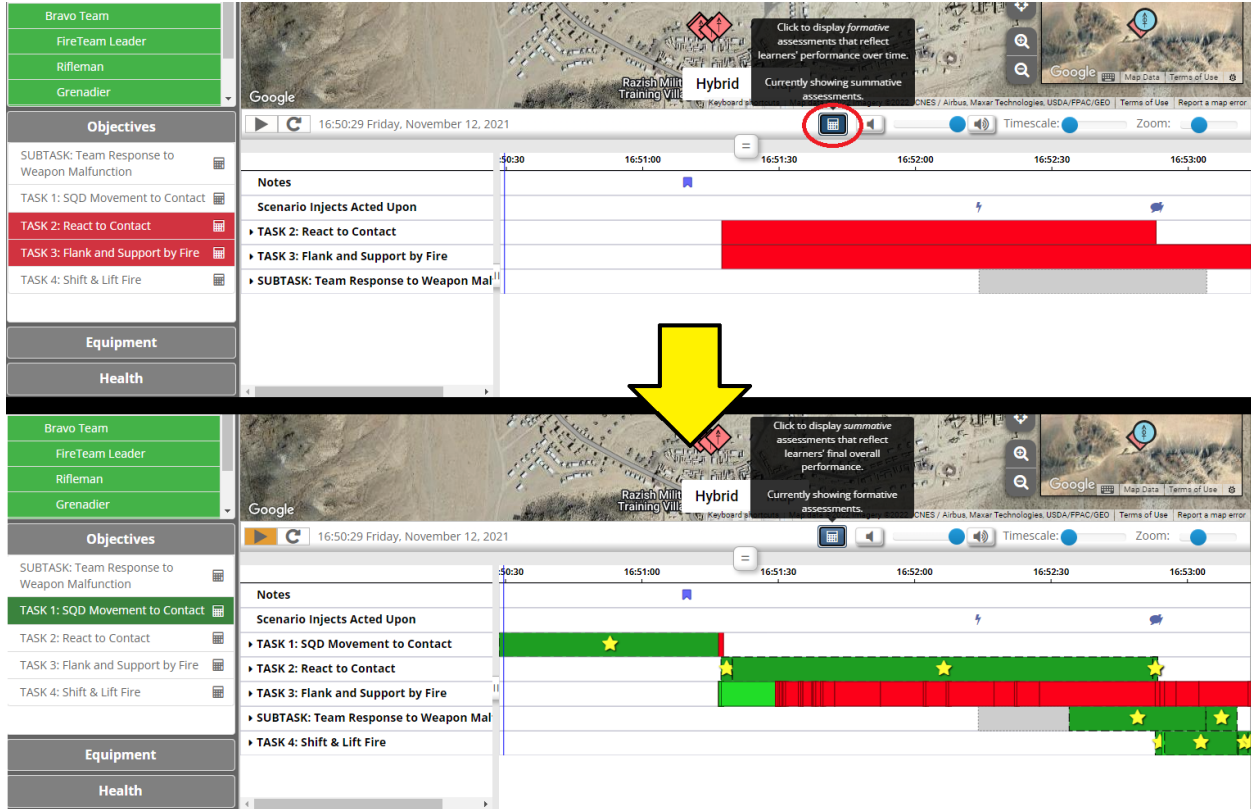


Figure 6. Editing the Summative Assessments for a Task

Changing a concept's summative assessment will “bubble-up” that concept's assessments to other “parent” tasks and concepts that are higher in the hierarchy, so assigning summative assessments to all of a task's child concepts will automatically update that task's summative assessment as well. While changing a summative assessment, the OC/T can also use the nearby text box to enter text describing the reason for the change, or they can make an audio recording of the reason for the change as an alternative. Notably, the timeline that is shown in Game Master when viewing a past session will now show the session's summative assessments by default, whereas it previously displayed the automated formative assessments that may have generated them. A session's formative assessments can still be shown in the timeline by clicking on a new button above the timeline to switch between showing summative and formative assessments (see Figure 7).

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)



**Figure 7. Switching the Timeline between Formative and Summative Assessments**

The aforementioned stress and difficulty values that are associated with tasks and strategies are also now displayed in Game Master. These values are updated in real-time during active knowledge sessions as strategies make changes to a task's initial stress and difficulty, allowing the OC/T to easily view a task's currently associated values in Game Master's Assessment panel (see Figure 8).

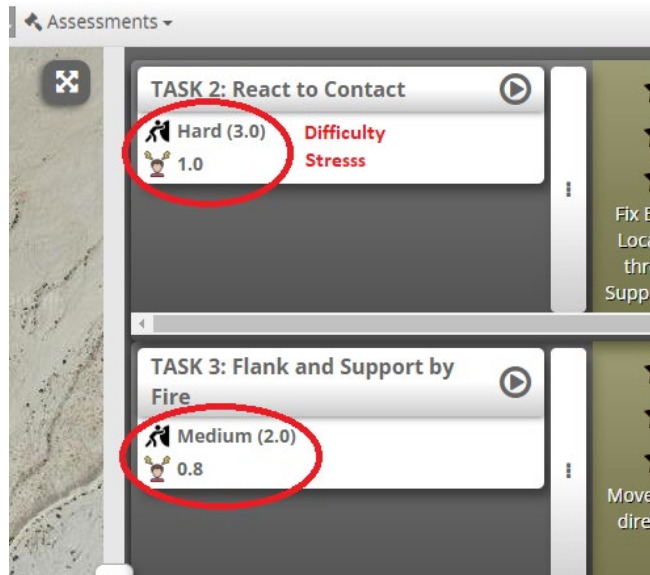


Figure 8. Viewing Tasks' Stress and Difficulty Values in Game Master

Similarly, the degree to which a strategy adjusts a task's stress and difficulty is also shown in the Scenario Injects panel in Game Master (see Figure 9).

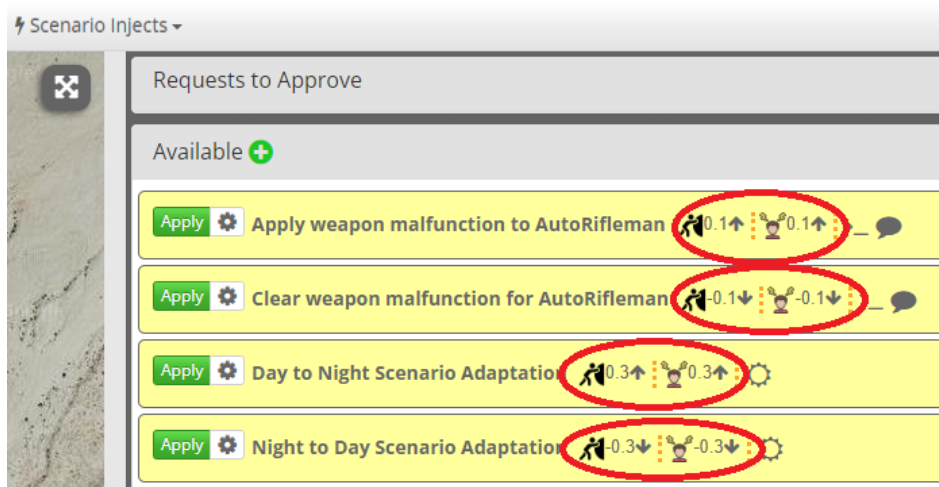


Figure 9. Viewing strategies' Stress and Difficulty Value Adjustments

Users in Game Master can also adjust the new "Map Icon Size" setting in the Settings panel to change the visual scale of map icons (see Figure 10). Using this slider, map icons can be made as small as ½ their normal size or as large as 2 times their normal size.





Figure 10. Changing the Size of Map Icons in Game Master

## Web Monitor

A web-based version of GIFT’s Monitor tool has been implemented in the GIFT Dashboard for use in headless operating systems and servers. The Monitor tool in GIFT’s Control Panel has long been used as a management tool for controlling GIFT’s modules and a debugging tool for developers interacting with GIFT’s internal messaging system, but it is limited in the sense that it can only be displayed in environments that support user interface elements, which some servers do not. The Web Monitor exposes this same functionality through GIFT’s webpages, allowing this Monitor functionality to be used on platforms where it would otherwise be unsupported. The Web Monitor can be accessed by clicking on the new Modules tab in the GIFT Dashboard’s header. The Status option underneath this tab will show the running status of GIFT’s modules and provide controls to start and stop them (see Figure 11), while the Messages option under this tab will display GIFT’s system messages and any messages that are being sent out as part of a domain session, i.e. a running course (see Figure 12).

# Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

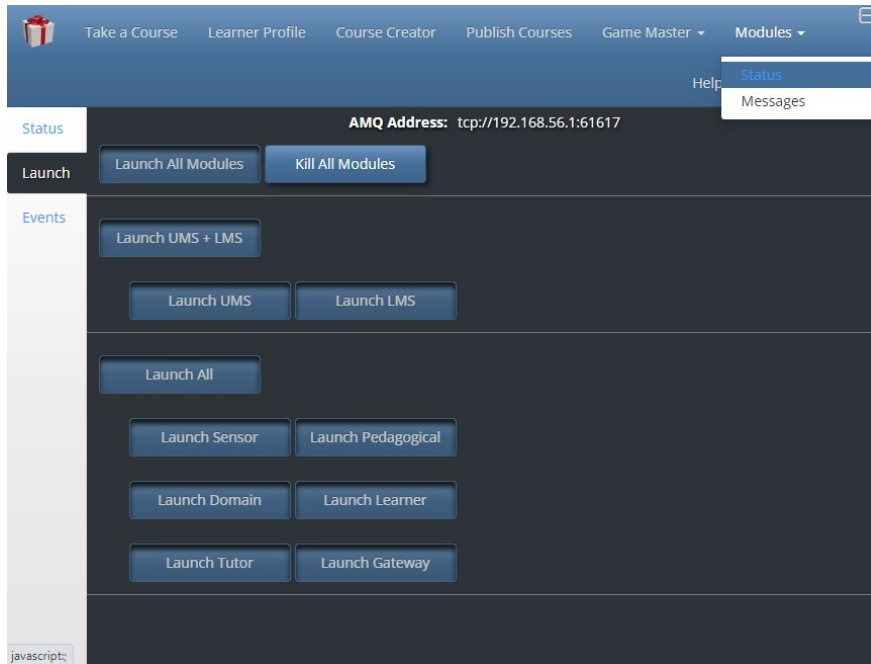


Figure 11. Module Status in the Web Monitor

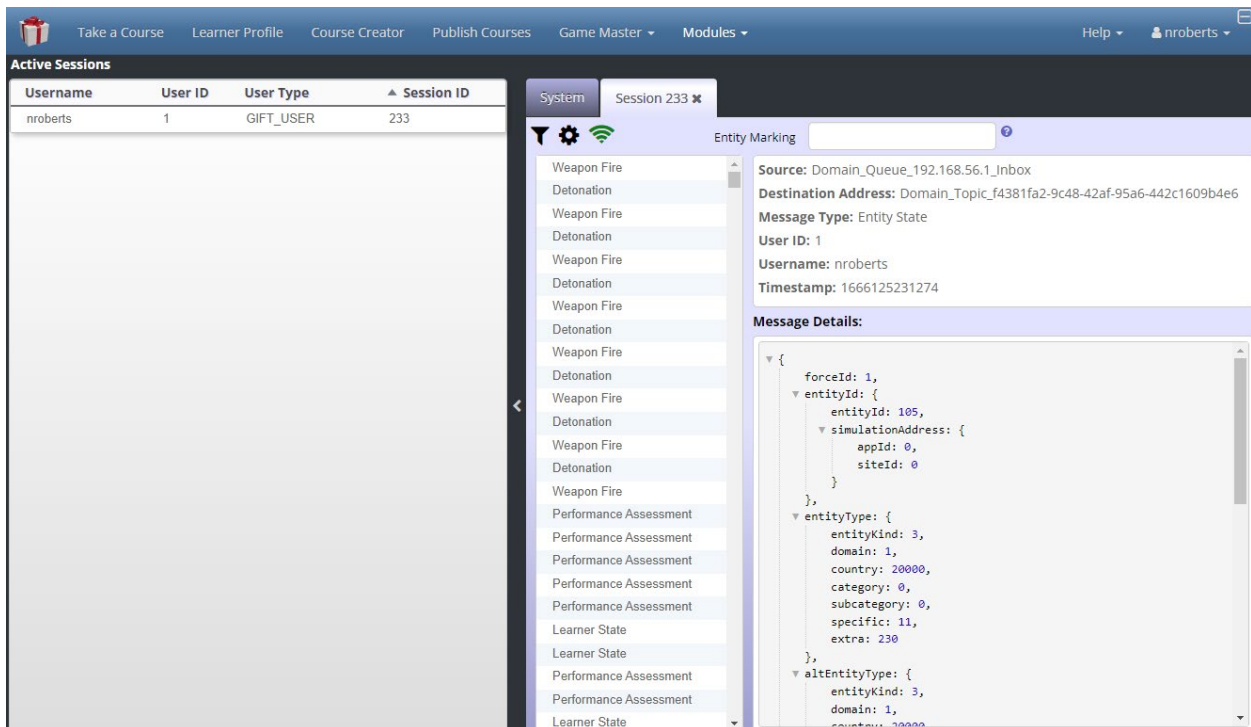


Figure 125. Domain Session Messages in the Web Monitor

## Java 11 Support

GIFT's support for Java 11 that began following the 2021-2 release has now been fully integrated into the 2022-1 release. With this change, GIFT no longer uses Java 8 and instead uses a distribution of OpenJDK 11 that comes packaged with this release. A notable impact of this change is that this Java distribution no longer supports 32-bit machines and is instead optimized for 64-bit platforms. Support for 32-bit machines has long been dwindling, to the point where Java no longer provides a 32-bit version that GIFT can use. Fortunately, 64-bit machines will benefit from this change, since GIFT now requires less configuration to use more memory and processing power if needed. Another impact of the switch to Java 11 is that Java has discontinued the use of Java Web Start (JWS), which has historically been used by GIFT Cloud to run GIFT's Gateway module during courses that use external training applications. To replace JWS, GIFT Cloud downloads a .zip file containing the Gateway module during these courses, which also removes the need for code-signing certificates that could sometimes trigger browser security warnings. Going forward, we plan on letting GIFT's Gateway module persist between courses to avoid needing to download the .zip file again upon starting another course.

The switch to Java 11 also required a number of changes to GIFT's software dependencies, which will impact developers more than researchers. The changes to GIFT's dependencies are outlined below:

### Updated Libraries:

- GWT → v2.7.0 to v2.9.0
- GWT Query → v1.4.2 to v1.5.0
- Google Guice → v3.0 to v5.0.1
- GWTBootstrap → v0.9.4 to v1.0.1
- GWTBootstrap Extras → v0.9.4 to v1.0.1
- Derby → v10.10.1.1 to v10.15.2.0
- Tincan With Dependencies → Tincan v1.1.2 jar is now standalone with dependencies added separately for easier maintainability
  - Lombok → v1.16.8 to v1.18.12
  - Dom4J → v1.6.0 to v2.1.0
- Javassist (JAVA programming ASSISTAnt) → v3.12.0 to v3.28.0

### Added Libraries:

- JAXB → v2.3.1
- JavaFX → v11.0.2
- Zip4J → v1.3.2
- Jackson Annotations → v2.12.5
- ActiveMQ → addition of 64 bit binary distribution for 64 bit JDKs

### Replaced Libraries:

- db-derby-10.10.1.1-bin
- gwt-2.7.0
- gwtbootstrap-0.9.4
- gwtbootstrap-extras-0.9.4
- tincan-1.1.2-jar-with-dependencies
- javassist-3.12.0.GA
- dom4j-1.6.1



Developers working in branches of GIFT that wish to merge changes back into trunk will need to update their branches to support Java 11. For the most part, this support will be gained automatically upon merging trunk into the branch in question, though branches that rely on additional libraries may need to find versions of those libraries that are compatible with Java 11.

### **Script File Changes and Linux Support**

Many GIFT script files had their logic converted into a portable format to allow GIFT processes to run on operating system platforms other than Windows, specifically Linux. This change has minimal impact on non-developers, and provides GIFT developers more flexibility in how and where to deploy GIFT. Historically, GIFT has relied on Windows-specific script files, such as .bat and .vbs files, to support the applications tied to its installer and core executables, but in an effort to support headless and server-based instances of GIFT, the reliance on these scripts has been toned down in favor of platform-agnostic configurations that can be handled by GIFT's JDK. This difference can be readily seen in the launchProcess.bat script in GIFT's scripts/tools/util folder, which contains settings that are modified for hosted distributions of GIFT. Most of the logic within this script has been moved to the launchProcess.xml file. These new platform-agnostic configurations use Apache Ant to handle their logic using GIFT's JDK (Java Development Kit). For specific execution logic that needs to be handled on an operating system (OS)-specific level, Ant is also used to determine the OS that GIFT is running on and make calls directly to the OS as needed.

In addition to these script file changes; a set of experimental developer scripts are included in GIFT's new scripts/dev-tools/Linux folder for developers that are interested in running GIFT in Linux. These scripts are in an early implementation stage and may require additional configuration to work on some distributions of Linux. The downloadable version of the release on [gifttutoring.org](http://gifttutoring.org) does not come packaged with Linux JDK, so developers wishing to use these scripts will need to use GIFT SVN (Subversion) to checkout the JDK from ThirdParty/Linux/x64. Once the JDK is in GIFT's external folder, GIFT can be built and installed on Linux by running the scripts/dev-tools/Linux/installGIFT script, which will also extract the launchGIFT script that can be used to run GIFT. These Linux scripts will be refined in future releases of GIFT.

### **Miscellaneous Improvements**

Several other changes were made to GIFT for this release. The list below highlights a few of these changes. Refer to the Release Notes 2022-1 wiki page on [gifttutoring.org](http://gifttutoring.org) for more information.

- A new Local Video course object was added to GIFT's Course Creator thanks to efforts from Ho and Donovan (2022). This course object functions similarly to the Image and objects Local Webpage in the sense that it displays content from a file in a course's course folder, though in this case, the file shown by the course object is a .mp4 file (see Figure 13). It should be noted that .mp4 files can use video codecs that are incompatible with certain web browsers, so authors attempting to use such files are encouraged to test their videos in browsers that the course's participants are likely to use.

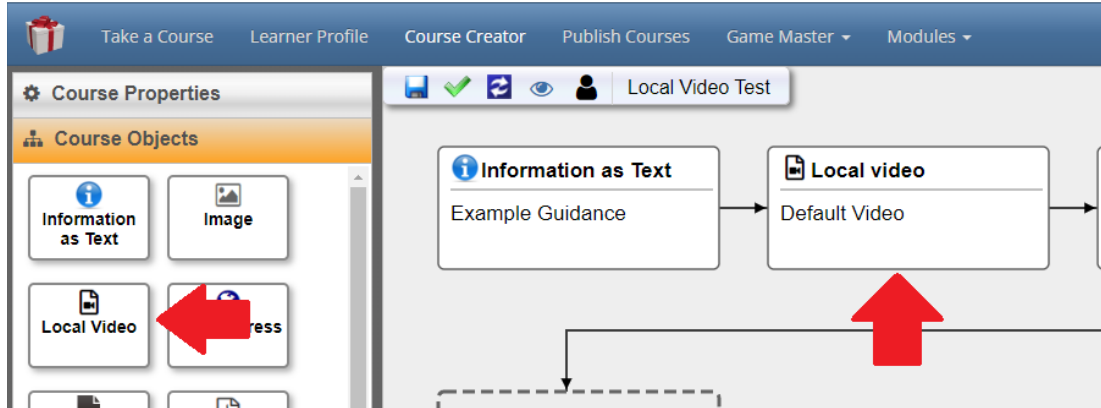


Figure 6. Local Video Course Object in Course Editor

- When editing task start or end triggers for a task in the course creator, authors can now use Places of Interest (PoI) from the same DKF to define location-based triggers that are activated based on an entity's proximity to a location in a simulated environment. Previously, if an author wanted to start a task when a player's-controlled entity reached a specific point, they would have to enter a completely new set of coordinates for the location to reach. Now, if the author has already defined this location in the PoI, then they can reuse the PoI where this was defined. Note that only Point objects from the PoI are eligible for this behavior, so Paths and Areas cannot be used to define location-based triggers.
- Generating a report from GIFT will now generate a .zip file containing both the report content that was generated as well as a file containing the settings that were used to generate the report. This change affects reports created using both the Event Report Tool (ERT) and the Publish Courses page within the GIFT Dashboard. The settings included in the export can be used to produce similar reports with the same settings using different domain sessions.
- Bugs that could occasionally prevent Game Master's active sessions list from populating have been fixed. One way to run into this problem was to visit the Active Sessions page in Game Master multiple times while a knowledge session was running. Issues like this have been fixed by cleaning up cached resources in the browser and on the server when users navigate to and from Game Master.

## REQUESTED FEATURES FROM GIFTSYM10

---

GIFT is community-driven, and we take pride in our user base, especially as it relates to functions and processes requested to support their research and content delivery needs. From last year's symposium, there were relatively few papers which actively requested or demanded features for development. This is good and shows a robust platform – the majority of papers presented describe an activity which is ongoing with GIFT, rather than addressing some weakness or shortfall.

## GIFT AND IEEE STANDARDS ON ADAPTIVE INSTRUCTIONAL SYSTEMS

---

The discussion continues on adaptive instructional systems through the IEEE Learning Technologies Standards Committee (LTSC). LTSC coordinates with other organizations that produce specifications and standards for learning technologies and we are excited to be a part of those endeavors. It is a big year with the xAPI becoming an official IEEE standard. The GIFT community invites the reader to join the

conversation on what data exchange standards for learning technologies might look like in the future. Interested readers are encouraged to go to the IEEE LTSC meetings to become involved.

## **CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

---

The GIFT program has seen significant advancement since its conception in 2011. Each year, the community continues to develop new features and use cases for adaptive instructional systems. With a near-term focus on utilizing GIFT to address data use and team tutoring challenges, we are excited to continue evolving the tools and methods to address critical capability gaps to drive future training requirements and system development. Stay tuned for continued improvements that address all facets of intelligent tutoring in today's education and training climate. Check back next year to see what kind of progress we make!

## **REFERENCES**

---

- Brawner K, Ososky S. The GIFT 2015 report card and the state of the project. In: Sottolare R, Sinatra A, editors. Proceedings of the 3rd Annual GIFT Users Symposium (GIFTSym3); 2015 Jun 17–18; Orlando, FL. Aberdeen Proving Ground (MD): Army Research Laboratory (US); c2015. ISBN 978-0- 9893923-8-9.
- Brawner, K., Sinatra, A. M., & Sottolare, R. (2017). Motivation and research in architectural intelligent tutoring. *International Journal of Simulation and Process Modelling*, 12(3-4), 300-312.
- Brawner, K., & Hoffman, M. (2018). *Architecture and Ontology in the Generalized Intelligent Framework for Tutoring: 2018 Update*. Paper presented at the Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6).
- Brawner, K., Hoffman, M., Nye, B., & Meyer, C. (2019). Architecture and Ontology in the Generalized Intelligent Framework for Tutoring: 2019 Update. Paper presented at the Proceedings of the 7th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym8).
- Goldberg, B., Brawner, K., & Hoffman, M. (2020). *The GIFT Architecture and Features Update: 2020 Edition*. Paper presented at the Proceedings of the 8th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym8).
- Goldberg, B., Owens, K., Hellman, K., Robson, R., Blake-Plock, S., Hoffman, M. & Gupton, K. (2021). Forging proficiency and readiness through an experiential learning for readiness strategy. In Proceedings of the 2021 Interservice/Industry Training Simulation and Education Conference (I/ITSEC). Orlando, FL.
- Hernandez M., Goldberg, B., Robson R., Owens, K., Blake-Plock, S., Welch, T., Ray, F. (2022) Enhancing the Total Learning Architecture for Experiential Learning. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Orlando, FL.
- Ho, R., & Donovan, B. (2022) *Applying intelligent tutoring systems to an Army Synthetic Training Environment*. Unpublished manuscript.
- Hoffman, M., Goldberg, B., & Brawner, K. (2021). The GIFT Architecture and Features Update: 2021 Edition. Paper presented at the Proceedings of the 9th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym9).
- Hoffman, M., & Goldberg, B. (2022) The GIFT Architecture and Features Update: 2022 Edition. Paper presented at the Proceedings of the 10th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym10).
- Ososky, S., & Brawner, K. (2016). The GIFT 2016 community report. Paper presented at the Proceedings of the 4th Annual GIFT Users Symposium.
- Sottolare, R., Brawner, K., Goldberg, B. & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Sottolare, 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, 1-19.
- Sottolare, 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.
- Sottolare, R., Holden, H., Goldberg, B., & Brawner, K. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottolare, R. (Eds.) *Fundamental Issues in Defence Simulation & Training*. Ashgate Publishing.

## ABOUT THE AUTHORS

---

**Nicholas Roberts** is a senior software engineer at Dignitas Technologies and the engineering lead for the GIFT project. Nick has been involved in the engineering of GIFT and supported collaboration and research with the intelligent tutoring system (ITS) community for nearly 10 years. Nicholas contributes to the GIFT community by maintaining the GIFT portal ([www.GIFTTutoring.org](http://www.GIFTTutoring.org)) and GIFT Cloud ([cloud.giftttutoring.org](http://cloud.giftttutoring.org)), supporting conferences such as the GIFT Symposium, and technical exchanges with Soldier Center and their contractors. He has also been heavily involved in integrating GIFT into TSS/TMT.

**Thomas Lenz** is a senior systems engineer and GIFT principal investigator at Dignitas Technologies. Thomas has been involved in various aspects of GIFT's development and testing for several years, recently shifting into a leadership role within Dignitas Technologies. His background includes over 10 years of instructional systems design for live and virtual training systems and systems engineering for Army and Navy use cases.

**Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.

# Supporting Future Learning Concepts: GIFT in the Year 2040

**Benjamin Goldberg and Randall Spain**

US Army Combat Capabilities Development Command (DEVCOM) - Soldier Center

## INTRODUCTION

---

Two years ago at GIFTSym9 we presented a use case that aligned (future) GIFT capabilities to the Army's Synthetic Training Environment (STE) modernization program (Goldberg & Owens, 2021). STE provided a mechanism to examine extensions of GIFT across the Plan, Prepare, Execute, Assess (PPEA) training management cycle. The resulting use case has informed design requirements in support of the STE Experiential Learning for Readiness (STEEL-R; Goldberg et al., 2021) project, which highlights GIFT's utility as an intelligent tutor data service in STE and how it builds alignment towards the Total Learning Architecture (Walcutt & Schatz, 2019). GIFT provides an ability to model and track STE's influence on competency development through an experiential learner modeling paradigm (Kolb et al., 2014). As stated in Goldberg and Owens (2021), these forward leaning use cases are vital for identifying adaptive instructional system (AIS) functions and requirements that are scalable and can meet organizational training and education needs.

Beyond STE modernization, the Army has also emphasized a broader enterprise level strategy to advance talent management and training practices. With a goal to objectively define a future training and education landscape, the Army is producing two concept papers, the Army Training Concept 2030 (ATC2030) and the Army Learning Concept (ALC) for 2030 - 2040. These long-term focused concept papers were introduced during a panel discussion at the 2022 US Army University Learning Science Symposium, hosted at Fort Leavenworth (Army University, 2022). The panel was comprised of concept leads and contributing authors, and the discussion covered high-level themes centered on a future Army training and education ecosystem. These papers are serving as an update and extension to the previously released Army Learning Concept for Training and Education 2020 – 2040 (TRADOC, 2017), with careful separation of learning and training required capabilities, and aligning to tenets introduced by the Advanced Distributed Learning (ADL) Initiative's book "Modernizing Learning: Building the Future Learning Ecosystem" (Walcutt & Schatz, 2019). They are being strategically developed across an interdisciplinary team of researchers, practitioners, and operators to establish directive themes on modernizing tools, methods, and technologies to optimize and interconnect development and skill progression across the Army's three learning domains (i.e., institutional, operational, and self-development).

These concepts provide an exciting perspective on how Army personnel will develop and sustain skills in a future climate, and lean against maturing tenets of organizational psychology, cognitive psychology, learning engineering, data science, artificial intelligence (AI), and learning technology standards. Of particular interest to the adaptive instructional system and learning engineering communities is the ALC 2030 - 2040's emphasis on learning technology and data-driven processes over an interconnected architecture, with direct alignment to objectives and required capabilities outlined in the ATC2030 (Army University, 2022). The ATC2030 establishes capabilities that will support high-fidelity synthetic and live training experiences that mimic operational environments and facilitate direct skill transfer, while the ALC will provide required capabilities to enable optimal utilization and assessment within those training opportunities. In addition, the ALC must extend across a broader spectrum of training and education programs and resources, establishing the goal of a broader Army talent ecosystem.

The purpose of this paper is to briefly review foundational themes of the ALC 2030 – 2040 and discuss GIFT’s potential role within the broader Learning Concept vision, with attention to design tenets that will guide research and development needs and priorities moving forward. We include considerations on implementable requirements to support integration and sustainment of learning technology within an ecosystem that conceptually will be continually evolving. In addition, we discuss capability considerations that will guide future research questions and investments that align to required capabilities to bring the ALC’s vision of a future learning ecosystem to reality.

## ARMY LEARNING CONCEPT 2030 - 2040

---

The ALC provides a conceptual framework of how future operational environments will impact the Army’s training and educational system and how the Army must respond to meet these challenges. The previous learning concept (TRADOC, 2017) being extended upon in the ALC under development establishes a central idea defined around the principles of a learning organization wherein employees continually create, acquire, and transfer new knowledge, to help the organization, adapt to changing demands and uncertain environments (Garvin et al., 2008). This vision calls for a model of continual improvement to systems, processes, and policies guiding talent management and learning requirements within an organization. To achieve this, an ALC framework was introduced at the US Army Learning Science Symposium that highlights five themes that will integrate across the three Army learning domains, including: Culture of Continuous Organizational Improvement; Workforce Development; Learning Strategies Focused on Learner Centricity; Integrate Enabling Technologies; and Data-Driven Assessment and Infrastructure (see Figure 1). These tenets lay the foundation for defining future-oriented capabilities that align to learning engineering principles that utilize foundations in learning science, leveraging advancements in technology and methodology to support talent competency needs, and establishing organizational strategies for iterative refinement and modernization.

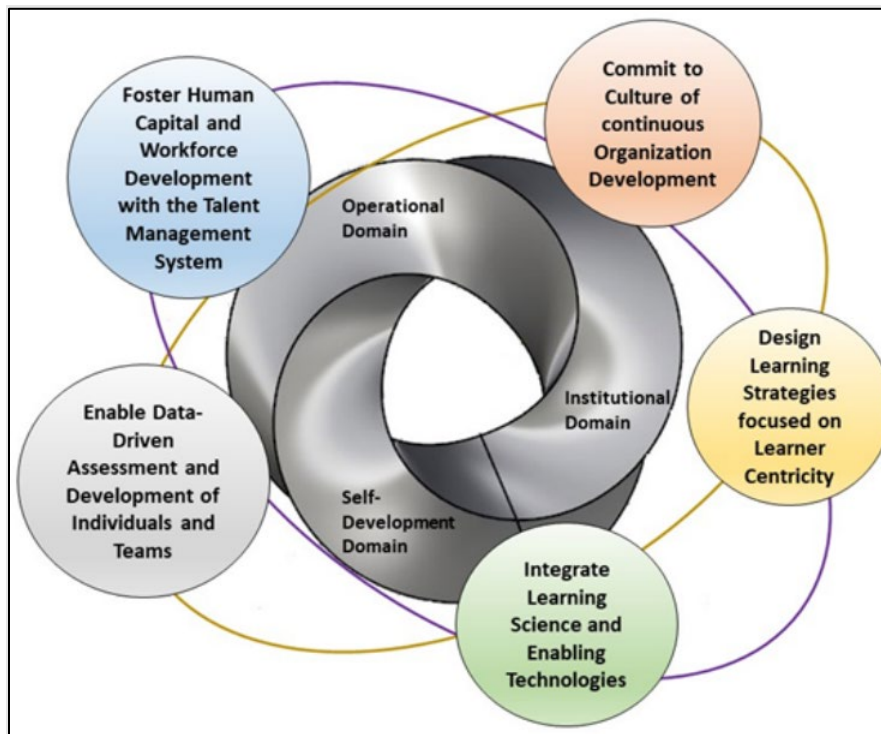


Figure 1. The Army Learning Concept 2030 – 2040 Framework (Army University, 2022)

In the following sub-sections, we briefly introduce and summarize the five themes covered during the ALC symposium panel. We then introduce implications for GIFT and STEEL-R implementation and future research and development needs. The discussion is not intended to be all encompassing. Rather, it is meant to educate the GIFT user community on these new concept papers, and to then begin the discussion on how a platform like GIFT/STEEL-R and a research program dedicated to advancing the application and utilization of AISs should react to these organizationally established goals and forward focused visions.

### **Theme 1: Culture of Continuous Organizational Improvement and Theme 2: Workforce Development**

For the sake of brevity, we combined the first two themes in this overview, as they are complimentary in the broader message discussed during the symposium. The Army aims to implement principles of a learning organization by establishing a learner-centric model that blends technology-based learning, self-paced learning, and face-to-face instruction across formal and informal instructional experiences. Adopting principles of a learning organization philosophy was a key element of the Learning Concept for Training and Education 2020-2040, and will continue to be a cornerstone in the new ALC. To transform into a learning organization, the Army must foster a culture of continuous improvement that empowers and enables its workforce to continuously develop new knowledge and skill, and readily apply these new skills to improve organizational outcomes. Successful learning organizations empower individuals and foster their sense of responsibility for learning by providing enabling systems and encouraging open discussions within chains of command to support talent development. The learning organization also encourages learners to seek out instructional resources and validate learning experiences developed by other organizations.

With these goals in mind, the second theme highlights workforce development and talent management practices needed to support the Army as a learning organization. As noted in the ALC, the Army must invest in its training and education workforce, which includes leaders, mentors, teachers, instructors, facilitators, training managers, and developers, to ensure these leaders have the required skills to be effective educators. A central goal will be building a workforce that is skilled with learning engineering principles and knowledgeable of effective instructional practices and theories (Goodell & Kolodner, 2022). Properly educating and hiring teachers, mentors, and instructors with digital skills will be essential for meeting the demands of the future. These teachers and instructors will serve as experts and authoritative sources for learners. Talent management tools will help the Army produce fully functional multidisciplinary learning development and employment teams comprised of experts in subject content, learning theory, instructional systems design (ISD), digital tutoring and tailored assessment, game and simulation design, and media development.

In summary, the Army's ALC framework calls for a culture of continuous workforce improvement to meet the challenges of multi-domain operations (MDO). Talent management practices will need to provide Soldiers and civilian personnel with opportunities to continuously develop their knowledge, skills, and abilities. At the center of these goals are education technologies and human capital investments that leverage the principles of learning engineering to diagnose, prescribe, and support tailored and meaningful learning experiences. These investments will allow the Army's training and education workforce to evolve and support formal, informal, episodic, and lifelong learning

### **Theme 3: Learning Strategies Focused on Learner Centricity**

In this subsection, we provide summary statements on how proven empirical learning science must be a foundational part of implementing sound training and education strategies. The ALC highlights the Army's commitment to implementing learner-centric principles and practices across its learning and development



domains. It highlights the importance of developing competencies required for facing the demands of complex and uncertain operational environments including problem-solving, creativity, and inquiry-based learning. It also addresses the use of advanced technologies, such as adaptive virtual learning environments to train and prepare Soldiers to operate in uncertain, contested, degraded, and limited information environments. The ALC discusses the importance of tailoring learning experiences to the needs of learners and using principles of instructional and synthetic training environments to ensure Soldiers learn how to improve decision making and performance under stress and periods of high cognitive load. Applying learning principles grounded in cognitive science is critical to development and sustainment of adaptable mental models that are required to perform in volatile settings. This involves careful design of training interventions that account for benefits of interleaved practice (Taylor & Rohrer, 2010) and spacing effects to reduce rates of skill decay (Kim et al., 2013). Additionally, explicit teamwork development is emphasized in addition to traditional work activities. Further, learning product developers and designers tailor learning strategies and tactics for desired learning outcomes.

Furthermore, the ALC highlights the need for instructional designers to familiarize themselves with the theoretical underpinnings of all major learning theories to develop instructionally effective learning and training experiences that will prepare Soldiers to successfully navigate the demands posed by multi-domain environments. It is critical to understand the connection between research-based theoretical principles and instructional techniques and strategies derived from those principles, as failure to do so could lead to erratic and inadequate instructional design when coupled with advances in areas such as neuroscience, AI, and other technology support systems. Overall, the Army is invested in improving learning outcomes and is willing to invest in the people and enabling organizational processes to achieve this goal.

#### **Theme 4: Integrate Enabling Technologies**

Technology is constantly changing the way we interact with our surrounding world, and we must project a future so that we can begin preparing the infrastructure now. Specifically forecasting how technology will impact how we develop knowledge, skill, and job-related competencies in the future is a fun exercise that requires a blend of creativity anchored by realism. The ALC establishes an integrative training and education approach that leverages advances in next-generation training technologies to support the development of behavioral, cognitive, and affective skills. The Army assesses learners' digital competencies with tailored competency frameworks, incorporating creative and critical thinking, leadership competencies, and evolving technological competencies associated with a future operational environment that promotes human-machine teaming. The Army applies a learning engineering and instructional design framework to its technology to provide opportunities to apply technical and tactical competencies in different contexts.

Furthermore, the ALC promotes use of immersive, experiential events in a controlled data-rich synthetic setting to facilitate individual and collective training tailored to optimize skill acquisition and competency development. This involves taking advantage of emerging technologies across the extended reality (XR) capability spectrum embedded with intelligent tutoring systems and AI to enhance learning, objective assessment and to provide the right type of learning opportunity at the right time within a designated learning curve. The Army's learning infrastructure uses interoperable standards to support the learning environment, allowing access to synthetic learning resources and virtual forums for on-demand learning across domains. The infrastructure is adaptive, flexible, and reliable to support learners circumventing obstacles in areas such as connectivity, bandwidth, on-site technicians, and network management.



## **Theme 5: Data-Driven Assessment and Infrastructure**

Behavioral assessments are essential for tailoring training and education to learners' needs and experience, shaping learner feedback, and verifying and validating that learning has occurred as intended. Assessments incorporate a composite of descriptive, quantitative, and qualitative sources, and must account for real-time and longitudinal learning processes (Biswas et al., 2019). This function emphasizes the importance of bridging the gap between data collection, analysis, and intervention with a goal of driving continuous organizational improvement through machine learning and data-informed processes.

To support a technology-centric concept grounded in learning science, the ALC panel recognized a critical need for a data-driven infrastructure to drive larger scale implementation and integration of learning resources across all the Army learning domains. This requires development of a holistic learning ecosystem that emphasizes competency-based learning strategies and a network-centric architecture to manage data across disparate learning events and interdependent technologies (Walcutt & Schatz, 2019). The learning ecosystem must be driven by common, standardized, shareable, and secure data standards that guide interoperability specifications and policies for data governance, privacy, and security. The panelists also highlight direct use of industry data standards, machine-readable competency frameworks, and learning science principles to support talent management across all learning domains.

The cornerstone of the learning ecosystem data strategy is the development of proponent managed competency frameworks. These frameworks define the required personnel capabilities for the force, set standards for assessment, and create a roadmap for aligning learning outcomes to career tracking. Competency frameworks will cover foundational KSBs (Knowledge, Skills, Behaviors) that are concentration specific, while accounting for common knowledge and skill profiles linked to leadership and team development. The data infrastructure will also support analysis for identification of emerging competencies for both individuals and teams. This requires the solution to support direct proponent authoring, management, refinement, and removal of competency frameworks to meet evolving operational requirements.

## **IMPLICATIONS FOR GIFT AND STEEL-R**

---

Without too much hyperbole, the ALC provides further justification for continued investment in the research, development and maturation of GIFT and STEEL-R capabilities aligned to Army training and education requirements. The ALC also provides guidance on future development priorities, with explicit definition of required capabilities and research needs. While our discussion just covers the tip of the iceberg in terms of our research program's alignment to the ALC vision, it identifies high-level categories that must be carefully accounted for in near-term research and development decisions. This accounts for GIFT's role as an AIS within the broader Army future learning concept, and how GIFT interconnects with learning ecosystems through established industry standards. These capabilities provide an excellent starting point to leverage decade plus government investments in advanced training technologies that can be directly steered towards the ALC capability goals. In the following paragraphs, we review three high-level design considerations that must be implemented to ensure we keep pace with a rapidly maturing industry and define an acquisition strategy to bring the ALC to reality.

### **Opens System Architecture**

With a future learning concept centered on competency-driven methods that interconnect learning domains and training experiences, providing a general and extensible adaptive instructional framework is critical. In this manner, GIFT is not responsible for directly maintaining concurrency of best practice intelligent

tutoring methods within its core architecture, but rather it must enforce interoperability across constituent capabilities and platforms that support varying domains and developmental phases within a domain. In other words, GIFT can provide a flexible infrastructure to integrate and interoperate core instructional technology components that are otherwise considered stand-alone solutions. This creates a competitive open market to continually evaluate weaknesses within the supported technologies and to provide innovative tools and solutions that can plug-in and meet a direct need within the learning infrastructure (e.g., new assessment techniques for modeling communication skills, leadership; new coaching techniques that focus on building resiliency and self-regulated learning strategies; new sensor classifiers that accurately monitor engagement and frustration during the learning process, etc.). Whatever the application or service, there needs to be an easy path of integration into an already existing ecosystem supporting training and education at the organizational level.

This broad requirement must be enabled through an open system architecture development strategy that continues to build upon the foundational principles of GIFT, while highlighting its role and streamlined connection to a designated training ecosystem. The requirement would involve using GIFT's extensible data models and schemas to manage the collection, retention and analysis of data produced during live, virtual, and XR-mediated training and education experiences. Vendors and practitioners control the training interaction, while leveraging GIFT to structure its assessment strategies and learning experience workflows. With the assessment schema linked to an xAPI (eXperience Application Programming Interface) profile that can generate xAPI statements, any learning experience leveraging this architecture should automatically connect to a cloud-enabled ecosystem and competency management system that will support visualization and diagnosis of training needs.

### ***Interoperable Assessment Structures and Embracing External Services***

To support STEEL-R's data strategy objectives, we are developing a modeling schema called the eXperience Training Support Package (XTSP; Hernandez et al., 2022). The XTSP is designed around an experiential learning paradigm and establishes standard data models for representing measurable experiences within a training and education event. The goal of the XTSP is to control how training and educational experiences are measured and assessed from a performance and proficiency standpoint, while also informing a data management strategy that links discrete training experiences together to support competency assessment and longitudinal modeling. As an example, a series of synthetic training resources can be used to train a single set of core competencies (e.g., game environment to learn cognitive and behavioral rules, a mixed reality environment to get exposure to physical requirements, and a live environment to replicate real-world execution). In this instance, the same XTSP assessment model can apply across all training experiences with different data sources informing measurement.

An important aspect is the XTSP can be used to directly inform GIFT's assessment model established within the Domain Knowledge File (DKF). This interoperability layer allows GIFT to import an XTSP, create a validated DKF from it, and then manage assessment in real-time through GIFT's connected gateway module. In addition, GIFT then auto-generates xAPI statements aligned to the assessment structure in the XTSP, providing a streamlined connection point between the learning event and the repository tracking all learning events from an ecosystem context.

### **Designing for Evolutionary Advancements in XR and AI**

Technology is maturing at a rapid pace. With recent advancements in generative AI tools and techniques, AI applications are in the mainstream media more than ever. As a result, industry sectors are required to investigate how these new tools and methods can directly impact business and educational practices. This is certainly true in learning engineering and development of next-generation training and education systems.

Content is becoming more immersive and engaging, with XR tools and methods becoming more common practice than ever. And because of new AI tools and methods, research and development communities are exploring new ways to utilize these interactive learning tools to drive truly impactful educational experiences. While adhering to open system architecture principles, understanding how AI tools can be applied to assist in the front-end design and development requirements of an effective learning event is critical.

### ***What does ChatGPT have to Say about This?***

We thought it would be an interesting exercise to see what ChatGPT thinks about the role of generative AI in the landscape of intelligent tutoring and adaptive training. We entered the below prompt and included the unedited answer directly below. We then share our own perspective on the answer provided.

***Prompt:*** You are a scientist funded by the Army to do research on intelligent tutoring systems. How will advancements in large language models and generative AI impact intelligent tutoring and adaptive training practices in the Army?

***Answer:*** Advancements in large language models and generative AI have the potential to significantly impact intelligent tutoring and adaptive training practices in the Army. These advancements can improve the ability of intelligent tutoring systems to generate personalized and adaptive training content, which can be tailored to meet the specific needs and learning styles of individual soldiers.

One major impact of large language models and generative AI on intelligent tutoring systems is the ability to generate natural language responses to soldier inquiries. This can help soldiers get immediate feedback on their performance and better understand the material being covered. Generative AI can also be used to create interactive simulations that provide soldiers with a realistic environment to practice critical skills.

Another important impact of generative AI on intelligent tutoring systems is the ability to generate personalized training content. By analyzing data on individual soldiers' performance and learning styles, AI can generate content that is tailored to their specific needs. This content can be dynamically adjusted in real-time to ensure that soldiers are continually challenged at a level appropriate for their current level of skill and knowledge.

Generative AI can also be used to provide soldiers with more realistic and engaging training experiences. For example, AI-generated scenarios can simulate complex real-world situations, allowing soldiers to practice their decision-making skills in a safe and controlled environment. Additionally, generative AI can be used to create virtual coaches that can provide feedback and guidance to soldiers during training exercises, offering personalized instruction and support.

Overall, the impact of advancements in large language models and generative AI on intelligent tutoring and adaptive training practices in the Army is significant. These advancements can provide soldiers with more effective and personalized training experiences, helping them develop critical skills and better prepare for the challenges they may face in the field (OpenAI, 2023).

***Response.*** There are two interesting insights to take-away from this answer. First, there is an emphasis for dialogue-based instruction and coaching. Examining approaches to interface learners and trainees with specially trained language models through conversational agent techniques can change the way we interface

with instructional materials and After-Action Reviews (AARs). These agents can be trained as domain experts with underlying pedagogical properties based on student individual differences and competency levels. Examining how GIFT can directly interface with agents linked to language model processes is an area of research worth investigating. This can support new interactive learning techniques with traditional didactic content. In addition, we can examine new innovative ways to implement technology mediated AARs and learning reflection periods. Through GIFT's interoperable assessment techniques, learner state information can be input into a conversational agent following an assessment event. This can inform strengths and weaknesses at the individual and team level, which would drive a more focused and personalized remediation.

The second interesting take-away is the role of generative AI to support development of realistic scenarios and more personalized content. The area of content creation and curation has always been a challenge for intelligent tutoring researchers and developers. The more adaptive and personalized you want your system or experimental protocol, the more content and variations in content required. Examining how AI techniques can take learning objective structures and generate engaging content and problem sets will be required to meet the vision of the ALC. Proponents can define and refine the exact concepts and relationships they want emphasized in their programs of instruction, and technology can build and iteratively improve content and practice scenarios. In the same vein, these proponents defined structures can be used to build operationalized competency frameworks that can be informed from data collected across the learning ecosystem.

### **Standards, Standards, Standards: Establishing Best Practices**

Across all learning concept themes and required capabilities exists a need for standards. This message has been stated several times throughout this paper. Within the DoD, this requirement is captured across several policy documents and sponsored publications, with the DoD Data Strategy serving as an initial guiding perspective (DoD, 2020). In our context, the goal is to establish requirements grounded in implementable and enforceable standards that align to training system development and acquisition strategies. Standards are also intended to drive scalability of solutions developed and matured within a laboratory setting, so that the tools, methods, and best practice workflows can be reproduced and reimplemented across a broad audience of learning engineers and practitioners.

With that said, there has been considerable investment in defining and updating standards across the learning technology spectrum that will directly influence a future ALC. As highlighted by the symposium panel, an influential resource guiding the expansion of the previous concept is ADL's Modernizing Learning book (Walcutt & Schatz, 2019). This publication was followed up by the release of the Learning Engineering Toolkit (Goodell & Kolodner, 2022), which focuses on operationalizing these modernization concepts into sound engineering recommendations. These resources establish guidelines for implementing an instrumented ecosystem of interconnected learning resources, with careful attention to a standardized data fabric that drives integration and interoperability requirements across disparate resources spanning a diverse community of providers, learners, and government/stakeholder sponsors.

This data fabric is informed through community driven requirements and aligns to investments by IEEE's Learning Technology Standards Committee (LTSCa, n.d.). A recent accomplishment within this community was the successful balloting of the eXperience Application Programming Interface (xAPI) data standard (LTSCb, n.d.). This is a big deal, as it creates an official standard that guides structured implementation of performance and experience reporting through a controlled data syntax. This ensures systems can output data that can be combined to drive more robust learning curve analytics within an ecosystem and drives the data fabric required for a robust and extensible learning ecosystem. In addition to xAPI, several working groups are in motion looking at data models and metadata requirements around

xAPI, Adaptive Instructional System definition and components for reuse and development, and universal learning records for tracking training and education variables in a precise manner. Visit the IEEE LTSC website to maintain awareness of these activities and to get involved.

Beyond the specification of a standard to guide development across a community of interest, the ALC requires a set of best practices for widescale implementation across varying audiences and learning strategies. Just because a standard exists does not mean that it applies to all use cases and that practitioners know how to best implement it. The learning strategies implemented across a future ALC are grounded in learning science theory and account for different interaction and assessment requirements. To meet this flexibility, best practices and technologies to implement those technologies will be required. For instance, the STEEL-R program examined the use of data standards and adaptive instructional components informed by experiential learning theory, which required explicit extensions to xAPI and the workflows required to configure a system to output meaningful data used to model experiential learning curves (Hernandez et al., 2022). We have learned several lessons over the course of the project including the importance of using xAPI as a mechanism for tracking and recording training experiences, the need to model tasks and competencies at different levels of granularity to support competency modeling and the need to leverage simulation data standards to help automatically craft synthetic training scenarios. From a best practice standpoint, the following considerations must be addressed when examining a new learning strategy: (1) building learning resources that support the science behind knowledge and skill development, (2) instrumenting those resources to collect data, (3) applying assessment structures to apply data-driven techniques for objective performance measurement, (4) outputting assessments with proper granularity and context/metadata, and (5) longitudinal modeling techniques that combine assessments overtime to infer proficiency and predict performance outcomes.

## CONCLUSIONS

---

In this paper, we discussed several overarching themes identified during the ALC panel discussion and their implications for future development of GIFT and STEEL-R. Advancing training and education strategies across the Army's operational, institutional and self-development will be vital for supporting a culture of continuous learning and meeting the demands of future operational environments. The tenets discussed in the ALC lay the foundation for defining future-oriented capabilities that align to learning engineering principles, learning sciences, and the use of next-generation technologies and educational best practices to ensure the Army effectively manages and develops Soldier and Army Civilian competencies. Moving forward, GIFT and similar training management tools will play an increasingly important role in bringing these concepts to reality.

## REFERENCES

---

- Biswas, G., Rajendran, R., Mohammed, N., Goldberg, B. S., Sottolare, R. A., Brawner, K., & Hoffman, M. (2019). Multilevel learner modeling in training environments for complex decision making. *IEEE Transactions on Learning Technologies*, 13(1), 172-185.
- Department of Defense. (2020). DoD Data Strategy. Arlington, VA. Accessed from <https://media.defense.gov/2020/Oct/08/2002514180/-1/-1/0/DOD-DATA-STRATEGY.PDF>.
- Garvin, D. A., Edmondson, A. C., & Gino, F. (2008). Is yours a learning organization?. *Harvard business review*, 86(3), 109.
- Goldberg, B., Owens, K., Gupton, K., Hellman, K., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging competency and proficiency through the synthetic training environment with an experiential learning for

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- readiness strategy. In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.
- Goldberg, B., & Owens, K. (2021). GIFT in a Blended Learning and Competency Development Continuum. In Proceedings of the Ninth Annual GIFT Users Symposium (GIFTSym9) (Vol. 89). US Army DEVCOM–Soldier Center.
- Goodell, J., & Kolodner, J. (Eds.). (2022). Learning Engineering Toolkit: Evidence-Based Practices from the Learning Sciences, Instructional Design, and Beyond. Taylor & Francis.
- Hernandez, M., Goldberg, B., Robson, R., Owens, K., Blake-Plock, S., Welch, T., & Ray, F. (2022). Enhancing the Total Learning Architecture for Experiential Learning. In Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).
- 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.
- Kolb, D. A., Boyatzis, R. E., & Mainemelis, C. (2014). Experiential learning theory: Previous research and new directions. In *Perspectives on thinking, learning, and cognitive styles* (pp. 227-248). Routledge.
- LTSCa (n.d.). Learning Technology Standards Committee. <https://sagroups.ieee.org/ltsc/>
- LTSCb (n.d.). P92741.1 xAPI Work Group. <https://sagroups.ieee.org/9274-1-1/>
- OpenAI. (2023). ChatGPT (April 29 version) [Large language model]. <https://chat.openai.com/chat>
- Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, 24(6), 837-848.
- U.S. Army TRADOC. (2017). Army Learning Concept for Training and Education for 2020–2040 (TRADOC Pamphlet 525-8-2). Ft. Eustis.
- U.S. Army University. (2022). Army Learning Concept Roundtable. Discussion Panel Presented at the 2022 Army University Learning Science Symposium. Fort Leavenworth, KS. See video at: <https://www.youtube.com/watch?v=I68ffgZ112U>. See Info at: <https://armyuniversity.edu/Organizations/LearningSymposium/Home>.
- Walcutt, J. J., & Schatz, S. (2019). Modernizing Learning: Building the Future Learning Ecosystem. Advanced Distributed Learning Initiative.

## ABOUT THE AUTHORS

---

*Dr. Benjamin Goldberg is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.*

*Dr. Randall Spain is a Research Scientist at the U.S. Army Combat Capabilities Development Command (DEVCOM) Soldier Center. His research focuses on designing, developing, and evaluating adaptive training technologies with a particular emphasis on investigating how data-driven coaching and feedback models can be used to support team training in synthetic training environments and exploring how natural language processing-based methods can support team communication analytics.*



**THEME II:  
MEASUREMENT AND  
ASSESSMENT**





# Investigating the Effect of Realistic Agents on Team Learning in Adaptive Simulation-based Training Environments using GIFT

Robert Sottolare<sup>1</sup>, Christopher McGroarty<sup>2</sup>, Christopher Ballinger<sup>1</sup>, and Timothy Aris<sup>2</sup>

<sup>1</sup>Soar Technology, Inc.; <sup>2</sup>U.S. Army Combat Capabilities Development Command – Soldier Center (DEVCOM SC) Simulation and Training Technology Center (STTC)

## INTRODUCTION

---

This paper examines how the process of designing, creating, training, and evaluating realistic intelligent agents can influence team engagement and thereby team learning in adaptive simulation-based training environments. Intelligent agents are entities that are able to perceive their environment, process information (e.g., make judgments), and take actions autonomously to achieve assigned goals (Russell & Norvig, 1995). Intelligent agents have mechanisms to improve their performance by acquiring and using knowledge about their environment (context) along with knowledge about the successful and unsuccessful outcomes resulting from actions taken.

Under the Learning and Readiness Intelligent Agent Testbed (LARIAT) project, sponsored by the US Army DEVCOM-STTC, we investigated the factors influencing realism along with the degree and modality of realism required to stimulate and maintain trainee engagement. The investigation yielded guidelines for designing intelligent agents to optimize their realism and thereby engagement in adaptive training environments intelligently guided by the Generalized Intelligent Framework for Tutoring (GIFT). Learner 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” (Schmieder, 1973, <https://www.edglossary.org/student-engagement/>).

The positive relationship between engagement and learning is well documented in the literature (Axelson & Flick, 2010; Rowe et al., 2010), as is the relationship between realism and engagement (Checa & Bustillo, 2020; Tashiro & Dunlap 2007). Engaging trainees in active learning increases their attention, motivates them, and can lead to higher-level critical thinking, the process of analyzing information, evaluating arguments, and making decisions based on evidence and logical reasoning. Active learning strategies such as case studies, group or collaborative learning, brainstorming, and peer instruction are designed to mentally stimulate and engage trainees in problem solving.

Problem-solving involves identifying a problem, generating possible solutions, evaluating the solutions, and implementing the best solution. Active learning can provide trainees with opportunities to apply their knowledge and skills to real-world problems, which can help them to develop needed problem-solving skills that will more easily transfer to operational environments. Simulation-based training methodologies present opportunities to problem solve use scenarios, artificial representations of real-world events to stimulate trainees to apply knowledge and skills to solving relevant problems and achieve assigned goals through experiential learning.

Significant elements of simulation-based training include the representation of entities in the form of intelligent agents. Designing an effective simulation scenario requires careful planning in order to stimulate trainees and focus their problem solving skills appropriately. Part of this design process is an appropriate degree of scenario realism. Realism in simulation refers to the degree to which a simulation accurately represents real-world phenomena or events.

Research has also shown that there is a positive relationship between realism in simulation and trainee engagement. When simulations are designed with a sufficient level of realism, trainees are more likely to be engaged in the learning experience (Issenberg et al., 2005). Realistic simulations can provide trainees with a sense of immersion and a feeling of being in a real-world environment, which can increase their interest and motivation to learn. Realism is also important from a doctrinal perspective. Trainees engaging in simulation scenarios that are sufficiently credible in that they replicate the need for actual procedures and reinforce needed operational behaviors are more effective in transferring learning from training to the real-world.

In this light, we put forward that the behaviors of intelligent agents should be designed and trained to use real-world processes and replicate real-world behaviors to sufficiently engage trainees, enhance their learning, and increase their transfer of training.

## PROBLEM STATEMENT

---

Throughout their careers, Soldiers participate in live training exercises where scenarios are crafted to engage them and apply their knowledge and skill to solve real-world problems and complete representative operational tasks. These simulated events provide opportunities for deliberate practice with the goal of improving their skills, coordination, team cohesion, and communication. The Army's primary live training experience is driven by the Combat Training Center - Instrumentation System (CTC-IS), an information technology-based communications, analysis, and feedback system located at the National Training Center (NTC) in Fort Irwin, CA and other Maneuver Combat Training Centers (MCTCs).

CTC-IS provides a realistic operational environment for training Brigade Combat Teams and echelons below in preparation for deployment to conduct decisive actions. It is comprised of voice, video, and data instrumentation subsystem networks that include software, hardware, workstations, base-station equipment, communications infrastructure, voice radios, data devices, and interfaces. The Instrumentation System (IS) also provides the Observer Coach/Trainers (OC/Ts) critical situational awareness for training safety, analysis, and feedback necessary to facilitate the conduct of After Action Reviews (AARs).

According to technical personnel at NTC, only a very small percentage of this data is currently used to support AARs and none of the data has been used to build realistic autonomous agents for use in virtual simulations (Hoffman, 2022). Given the very large multi-modal datasets generated when training brigades engage in live training exercises at the NTC, has no one thought to use this data to develop realistic autonomous agents or has there been other impediments? The data is infused with perceptions (information about the environment) and documents the actions based on those perceptions. This data could form the basis for modeling a realistic agent.

One reason for the lack of progress in developing autonomous agents is that while the data acquisition processes at the NTC are largely automated, the data analysis processes are largely manual and limited to statistical analyses conducted on simple spreadsheets. There are also several challenges associated with the transfer of learning from live simulations to the production of realistic autonomous agents in adaptive virtual team training environments. Since no agents have been generated from these massive datasets, there is a lack of understanding of the precise effect of more realistic perceptions and actions on trainee engagement. This lack of understanding can only be overcome through experimentation.

In seeking to understand the barriers to using massive datasets to create realistic autonomous agents that positively influence trainee engagement, we pose the following research questions:

- What analysis approach will be most efficient in identifying symbolic anchors as features in the datasets that will be transferred to autonomous agents?<sup>1</sup>
- What approach will result in the best possible transfer of live unit attributes to realistic perception, planning, and action-taking in autonomous agents in virtual environments?
- If realism can be optimized, what is the expected impact on team engagement and learning during adaptive team training using GIFT?

## ANALYSIS OF MASSIVE DATASETS

Reflecting on our first research question, what approach can we use to efficiently analyze massive NTC datasets? Our key objective is to establish an efficient process that will enable analysts to extract features in the form of symbolic anchors from the datasets. A generic data pipeline for processing large datasets is shown in Figure 1, but may need to be adapted to support a more efficient analysis of massive, multi-modal datasets. Data pipelines are composed of a set of interconnected data processing elements that include data management (data cleaning, data enrichment, and feature extraction), knowledge representation, and model building and testing.

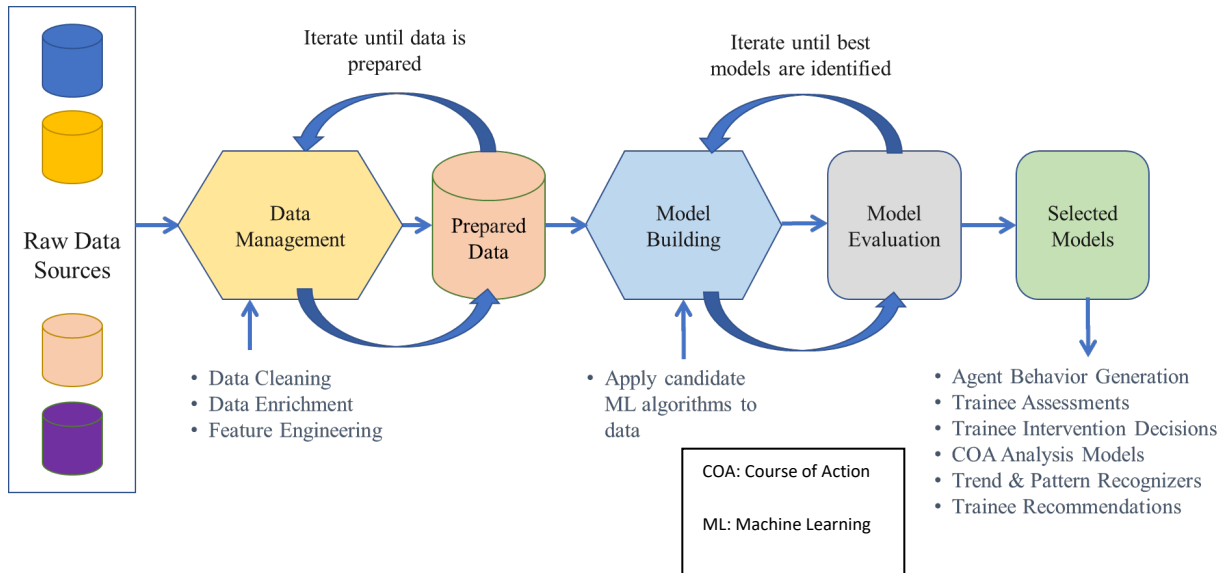


Figure 1. Generic Data Pipeline for Model Building & Testing

There are several barriers to analyzing massive datasets and extracting meaningful features. The sheer size of the dataset requires significant amounts of storage space and computational power to process and analyze the data to extract meaningful features. Another barrier is the quality of the data itself. Large datasets may contain errors, inconsistencies, or missing values, which can impact the accuracy of any analysis or feature extraction. They may also be poorly documented and/or not well-formed and annotated. In addition, large datasets may be complex and diverse, with a wide range of variables and data types. This can make it

<sup>1</sup> Symbolic anchors are high-level concepts or knowledge about the conduct of a task expressed in a symbolic or declarative form (e.g., rules, concepts, or logical statements extracted during agent training in the source environment). To enhance domain adaptation (transfer of skills from one domain to another), this symbolic or declarative knowledge should also be made available to the target environment to facilitate agent training. Transfer of successful agent skills such as perception, planning, and behaviors, can be more easily transferred to new experiences in the target environment when a more comprehensive set of declarative knowledge is also transferred.

difficult to identify meaningful patterns or features within the data. Analyzing large datasets often requires a deep understanding of the domain and the specific data being analyzed. Without this expertise, it can be challenging to identify relevant features or patterns within the data. Finally, large datasets may contain sensitive information that needs to be protected. This can limit the ability to share or access the data, which can make it difficult to perform large-scale analysis or feature extraction. Assuming we are able to overcome the barriers noted above, the next step is to improve transfer learning from real entities to autonomous agents.

## IMPROVING TRANSFER LEARNING

Transfer learning is a machine learning technique that involves using a pre-trained neural network to accelerate the training of a new task or target domain that usually has limited data. The purpose of transferring features from real entities to autonomous agents is to embed the attributes, knowledge, processes, and capabilities inherent in the real entities in autonomous agents. Symbolic anchors are commonly used in human instruction as a stimulus (e.g., feedback, reinforcement, guidance) that becomes associated with a learning experience to the degree that learning can be transferred if the stimulus (prior knowledge) can also be transferred.

Symbolic anchors are usually verified or generated by a Subject Matter Expert (SME) or extracted through authoritative sources of knowledge, such as textbooks, manuals, or online resources (e.g., a repository). A data management process is needed to aid the extraction of symbolic knowledge; data pipelines are common tools for identifying features. Feature extraction is a data mining process that uses induction, concept learning and logic programming to discover underlying knowledge in the data (e.g., events, patterns/trends, rules, or concepts). The key is to extract features that can supplement the knowledge required for the agent to operate effectively and efficiently in new environments. Intuitively, this means the agent will require broad features that include linguistic, domain-specific, common-sense, cultural, historical, social, and technical knowledge. Extracting this knowledge from massive datasets will be time consuming and expensive unless there is a method to mine symbolic knowledge automatically. In our study, we seek to transfer knowledge from one setting (live training) or environment to another (virtual training) and this can be challenging if the dataset is small, albeit this is not the case for our analysis. An example of our model building process for the development of a classifier (Figure 2) shows that a much smaller dataset is required by the target environment due to the extraction and transfer of knowledge from the source environment.

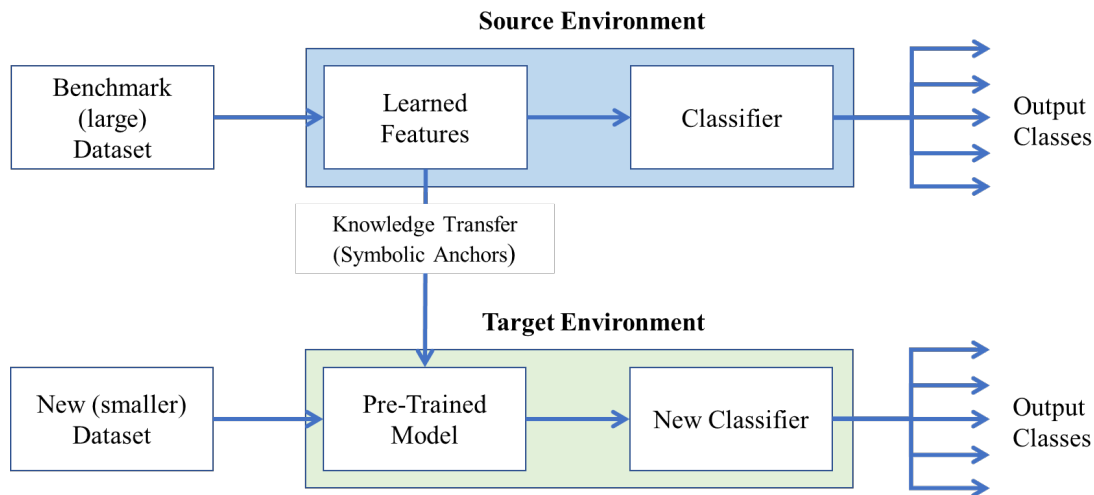


Figure 2. An Example of Transfer Learning from One Environment to Another

In developing our autonomous agents, a large list of tasks may be required to support adaptive team training. For each task required of the agent in the target environment, we will be required to create a model and transfer the necessary symbolic anchors from the source environment to ensure successful performance. A set of sample agent tasks can include:

- **Object recognition:** Identifying and classifying objects in images or videos, such as detecting and categorizing different types of objects, such as vehicles, individual Soldiers, or weapon platforms.
- **Natural language processing:** This includes tasks such as the conversion of spoken language into written text, understanding written text, or generating spoken language.
- **Sentiment analysis:** Analyzing text data to determine the sentiment or emotional tone to identify stress levels.
- **Event recognition:** Recognizing events (actions or activities) from video or other data sources to identify events during training to improve assessment accuracy.
- **Error detection:** Identifying activities from data that directly or indirectly classify errors, such as detecting variances from expected performance.
- **Recommendation generation:** Providing personalized recommendations to users, such as suggesting experiences or content based on user preferences, behaviors, or demographics.

In our case, transfer learning is focused on massive source datasets generated during live training sessions at the NTC and a target simulation. The methods described in this section are used to improve transfer learning by better aligning the source task, or domain, with the target task, or domain. The massive NTC datasets are assumed to have a high degree of alignment with the features and tasks required by intelligent agents in the target simulation. Therefore, the methods described below may or may not be suitable for our investigation.

**Task-specific data collection** is used to add task-specific data for the target task or domain and can greatly improve the performance of the transferred model. More data allows the model to learn more about the specific task or domain, reducing the reliance on the limited available data, and improving the transferability of the pre-trained features. Although the NTC datasets are massive, their adequacy to describe all of the tasks in the domain are not known, and may require the collection of additional data during NTC rotations.

**Data augmentation** artificially increases the size of the limited available data by applying various data transformations, such as rotation, translation, scaling, and flipping. Data augmentation helps in introducing more diversity into the training data, which can improve the transferability of the learned features from the pre-trained network to the target task or domain.

**Feature extraction** involves using the pre-trained neural network as a fixed feature extractor, where the learned features from the pre-trained network are used as input to a new classifier or regressor. This allows the model to leverage the high-level abstract features learned by the pre-trained network. Experimenting with different layers in the pre-trained network for feature extraction, and different classifiers/regressors on top of the extracted features, can improve the transfer learning performance.

**Fine-tuning** uses a pre-trained neural network and trains it on the target task or domain with the limited available data. Fine-tuning allows the network to adapt to the specific task or domain, leveraging the learned features from the pre-trained network. Experimenting with different learning rates, batch sizes, and optimization algorithms during fine-tuning can help improve the performance of the transferred model. Once the machine learning model is trained, its behavior needs to be defined and this involves mapping the output of the model to actions that the agent can take in the target environment.

**Domain adaptation techniques** are specifically designed to address the domain shift between the source task or domain (pre-trained network) and the target task or domain (limited available data). Domain adaptation methods aim to align the feature distributions between the source and target domains so that the transferred features are more relevant to the target domain. Domain adaptation techniques, such as domain adversarial training, domain-specific batch normalization, etc., can be used to improve transfer learning performance.

**Ensemble models** combine multiple models to make predictions. In transfer learning, using an ensemble of multiple pre-trained models or multiple variants of the same pre-trained model can improve the overall performance. Ensemble methods, such as model averaging, stacking, and boosting, can be used to combine the predictions of multiple transferred models, leading to improved performance.

**Hyperparameter tuning** involves changing the learning rate, batch size, and regularization strength, and plays a critical role in the performance of a transferred model. Experimenting with different hyperparameter settings using various techniques, such as grid search, random search, or Bayesian optimization, can help identify optimal hyperparameter values, leading to improved transfer learning performance.

By carefully applying these methods, transfer learning performance can be significantly improved, allowing for better utilization of pre-trained neural networks in various practical machine learning scenarios with limited available data. Assuming transfer learning has been optimized and we have an autonomous agent that represents the proficiency of the real entity, our next step is to examine the effect of realistic agents on team engagement in an adaptive training environment.

## **THE EFFECT OF REALISM ON ADAPTIVE TEAM TRAINING**

---

To promote learning, training environments must enable trainees to interact within the environment in a manner similar to their interactions in an operational or work environment. A lack of realism may cause trainees to discount the validity and the value of the training environment. This could cause them to disengage from the learning process (Greitzer et al., 2007) and miss opportunities to enhance their knowledge or grow their skills. The association between trainee engagement and learning outcomes, such as retention, is well documented (Axelson & Flick, 2010; Rowe et al., 2010), as is the relationship between realism and engagement (Checa & Bustillo, 2020; Tashiro & Dunlap 2007). Realism that is relevant to training objectives and fidelity is also related.

Fidelity refers to how close the training experience is to operational reality and how the experience influences trainee engagement in terms of its physical fidelity, conceptual fidelity, and experiential fidelity (Choi et al., 2017). Physical fidelity refers to the degree to which the environment approximates visual, tactile, auditory, and olfactory cues. Conceptual fidelity refers to the degree to which the training events flow compared to operational tasks. Finally, experiential fidelity refers to the degree to which the training generates responses (e.g., cognitive load, stress or emotions) that a trainee would expect in a similar operational context in the real world.

In our study, we are examining the physical fidelity of the autonomous agent in terms of their perception of the environment and their resulting behaviors. More abstractly, we are examining the agent's experiential fidelity in terms of decision making processes during the execution of assigned missions/tasks during the training. Autonomous agents perceive their environments (take in information), make decisions (e.g., formulate plans for future actions), and then act to achieve their assigned goals (Russell & Norvig, 1995). Learning agents often use rewards and penalties to drive decision making and actions to achieve increasingly better outcomes with more experience. While agent performance may be improved though



knowledge gained through new experiences, agent performance may also be enhanced by transfer learning in much the same way that humans develop skillsets that can be applied to solve a variety of problems.

Although fidelity is important in training simulations, and specifically in transfer learning, its importance may depend on the relationship between fidelity and the training objectives relevant to achieving assigned tasks. In general, higher fidelity simulations are more likely to promote realism in training and may enhance the effectiveness of the training. However, if the specific task objectives do not require a high level of fidelity, lower fidelity simulations may be sufficient and may also be more cost-effective. For example, if the task objective is to train Soldiers how to clear a building, high fidelity simulations that accurately represent a building and the surrounding environment may be necessary to provide an immersive training experience. On the other hand, if the task objective is to train for an evaluation mission, a lower fidelity simulation that focuses on the communication skills and response time may be sufficient to meet those objectives. For autonomous agents, the realism or granularity of the environment must be sufficient to support perception, decision-making, and the conduct of actions.

The degree and modality of realism in a training simulation can also have a significant impact on trainee engagement. In general, higher levels of realism can enhance trainee engagement by creating a more immersive and engaging training experience. Moreover, the specific modality of realism (i.e., visual, auditory, haptic) may also play a role in trainee engagement. For example, a highly realistic flight simulator can provide an immersive and engaging training experience for pilots by accurately simulating the visual and auditory cues of flying a plane. Similarly, a haptic feedback system that simulates the physical sensations of operating equipment can enhance trainee engagement in tasks that require hands-on skills.

Ultimately, the level of realism required to maintain trainee engagement depends on the task objectives and the expertise of individual trainees. Some trainees may require a high degree of realism to maintain engagement, while others may find a lower level of realism to be sufficient. It is also important to note that simply increasing the level of realism in a training simulation is not always sufficient to maintain trainee engagement. The simulation must also be designed to be interactive, challenging, and relevant to the trainee's goals and interests. Providing timely and relevant feedback can also enhance trainee engagement by giving them a sense of progress and accomplishment.

The relationship between realism requirements and trainee competency in training simulations is complex and can depend on several factors. In general, the degree of realism required in a simulation can impact trainee competency by providing a more accurate representation of real-world conditions and tasks. A well-designed training simulation should balance the level of realism with the trainee's needs and abilities to create an effective training experience that improves their competency.

## CONCLUSION

---

While an approach to transferring learning from real entities to autonomous agents has been put forth in this paper, the guidelines developed herein should be empirically validated in a variety of task domains and under a diverse set of environmental conditions (e.g., weather, terrain, scenario difficulty, etc.).

Transfer learning in a Machine Learning (ML) context affects the ability of an autonomous agent to adapt to new environments and influences the realism of the agent derived from real entity behaviors. This influences the learning of a trainee in an Adaptive Instructional Systems (AIS) architecture, such as GIFT, since the ability of the agent to adapt in realistic ways impacts the team training experience. Developing agents to be realistically proficient and not optimally proficient ensures that the agents will vary in proficiency in ways that at times stress trainee skills and cause failures resulting in lessons that might not otherwise be learned.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

When designing autonomous agents to optimize their realism and engagement in a GIFT-managed training environment, consider the following guidelines:

- The design of autonomous agents should align with the specific training task objectives to ensure that the agents support the learning goals.
- The agents should exhibit realistic behavior models that align with the real-world scenarios to provide an immersive training experience.
- The trainee's cognitive load should be considered to design effective agents.
- The trainee's cognitive load should be managed to avoid overwhelming them with too much information.
- Natural language processing can help create more engaging and immersive interactions between the trainees and autonomous agents.
- Incorporating social cues, such as facial expressions and gestures, can help create more engaging and realistic experiences for the trainees.
- A plan should be developed to test and evaluate agents to ensure that they are effectively supporting the training objectives and providing an engaging and realistic experience for the trainees (Hanks et al., 1993).

## ACKNOWLEDGMENTS

---

The research described in this paper is sponsored by the U.S. Army (contract W912CG21C0002). Statements and opinions expressed in this paper do not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

## REFERENCES

---

- Axelson, R. D., & Flick, A. (2010). Defining student engagement. *Change: The Magazine of Higher Learning*, 43(1), 38-43.
- Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 79, 5501-5527.
- Choi, W., Dyens, O., Chan, T., Schijven, M., Lajoie, S., Mancini, M. E., ... & Aggarwal, R. (2017). Engagement and learning in simulation: recommendations of the Simnovate Engaged Learning Domain Group.
- Greitzer, F. L., Kuchar, O. A., & Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *Journal on Educational Resources in Computing (JERIC)*, 7(3), 2-es.
- Hanks, S., Pollack, M. E., & Cohen, P. R. (1993). Benchmarks, test beds, controlled experimentation, and the design of agent architectures. *AI magazine*, 14(4), 17-17.
- Hoffman. Interview. Conducted by Robert A. Sottolare. 20 October 2022.



- Issenberg, S., Mcgaghie, W. C., Petrusa, E. R., Lee Gordon, D., & Scalese, R. J. (2005). Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Medical Teacher*, 27(1), 10-28.
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Integrating learning and engagement in narrative-centered learning environments. In *Intelligent Tutoring Systems: 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part II 10* (pp. 166-177). Springer Berlin Heidelberg.
- Russell, S., & Norvig, P. (1995). A modern, agent-oriented approach to introductory artificial intelligence. *Acm Sigart Bulletin*, 6(2), 24-26.
- Schmieder, A. (1973). A glossary of educational reform. *Journal of Teacher Education*, 24(1), 55-62.
- Tashiro, J. S., & Dunlap, D. (2007, November). The impact of realism on learning engagement in educational games. In *Proceedings of the 2007 conference on Future Play* (pp. 113-120).

## ABOUT THE AUTHORS

---

**Dr. Robert Sottolare** is the Director of Learning Sciences at Soar Technology, Inc. His research interests include using machine learning approaches to automate authoring, instructional management, and evaluation processes to enhance adaptive instructional tools and methods. Dr. Sottolare is the father and co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source adaptive instructional system architecture. He has published nearly 300 technical papers in the training and learning sciences field and has 3000 citations. He is a senior member of the IEEE. Dr. Sottolare is the recipient of multiple lifetime and technical awards from the US Army and the National Training & Simulation Association. He currently serves as the principal investigator for the LARIAT project.

**Christopher McGroarty** is the Chief Engineer for Advanced Simulation at the US Army Combat Capabilities Development Command, Soldier Center, Simulation and Training Technology Center (DEVCOM SC STTC). His research interests include distributed simulation, simulation architectures, applications of Artificial Intelligence Technologies to simulation, novel computing architectures, innovative methods for user-simulation interaction, methodologies for making simulation more accessible by non-simulation experts, future simulation frameworks, and the application of videogame industry technologies. He manages and leads a variety of research efforts that mature, integrate and demonstrate these technologies in a relevant Army and Department of Defense context. He received his Master of Science and Bachelor of Science in Electrical Engineering from Drexel University in Philadelphia, Pennsylvania.

**Dr. Christopher Ballinger** graduated from the University of Nevada, Reno with his PhD in Computer Science and Engineering. His thesis focused on using generative adversarial algorithms to solve production pipeline problems in a strategic real-time combat simulator. Since joining SoarTech, Christopher has continued to apply his expertise as an Artificial Intelligence (AI) Engineer. Christopher has demonstrated his proficiency with a multitude of AI and machine learning methodologies, including novelty search, SOAR, and neural networks, across several projects. One such project utilized novelty search to generate unique content for adaptive military training. On another project, Christopher helped author rules for an expert system that allowed entities in a training simulation to react dynamically to a trainee's decisions and behavior. Christopher has published some of his work in highly-regarded IEEE conferences and journals, including "IEEE Transactions on Computational Intelligence and AI in Games" and "IEEE Conference on Computational Intelligence and Games".

**Timothy Aris** is a Computer Engineer working for the U.S. Army Combat Capabilities Development Command - Soldier Center (DEVCOM SC) Simulation and Training Technology Center (STTC) in Orlando, FL. He has a Master of Science in Computer Science from the University of Southern California (USC) in Los Angeles, California, and is continuing his studies there towards a PhD. His research to date is focused on training reinforcement learning artificially intelligent agents to take cover from an enemy in a military simulation environment. He has two publications in the Florida Artificial Intelligence Research Society (FLAIRS) highlighting this work. He is currently helping manage the Dynamic Synthetic Behavior Generation research project, which includes a portfolio of efforts applying artificial intelligence technologies to military simulation behaviors.



# Using GIFT to Develop an Adaptive Distributed Learning Environment Supporting Data Science Competencies

Fadjimata Anaroua<sup>1</sup>, Hong Liu<sup>1</sup>, and Naomi Malone<sup>2</sup>

<sup>1</sup>Embry-Riddle Aeronautical University; <sup>2</sup>Cybermedia Technologies Inc.

## INTRODUCTION

This paper presents our latest research on the efficient support of Data Science (DS) students in higher education, particularly focusing on underserved universities. The need for new graduates and professionals to upskill in DS surpasses the capacity of universities to offer conventional classes, particularly in underserved universities (NASEM, 2018). Our solution provides otherwise unavailable DS courses for all students by implementing the Generalized Intelligent Framework for Tutoring (GIFT, Sottolare et al., 2012) to develop a multi-university adaptive distributed learning (ADL) environment that can share DS courses and facilitate student learning from anywhere at any time. This distributed learning ecosystem using Department of Defense (DOD)-initiated technologies (ADL, 2018) allows students from 11 networked universities to share courses and resources, providing equal access to underserved and better-equipped research universities within the system. Besides GIFT, the ADL environment integrates the learning management system (LMS), Moodle, competency management software such as Competence and Skill System (CaSS, 2021), and Learning Record Stores (LRSs) to collect and analyze data for personalized learning. Our instructional design and course development efficiently align learning objectives, activities, and assessments of DS student competencies based on the Edison DS Competency Framework (Edison, 2017).

In previous research (National Science Foundation Grants 1244967 and 1626602, 2014-2020), we developed and delivered five distributed learning (DL) courses using online and teleconferencing tools for 3 Underrepresented Racial Minority (URM) universities. Our latest National Science Foundation (NSF) grant (NSF Grant 2145214 2022-2025) expands this project to develop and deliver an program of 10 courses in DS for students in 11 universities, including 8 URM universities. The scope of the funded project is shown in Figure 1. The overarching goal of the newly NSF-funded project is to adapt GIFT and other DOD-initiated educational technologies to an academic setting to develop and analyze their efficacy in producing DS graduates who can succeed in the expanding DS workforce.

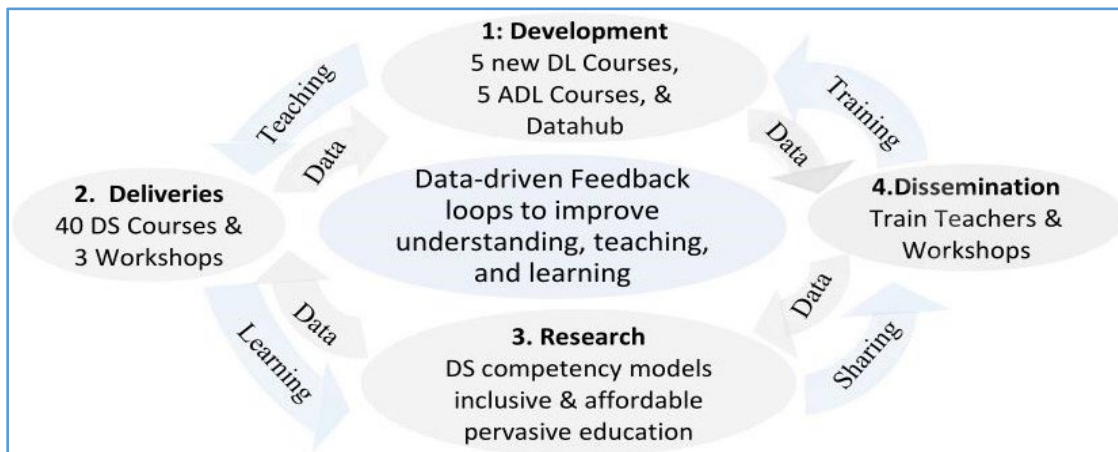


Figure 1. Data-driven formative evaluations are used to inform project progression.

There are several challenges to developing DS courses in a multi-university distributed environment. First, due to increasing class sizes, universities must already accommodate large classes that often mix in-person and remote students. Second, the interdisciplinary nature of DS courses attracts students from diverse academic backgrounds, bringing another challenge to offering courses that are customizable to student backgrounds. ADL courses enabled by GIFT not only can tailor the course to fit the individual student's learning paces and preferences but also increase teaching efficiency. Our current work focuses on automating non-cognitive tasks (Liu et al., 2022). The next phase will automate cognitive aspects of formative learning assessments: grading homework, monitoring team collaboration, and providing feedback. The goal is to automate all repeating routines and reduce instructor tasks.

This paper is organized as follows. Section 2 reviews the literature on team competency assessment in DS applications. Section 3 explains the instructional design and strategies to automate part of the assessment and aggregate lower-level competencies into higher-level ones. Section 4 describes the system integration of Moodle, GIFT, CaSS, and LRS. Section 5 presents three use-cases for students, instructors, and educational researchers, respectively. Section 6 discusses the innovation, impact, challenges, and future work of the project.

## **LITERATURE REVIEW ON TEAM COMPETENCY ASSESSMENT**

---

The primary objective of our previous and current NSF-sponsored projects is to use DL and ADL technologies to enhance active learning that is competency-based, inclusive, and cost-sensitive. Experiential learning is a popular active learning approach that boosts metacognition and reflection (Kolb & Kolb, 2009). Developed in our previously funded projects, the instructional design of our DL course materials followed the Online Learning Initiative (OLI) framework originated from Carnegie Mellon University to align learning objectives, learning activities, and learning assessments (Brooks 2017; Lovett et al., 2008). The primary tasks for content development were to supplement current non-DL course materials with case studies, applications, and data analytics projects. In addition, these DL courses included a Course-based Undergraduate Research Experience (CURE) component to promote deep experiential learning (Bangera & Brownell, 2014; Liu et al., 2018).

A major task of our current project is to transfer existing DL courses into ADL courses that follow current data interoperability standards and provide personalized and flexible learning leveraging AI, big data, and communication technologies. ADL courses also collect and analyze learner data to provide feedback loops that support learners and educators, improve courses, and advance our understanding of human learning (ADL, 2018; Bienkowski et al., 2012; Sottolare et al., 2018). In Liu et al. (2020), we presented a preliminary ADL matrix algebra course using GIFT and the Moodle LMS and stored learning activity data in the LRS.

## **INSTRUCTIONAL DESIGN AND ASSESSMENT STRATEGIES**

---

The 10 courses include five domain-agnostic courses in Introduction to Data Science (IDS), Data Visualization (DV), Mathematical Modeling & Simulation (MMS), Data Mining (DM), and Cloud Computing, as well as five domain-specific courses in Genomics and Bioinformatics (GBI), Advanced Computing Resources in Biology (ACRB), Data-driven Decision Making (DDM), Environmental Data Analytics (EDA), and Data Analytics for Public Health (DAPH). Five courses - MMS, DV, DM, GBI, and ACRB, were DL courses that were developed and offered in hybrid learning modes under the two prior funded projects from 2014-2020. The other five courses have only been offered before as traditional lecture classes in a single university with the corresponding instructor. In year 1 of the project, Embry-Riddle Aeronautical University (ERAU) will develop the first adaptive DL course (MMS). In year 2, ERAU will

create two more adaptive DL courses (DV and DM). All 5 courses listed above will be upgraded as adaptive by year 3 and adjusted based on student outcomes over the final two years of the grant.

We aim to teach students effectively by building on their prior knowledge and addressing their learning gaps. Our research method uses competency-based learning assessment and learning analytics, defined as "the use of data and analysis to understand and improve learning and its environment". We foster teamwork and timely feedback among students during problem-solving activities and promote student motivation through collaborative learning, both online and in-person. The primary theoretic challenge for our DS project is how to assess team competency in the context of DS applications (Owens & Goldberg, 2022; Salas et al., 2017; Vatal et al., 2022;).

Our competency-based learning strategy is to decompose the top-level competency from the Edison DS Competency Framework into tasks and roles of teamwork and then roll up lower-level competencies into high-level competencies reversely. At the top level, the DV course targets two competencies: 1) Applying visualization tools and DS ethics to communicate effectively for diverse audiences; and 2) identifying relevant data sources, retrieving data, cleansing data, transforming data, and warehousing data. While homework assignments, quizzes, and tests assess the basic skills and knowledge, the overall competencies are trained and evaluated through teamwork in the last five weeks of classes. The two objectives are decomposed into six lower-level weekly learning objectives, as the project process is shown in Figure 2.

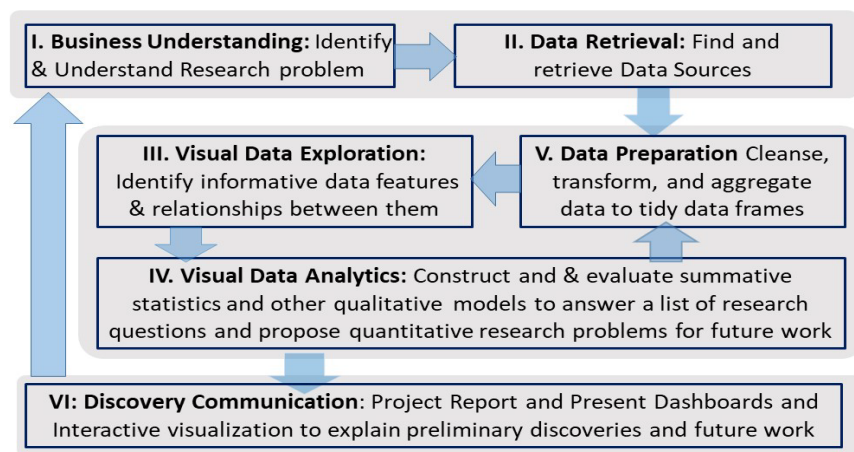


Figure 2. Team Project Process for Data Visualization Course

As Shown in Table 1, the weekly learning objectives are further divided into tasks and roles in the size that a student can complete in a week. Each team, typically 3-5 students, must use a teamwork tool (e.g., Slack) to record the team meetings and task assignments. The instructor tracks the team process and collects data for evaluation using the teamwork tool. Due to time restraints, the instructor often only has time to check records at the end of the term. Therefore, a software application called the BotCaptain is assigned an email and added to each team to perform formative assessments, monitor progress, and provide feedback. BotCaptain collects data using an xAPI (eXperience Application Programming Interface) format and stores it in a noisy LRS.

Table 1. Tasks and Roles for DV Team Project

Weekly Objectives	Roles and Task Description	Deliverables to store in the team-shared folder
<p>Week 1: Business Understanding</p>	<p><i>Domain Researchers:</i> One to two members with DS application domain knowledge other than DS, Math, or Computer Science (CS). Start literature review and internet search, brainstorm with other members and instructor to identify the research problem and possible data sources.</p> <p><i>Data Retrievers:</i> Two members with Database and programming skills. Identify, extract, and pull together available and pertinent heterogeneous data, including modern data sources such as social media data, open data, and governmental data.</p>	<p><b>Submit (1):</b> Team members and weekly meeting times are set up and posted in the shared folder.</p> <p><b>Submit (2):</b> Sign team agreement in page 7 and upload to shared folder!</p> <p>Team members and weekly meeting times are set up and posted in the shared folder.</p> <p>Who serves as domain researchers?</p> <p>Who serves as data retrievers?</p>
<p>Week 2: Data Preparation</p>	<p><i>One domain researcher</i> may continue to propose business questions that DM may help to get answers.</p> <p><i>One Data Retriever</i> may continue to retrieve more datasets and merge datasets.</p> <p><i>Data Wranglers</i>, 2 (+ depending on the need above) members with strong programming skills. Use Numpy, Pandas, and Microsoft Excel to cleanse and transform data into a tidy data frame.</p>	<p><b>Submit (3):</b></p> <p><i>One page proposal</i> about the topics of the project and data sources that your team agrees to work on.</p> <p>Who serves as data wranglers and what are the roles of other members?</p>
<p>Week 3: Data Visual Exploration</p>	<p><i>Data Wranglers:</i> Continue to cleanse and transform data based on the inputs from the explorative analyzers.</p> <p><i>Exploratory Analyzers:</i> Two members. Use Tableau, Pandas, Matplotlib, and Seaborn to identify and rank the informative values of the features to be modeled for visual analytics.</p> <p>Questions are asked to explore the data distributions, relationships, and their implications for business actions.</p>	<p><b>Submit (4):</b> upload a tidy data frame (table) that is ready for data mining.</p> <p>Who serves as <i>Exploratory Analyzers</i>, and what are the roles of members?</p> <p>What questions are to be explored?</p>

Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

		What features are selected to answer the questions?
Week 4:	<p><i>Visual Developers:</i> Each member needs to develop visual analytics <i>independently</i> to answer the questions that the team developed.</p> <p>The team will agree on which graphs will be selected and merged for the project report and final presentation.</p>	<p><b>Submit (5):</b> Preliminary Visual analytics to explain the answers to the questions of research interest.</p> <p>Have you answered the questions?</p>
<p>Week 5: Discovery communication.</p> <p>Final presentation and report due Date.</p>	<p><i>Dashboard developer:</i> Two members will be selected to develop the dashboard and merge the selected visual analytics and graphs.</p> <p><i>Writers:</i> All members will be responsible for providing technical details of the visual analytics you developed, and the writer will help merge and polish writing.</p> <p>Future work for the data mining course is encouraged.</p> <p><i>Presenters:</i> All members are in-person. The team will give a final presentation.</p>	<p><b>Submit (6):</b> Email your peer reviewer!</p> <p><b>Submit (7):</b> a white paper/report about the findings and</p> <p><b>Submit (8):</b> presentation slides, and seven meta questions to be answered. See rubrics below for evaluation.</p> <p><b>Submit (9):</b> Executable visualization code in Python and Tableau.</p>

Using CaSS to roll up lower-level competencies into higher-level competencies and match compatible competencies is a work in progress. Since the bottom-level skills and competencies associated with every learning activity are all recorded in xAPI format (“Who *did* what”), we need to configure the taxonomy of the hierarchically organized competencies in CaSS roll-up rules and compatible rules to certify higher-level competencies after a student completes a set of lower-level competencies. This system will allow automated learning assessment to make adaptive content recommendation during learning.

## SYSTEM ARCHITECTURE AND TOOL INTEGRATION

---

Universal data exchange and Learning Tool Interoperability (LTI) for long-term sustainability were our primary concerns in the development of platforms and tools. We selected the GIFT tutoring system because it is an integrated component of the Future Learning Ecosystem (Smith & Ram, 2019). Moodle was selected as a LMS because Moodle is an LTI tool that facilitates the xAPI (<https://github.com/adlnet/xAPI-Spec>) standard for data exchanges. Moodle and GIFT, can not only directly exchange data as consumers or producers but also share mutually accessible data through the xAPI data format (Hruska1 et al., 2015) and LRSs. In this section, we present the tool configuration, content deployment, and use cases, as well as the components that map the pedagogy and instructional design into the artifacts of the GIFT tutoring system.

We installed Moodle ([www.icycle.cloud](http://www.icycle.cloud)), GIFT (<http://3.12.146.191:8080/dashboard/#login>), and a LRS (<https://erau.xapi.io/>) in the EC2 servers of the Amazon Web Service (AWS) cloud. Veracity Learning, Inc.



donated the LRS and its accompanying data visual analytics tools. Figure 3 shows the system components distributed in cloud services and data exchange through three levels of the LRS. BotCaptain is an in-house developed web bot that helps instructors to collect teamwork data and perform primitive learning assessments (Liu et al., 2020). In our recent implementation, we replaced the in-house developed Natural Language Processing (NLP) component with ChatGPT. The course contents are posted on the Moodle site mostly through linked pages, while the online exercises, formative assessment, and content recommendations are delegated to GIFT (using its default pedagogical model). The learner state transition uses the three default levels - below, meet, or exceed, for the students to change state and move to the next learning activity. The student learning records are stored in a noisy LRS.

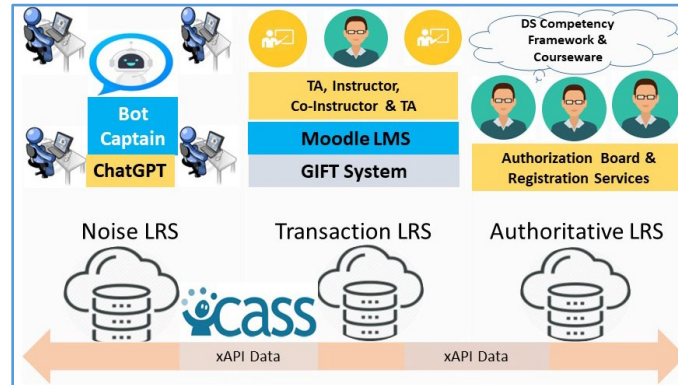


Figure 3. The Conceptual Architecture Design of the Distributed Learning Environment

At the current stage, we are focusing on automating the non-cognitive tasks, such as collecting data or reminding students about a deadline (Liu et al., 2022). Next, we will gradually automate the cognitive aspect of formative learning assessments, such as grading homework, monitoring team collaboration, and providing feedback. The ultimate goal is to automate all repeating routines and reduce instructor tasks to grading test papers and project reports at the end of the project.

## USE CASES FOR STUDENTS, INSTRUCTORS, AND RESEARCHERS

This section illustrates three use cases of our ADL environment to support students with self-paced and personalized education, instructors with evidence for iterative course improvement, and educational researchers with the data for learning analytics.

The first example use case shows how GIFT provides learning anywhere at any time. Joe is a junior Biochemistry major and student-athlete at Hampden-Sydney College (HSC). Under recommendation by his teacher, Dr. Smith, for Data Analytics for Public Health (DAPH) at HSC, Joe watches five video lessons from ERAU instructor and project Principal Investigator (PI) Dr. Liu’s Data Visualization GIFT courseware from his smartphone while traveling for a game at another institution. GIFT recommends progression from one lesson to another based on its real-time learning assessment. After Joe returns to his room that evening, he completes the assigned online homework. Each activity triggers an xAPI assertion in a noisy LRS for Joe. Several assertions collectively trigger the summative assertion in a transactional LRS, “Joe completed the Visual Data Analytics Module (Novice).” Following up on the DS courses, we form a hybrid team consisting of one pair from Dr. Smith’s DAPH class (e.g., including Joe), and the other pair from Dr. Liu’s



Data Mining Class. The collective skills of students cover both domain knowledge in DAPH and Data Mining. After Joe completes the workshop, an authoritative xAPI assertion, “Joe demonstrated expertise in visual data analytics of DAPH (Intermediate),” is certified.

The second use case demonstrates how the instructors use visual analytics derived from data in the LRS to iteratively improve the instructional design and course content. As illustrated in Figure 2, this project's emphasis is on formative measures that are likely to improve both students' and teachers' performance. For DL courses, an impact study will be conducted that includes measures of student performance and standardized tests. Each individual course will be designed to have pre- and post-tests to measure content competencies among students. Pre-tests will help instructors to develop personalized learning activities. Student preference questions on pre- and post-tests will be used to determine the degree to which learner engagement and motivation efforts were successful. We can use a dashboard LRS (donated by Veracity Learning) to visualize the learning outcomes in a transactional LRS for individual learners, teams, and cohorts to identify the problematic content and ineffective instructional designs.

Multiple delivery modes and heterogeneous student backgrounds make the instructional design of our courses particularly challenging. A one-size-fits-all best model is unrealistic. As a practical incremental approach, the instructors and researchers of this project meet in a symposium each summer to share findings and identify the “best practice so far” for each network course, allow peers to duplicate and observe the outcomes, and determine what is effective and what requires improvement.

The third use case shows how the researchers use the data to conduct learning analytics research and identify repeating routines to be automated. The proposed project related to this research will require us to log our time efficiencies to use Artificial Intelligence (AI) to relieve repetitive routines for teachers. Time cost-effectiveness is crucial for the sustainability and adaptability of proposed DS courses. In the first two years of this project, we will determine which instructor routines are suitable for potential replacement by a virtual Teaching Assistant (TA). Specifically, we will measure the instructor's time in mentoring classes and workshops as well as student learning outcomes and feedback. To accomplish this, we will classify instructors' duties into two categories: domain-agnostic routines and domain-specific instruction. The data for time efficiency will be collected from surveys of key personnel that show the hours spent on creative tasks on lecture preparation, learning assessment, lecturing, mentoring, and tutoring.

## **CONCLUSIONS AND FUTURE WORK**

---

This paper presents the large-scale practice of transferring DOD's advanced distributed learning technologies from military training into academic settings to promote pervasive DS education. Based on student-centered pedagogy, it was built on the previous work described by Liu et al. (2020). In the instructional design of this paper, we illustrated how a higher-level competency is decomposed into a set of lower-level competencies for delivering learning activities in the size of launchable units LMS compatible with the Sharable Content Object Reference Model (SCORM). Such an instructional design enables us to apply the associated technology, such as CaSS, to automate part of learning assessment and content recommendation. The progress of technology and applications could be faster, but steadfast in improving online courses' quality and reducing human intervention costs. The three use cases of the ADL environment illustrate how it provides students with self-paced and personalized education, instructors with evidence for iterative course improvement, and educational researchers with the data for learning analytics.

This NSF-funded project involves 11 institutions, 23 researchers and instructors, and at least two hundred DS mini-bachelor's degree awardees. We are developing 5 ADL courses, the first ADL course this year, two more courses in 2024, and the last two in 2025. The courses will be revised annually based on student feedback and learning outcome data. As an ongoing work in the next three years, the instructors must

transmit their course contents into the Moodle LMS so that our DS courses can be integrated with GIFT and LRSs. The AWS cloud hosting solution has the long-term benefit of using a small business model to sustain the project after the NSF fund terminates.

## ACKNOWLEDGEMENT

---

This research was partially sponsored by the National Science Foundation (NSF) of the USA under the IUSE Grant NSF Grant 2145214 2022-2025. The authors would like to thank Veracity Learning Inc. for donating the LRS software system and the colleagues in GIFT research and development for their free online technical support.

## REFERENCES

---

- ADL (2018). DOD's Advanced Distributed Learning Initiative: Total Learning Architecture: 2018, Reference Implementation Specifications and Standards, W900KK-17-D-0004, <https://adlnet.gov/projects/tla> .
- Bangera G & Brownell S.E. (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE-LSE* 13(4), 602-606.
- Bienkowski, M., Feng, M., & Means, B. (2012). Center for Technology in Learning, SRI International, U.S Education Department, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, <https://files.eric.ed.gov/fulltext/ED611199.pdf> .
- Brooks, J. (2017) Learning Design: The Simon Approach, LearnLab Summer School, Open Learning Initiative Track, July 10 - 14, 2017, [https://docs.google.com/presentation/d/1kiwK2L0WYYW0WJMJapc\\_tuNV\\_NrsFwhegJY6Y9YXM8c/edit#slide=id.p4](https://docs.google.com/presentation/d/1kiwK2L0WYYW0WJMJapc_tuNV_NrsFwhegJY6Y9YXM8c/edit#slide=id.p4)
- CaSS (2021), Competencies and Skill System by EduWorks Inc., <https://www.cassproject.org/>
- Edison (2017). Edison Data Science Framework: Part 1. Data Science Competence Framework (CF-DS), <https://creativecommons.org/licenses/by/4.0/>.
- IBM Analytics, (2020). The Data Science Skills Competency Model, A blueprint for the growing data scientist profession, <https://www.ibm.com/downloads/cas/7109RLQM#:~:text=A%20data%20scientist%20must%20demonstrate,must%20be%20good%20problem%20solvers>.
- Kolb, A. Y., & Kolb, D. A. (2009). The learning way: Meta-cognitive aspects of experiential learning. *Simulation & gaming*, 40(3), 297-327.
- Liu, H., Bernard, T., & Acharya, K. (2020). Using GIFT to Develop Adaptive Remedial Courses for Graduate Degree Programs in Data Science, Proceedings of the 8th Annual GIFT Users Symposium, 61-68.
- Liu, H., Bernard, T., Cankaya, E., & Hall, A. (2022). Task-agnostic team competence assessment and metacognitive feedback for transparent project-based learning in data science. *International Journal of Smart Technology and Learning*, 3(2), 138-162.
- Liu, H., & Malone, N. (2021). Building a Competency-Based Data Science Certificate Program Using CASS – A Prototype, Presentation accepted by iFest 2021 “Learning and Thriving in the New Normal” Virtual and In-person Conference organized by National Training & Simulation Association, Washington DC, 8/31 - 9/2, 2021.
- Liu, H, Warner, T., Ikle, M, and Mitten S. (2020) Harness Big Data by iCycle - Intelligent Computer-supported hybrid collaborative learning environment. *International Journal of Smart Technology and Learning*, 2(1), 31-47.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Lovett, M., Meyer, O., & Thille, C. (2008). The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning. *The Journal of Interactive Media in Education*, 14, 1-16.
- NASEM 2018, National Academies of Sciences, Engineering, and Medicine 2018. Data Science for Undergraduates: Opportunities and Options, Washington, DC: The National Academies Press. <https://doi.org/10.17226/25104>.
- NIST Big Data PWG (2015-1) NIST Big Data Public Working Group , Definitions and Taxonomies Subgroup, NIST Big Data Interoperability Framework: Volume 1, Big Data Definition, <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>.
- Owens, K., & Goldberg, B. (2022). Competency-Based Experiential-Expertise. In Sinatra, A., Graesser A., Hu, X., Goldenberg, B., Hampton, A., Johnson, J., (eds.) Design Recommendations for Intelligent Tutoring Systems, Volume 9, Competency-Based Scenario Design.
- Salas, E., Reyes, D.L., & Woods, A.L. (2017). The Assessment of Team Performance: Observations and Needs, In A.A. von Davier et al. (eds.), Innovative Assessment of Collaboration, Methodology of Educational Measurement and Assessment, Springer International Publishing Switzerland.
- Smith, B., & Ram, P. (2019). Interoperability. In Walcutt, J.J. & Schatz, S. (eds), Modernizing Learning: Building the Learning Ecosystem, pp. 107 – 128.
- Sottolare, R.A., Brawner, K., Goldberg, B., & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT), <https://giftutoring.org/projects/gif/documents>.
- Sottolare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of AI in Education*, 28(2), 225-264.
- Vatral, C., Biswas, G., Mohammed, N., & Goldberg, B.S. (2022). Automated Assessment of Team Performance Using Multimodal Bayesian Learning Analytics, 2022 Interservice/Industry Training, Simulation, and Education Conference (IITSEC).

## ABOUT THE AUTHORS

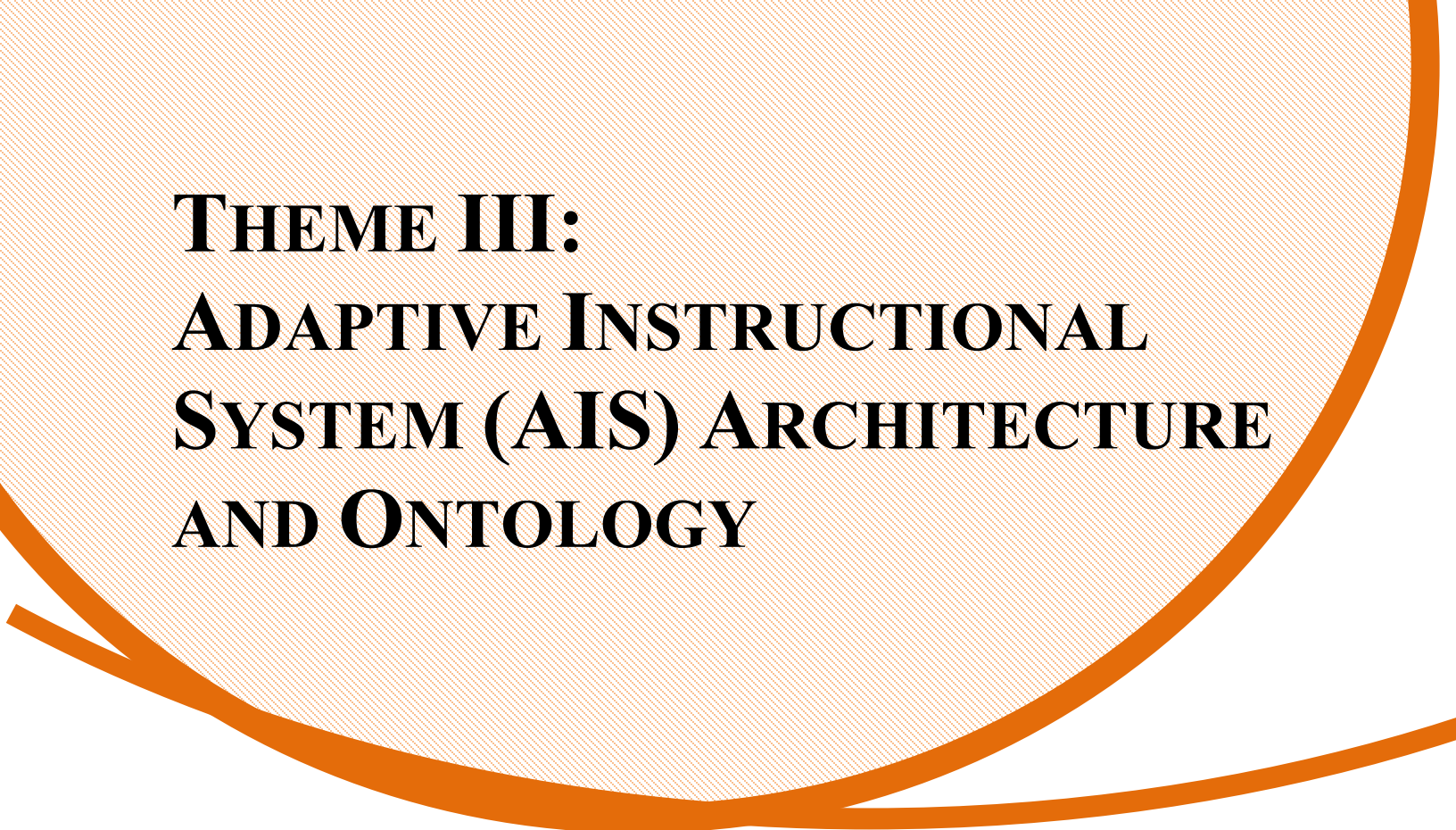
---

**Fadjimata Anaroua** is a Ph.D. student and research assistant at Embry-Riddle Aeronautical University. She was awarded a master's degree in software engineering at Embry-Riddle Aeronautical University, a bachelor's degree in Aeronautical Science, and a bachelor's degree in industrial engineering.

**Dr. Hong Liu** was awarded a Ph.D. in Mathematics and MS in Computer Science at the University of Arkansas, Fayetteville in 2000. He serves as a professor in Mathematics and Computing at Embry-Riddle Aeronautical University. He served as PI and Co-PI of 14 sponsored projects and published numerous articles in mathematics, data science, and STEM education. His current research interest is to develop a smart learning environment to promote peer learning.

**Dr. Naomi Malone** was awarded a Ph.D. in Instructional Design and Technology from the University of Central Florida in 2017. She is currently a Senior Instructional Designer at Cybermedia Technologies, designing online training for the DOD. Her primary area of interest is promoting self-regulated learning practices in online learning environments using evidence-based learning strategies and distributed learning technologies.





**THEME III:  
ADAPTIVE INSTRUCTIONAL  
SYSTEM (AIS) ARCHITECTURE  
AND ONTOLOGY**



# Cloud Technology Evolutions and Their Influences to GIFT and the STEEL-R Cloud

Ruth Feliz<sup>1</sup>, Michael Kalaf<sup>1</sup>, Christopher Meyer<sup>1</sup>, Nicholas Roberts<sup>2</sup>, and Benjamin Goldberg<sup>3</sup>

<sup>1</sup>Synaptic Sparks; <sup>2</sup>Dignitas Technologies; <sup>3</sup>US Army Combat Capabilities Development Command (DEVCOM) – Soldier Center

## INTRODUCTION

---

Adaptive Instructional Systems (AISs) have become a popular tool for personalized learning in various sectors, including education and the Department of Defense (DOD). With the advent of cloud technology, the potential for AISs to become even more effective has increased. In recent years, the DOD, private and public sector industries, and open source campaigns have made significant technological advancements in cloud technology. These advancements have provided new opportunities to deploy larger systems such as the Synthetic Training Environment Experiential Learning for Readiness (STEEL-R; Goldberg et al., 2021) and have had an undeniable impact on the deployment and availability of AISs, including the Generalized Intelligent Framework for Tutoring (GIFT; Sottolare et al., 2012) software suite.

This paper presents a subset of the technological evolutions made since 2020 in cloud technology and explores both the direct and potential impacts on AISs, specifically GIFT. We examine how cloud technology advancements can enhance GIFT capabilities, including additional training data collection methodologies, faster access methods and distribution of training data, and the potential impacts of learning technology data standards and artificial intelligence algorithms now operating using cloud technology.

Cloud technology advancements have also naturally evolved the nature of computing systems in general, including the environments that GIFT and all additional STEEL-R Cloud subsystems currently operate in. This paper presents the research conducted to explain the potential of greater scalability, flexibility, and cost savings of utilizing cloud technology. Additionally, we compare operational and deployment methodologies with regards to traditional vs. cloud technology to demonstrate the positive impact of cloud technology on GIFT and broader learning ecosystems leveraging AISs. The research conducted in this paper used mixed-method approaches, including literature review of publications authored by many current and past GIFT stakeholders, and surveys.

## Background and Purpose of Studying Cloud Technology for GIFT Enhancements

In this paper, we present how GIFT has evolved with the advent of online and cloud technologies, and how the interconnectivity of the STEEL-R Cloud system-of-systems approach has resulted in a platform that combines and adds-to the value of each individual subsystem. As a high-level motivation for describing the purpose behind the effort of the cloud modernization, to quote Hernandez et al. (2022, p.1):

“The STE (Synthetic Training Environment) Experiential Learning for Readiness (STEEL-R) project addresses the challenge of gathering and analyzing longitudinal training and performance data by establishing a common data interoperability layer that collects evidence through a competency-based experiential learning model. The STEEL-R architecture is based on and extends the US Advanced Distributed Learning (ADL) initiative’s Total Learning Architecture (TLA) to function across an ecosystem of synthetic and live training environments. This approach provides data traceability, supports evidence-based training decisions, and results in datasets that can inform acquisition teams and reduce the need to manually collect data when transitioning from research to acquisition.”

This description used in its entirety directly explains some of the STEEL-R Cloud efforts' motivations for providing a new and cutting-edge research platform by which the US Army, DOD, and educational institutions are able to access and create training content, provide adaptive tutoring and remedial feedback, plan simulated training scenarios, store and access personalized, secure, and persistent learner data, define and match competencies and skills matrices, and finally provide information to external systems (such as the Total Learning Architecture (TLA); Walcutt & Schatz, 2019) in a modular plug-and-play fashion.

We continue with a brief Literature Review and Background section, and follow with high-level details on the STEEL-R Cloud implementation, lessons learned, and recommendations for GIFT moving forward as part of a modernized and online ecosystem.

## LITERATURE REVIEW

---

### Analysis of GIFT Operations in Desktop, Locally Networked, and STEEL-R Cloud Architectures

Since GIFT's inception and initial release in May of 2012, AISs have continued to increase in popularity in educational and training settings, including in the military. This is in part due to a true AIS's ability to provide personalized and effective learning experiences through data-driven methods informed through intelligent tutoring methods. The method of deployment for GIFT has also continued to evolve, whether it is operating on a single desktop computer in a stand-alone capacity, on a local area network (LAN) for a secure training event, or online in the cloud using Amazon Web Services (AWS) in GIFT's specific case. This evolution of GIFT's local-to-cloud capability over the past decade coincides with increases in the use of adjacent and highly-correlated systems such as online learning platforms, personalized learning systems, and learning management systems (Bouchrika, 2022).

To offer a brief history of the evolution of GIFT; AISs operating on a single desktop computer with a single trainee is a common deployment tactic present across government programs. GIFT's Offline Server mode (still an available option at the time of this writing) offers the advantage of not requiring an internet connection, which can be important in certain settings where internet connectivity may not be reliable or available. However, it can also be limiting in terms of accessibility and scalability, as only one user can access the physical machine hosting the software at a time and the software cannot be easily updated or modified (sometimes even when online access is present). For more documentation regarding GIFT's different deployment strategies, please consider making a free account and refer to documentation on GIFT's homepage at [www.gifttutoring.org](http://www.gifttutoring.org).

GIFT provides a number of deployment approaches that have had profound implications on cloud-based instances of GIFT and their networking with other cloud-based applications. GIFT instances that are downloaded from GIFT's websites or checked out from GIFT's repositories typically come preconfigured to use GIFT's Desktop mode by default, which is GIFT's original mode of deployment that has existed since the original releases in GIFT. In 2015, GIFT Cloud was introduced, a publicly hosted version of GIFT that could be accessed by multiple users at any time without downloading any GIFT software. This release saw the introduction of a new Server deployment mode, which was designed to address limitations of Desktop mode that were found upon attempting to deploy GIFT to a cloud-based solution. Server mode uses a dedicated content management system (CMS) that allows users to have their own workspaces within GIFT that other users cannot use without appropriate permissions, whereas Desktop mode uses the file system of the machine GIFT is on and has no limitations on workspace access. Server mode also handles GIFT's Gateway module differently, since Desktop mode only allows one user to use the Gateway module



at a time to interact with training applications during a course, while Server mode allows multiple users to each connect to the Gateway module at once in different training sessions.

The GIFT software suite can also be deployed to offer training content and adaptive tutoring to a LAN as well as via a single machine on the network, which allows multiple users to access the software simultaneously within a restricted area which may also be made available online if the LAN/web servers are configured correctly. This method can provide greater accessibility and scalability than operating on a single desktop computer, but it still requires physical access to the LAN and may not be accessible from outside the network.

Finally, to increase general security, availability, and accessibility even further, the GIFT software suite was deployed to the AWS cloud computing platform and iteratively enhanced for cloud compatibility during 2017-2019. The COVID-19 pandemic ushered in multitudes of performance-based studies and research regarding instruction in cloud or hybrid-cloud environments, and one example of a representative study that directly mirrors considerations for GIFT was presented by Barbierato et al. (2021). Of special note from Barbierato et al. (2021) is the proposition of a mathematical formula to gauge the overall “success” of local, hybrid, and pure-cloud framework deployments that could be used to measure GIFT’s effectiveness in the future. Thankfully, GIFT was deployed in a mature state and operating smoothly in AWS by the time the COVID-19 pandemic arrived, as many benefits that were proposed as hypothetical in similar educational research studies were already proven to be true for GIFT’s operations in the AWS cloud. STEEL-R Cloud predictably follows the deployment approach taken by GIFT Cloud, albeit with some evolutions to facilitate GIFT’s interactions with other STEEL-R applications in the cloud. The measure of these benefits are discussed later in the paper, and cover objective measures of cost, availability, security, and performance.

## **Overview of Porting GIFT to Cloud Virtual Machines and Recent Cloud-Based Technological Evolutions**

During the initial modernization of GIFT from desktop/web server deployments to an AWS cloud deployment, the benefits of cloud-based software, such as increased scalability, flexibility, and accessibility, were primary motivations. However, the process of modernizing a desktop application to a cloud-based implementation can be challenging and requires careful planning and execution.

One of the key changes that needed to be made when transitioning to the cloud was to move away from local data storage to a distributed storage system. In AWS, these considerations led the development team to create a trio of GIFT instances in the cloud to better-manage these architectural updates. One primary and upper-tier AWS EC2 instance running Windows was initially required for the GIFT server, another lower-tier EC2 instance running AWS Linux was deployed to host GIFT’s Nuxeo instance (responsible for administration, individual user course content and tracking, among other features), and finally an AWS Relational Database instance (RDS) was required running PostgreSQL. This shift allowed for better scalability and redundancy, as well as improved data access and security. For years, GIFT Cloud has traditionally been run using an AWS-based Windows Server instance, but based on recent changes to GIFT’s scripting in the 2022-1 baseline, it was decided to use a Linux-focused distribution of GIFT running in Server mode on an Ubuntu AWS virtual machine. This decision was partly motivated by standardization, as most of the STEEL-R Cloud applications other than GIFT have traditionally run on Linux-based platforms, either directly on a Linux server or through Docker containers. Using a Linux-based version of GIFT allowed settings and access rules to be more easily shared between GIFT and the other STEEL-R applications, easing the shared use of these applications in a cloud environment. As a secondary motivation, GIFT’s ongoing effort to integrate GIFT with STE-IS (Synthetic Training Environment Information

System) helped shape GIFT's Linux support in the first place, making a Linux-based distribution a logical choice to use with STE-focused systems as STEEL-R evolves going forward.

GIFT Cloud already relied on some Linux Virtual Machines (VMs) to handle its CMS and other supporting tools, and with the scripting changes from the recent release, nearly all build and execution settings are now configured through Java, allowing them to work identically on Windows and Linux. The most involved portion of this porting process surrounded translating GIFT Cloud's Internet Information Services (IIS) settings in Windows Server to similar configurations in a Linux-based Apache server. Initially, it was unclear how to port GIFT Cloud's rules for redirecting URLs and enforcing HTTPS connections when connecting to GIFT's internal Java servers, but after cross-referencing other Linux-based servers like the one hosting the GIFT Portal, a solution was found by using Apache proxy rules in the server's launch configuration.

The GIFT development team also needed to ensure that the GIFT software suite was optimized for the cloud environment, including considerations for auto-scaling, load balancing, and fault tolerance. While GIFT utilizes only the cloud features that make functional sense, the modernization process required a deep understanding of the underlying cloud platform and its capabilities, as well as the ability to make the necessary changes to the application code and configuration. Finally, the GIFT development team had to consider operational aspects of the cloud-based implementation, including deployment, monitoring, and management. The move to AWS required new tools and processes to be put in place (for instance, regression testing processes, branch/feature testing, continuous integration/continuous deployment processes), as well as new skills and expertise to be developed and maintained within the team.

For more information on setting up your own GIFT software suite in AWS, please refer to the AWS GIFT Deployment Guide here: [https://gifttutoring.org/projects/gift/wiki/Deploying\\_to\\_AWS](https://gifttutoring.org/projects/gift/wiki/Deploying_to_AWS), and the reader may also refer to <https://docs.aws.amazon.com> for in-depth AWS documentation on specific topics.

## **Analysis of Modern Cloud Computing Capabilities for Adaptive Instructional Systems**

AISs are designed to provide personalized learning to individuals based on their abilities, preferences, and learning styles. The effectiveness of an AIS is influenced by the availability of data, computational resources, and software capabilities – of which AWS offered completely flexible options that the GIFT development team was able to customize to programmatic requirements. In recent years, the DoD has also been exploring the use of AWS cloud computing to enhance their capabilities and effectiveness (<https://aws.amazon.com/government-education/defense/>).

Another advantage of cloud computing for AISs is the ability to provide real-time access to training data and insights. In GIFT, these features are enabled through Course Authoring, Experiment Setup, and After Action Reviews to name just a few capabilities. This enables instructors to monitor learners' progress and provide timely feedback, sometimes during live (and simulated) external training application scenarios. The availability of data on a cloud platform also allows for collaboration among developers and trainers, leading to more effective training programs. Moreover, cloud computing platforms can also provide institutionalized security and reliability for AISs by enabling these services in the cloud computing platform itself (such as the capability to link login keys to development computers in AWS, or create IP whitelists for access control). Cloud providers invest heavily in security infrastructure and offer reliable data backup and disaster recovery solutions, which are critical for AISs that handle sensitive military data. The GIFT software suite is not currently setup to provide Sensitive or Classified data handling, but the software suite has been architected with the framework for enabling such data protection services in the future, independent of cloud capabilities.

Cloud computing offers several benefits for the development and deployment of AISs for the DoD. The cloud provides access to on-demand computing resources, storage, and data analytics capabilities, which have been proven in GIFT's case to enhance the effectiveness and efficiency of AISs. The GIFT software suite is just one example of a successful AIS that leverages cloud computing to provide personalized training to military personnel. The DoD is continuing to explore the use of cloud computing for AISs to further enhance their capabilities and effectiveness in training and mission support. To learn more individuals with interest in cloud technology or standards development may participate in working groups such as the Adaptive Instructional Systems (C/LT/AIS) P2247 Working Group enabled by Advanced Distributed Learning (ADL, via <https://adlnet.gov/working-groups/adaptive-instructional-systems/>), or through the Army DEVCOM Soldier Center's current efforts, including GIFT, via <https://sc.devcom.army.mil/>.

## DISCUSSION

---

To fully understand the following discussion it may be beneficial to also study past efforts regarding GIFT and the STEEL-R project. The most recent local network implementation for a system such as the STEEL-R Cloud as of the time of this writing was the demonstration setup for the I/ITSEC November/December 2022 conference, where a LAN was setup on the trade show floor physically hosting the STEEL-R Cloud subsystems defined in Hernandez et. al. (2022) and Owens et al. (2022).

The new research framework setup under the STEEL-R Cloud umbrella can be accessed at <https://steel-r.org>. Please note that this web address is currently hooked to the STEEL-R Cloud system in the early stages of research and development. This means that the user experience may change, go offline at unannounced times, be reset with no notice, and data stored by any user may be purged, until further notice is given, as the STEEL-R Cloud platform matures.

We will now present and briefly explain the processes that were necessary to bring the following to the AWS Cloud: GIFT, xAPI/XI (eXperience Application Programming Interface/eXperience Index) and Learner Record Stores (LRSs), the Competency and Skills System (CaSS), the Experiential Design Tool (XDT), and supplementary support systems enabling further capabilities in STEEL-R Cloud.

### **Building and Maintaining a STEEL-R Cloud Platform**

Overall, the move to the new cloud platform represented a significant step forward for GIFT as part of the STEEL-R Cloud and its ability to deliver effective and personalized training content to service members. By leveraging the capabilities of cloud computing, GIFT continued to improve its scalability, flexibility, availability, and security as definitively measured in the GIFT monitoring instance and via AWS Cloud Monitor tools. This ensured that the GIFT (and STEEL-R) learning community and specifically service members will, once all STEEL-R Cloud systems reach a mature and integrated state, have access to high-quality training that is tailored to their needs and delivered in a timely and effective manner that is now linked to specialized content, skills and competencies, persistent learner data storage, and centralized accounts allowing for navigation by the same user across STEEL-R Cloud subsystem domains.

As part of this effort, GIFT and STEEL-R Cloud developers needed to maintain modular software in AWS, and were able to use several tools and services provided by AWS, such as AWS CloudFormation, EC2/RDS, CloudWatch, Route 53, Certificates and Systems Manager, and AWS Lambda. AWS CloudFormation is a service that allowed STEEL-R Cloud developers to define and manage infrastructure resources as code. It provided a simple way to manage and deploy infrastructure changes across multiple environments, ensuring consistency and reducing the risk of human error particularly with the LRS portion of the STEEL-R Cloud. Finally, AWS Lambda is a serverless computing service that enabled developers

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

to run code without provisioning or managing servers, providing a scalable and cost-efficient way to maintain modular software in the AWS STEEL-R Cloud.

Each STEEL-R Subsystem maintains its own operational environment in order for each software suite to acquire necessary system resources that may cause conflicts if attempting to integrate the software suites too tightly together (such as compiling and running a software suite from source code on the same personal computer (PC)/instance as another software suite).

The development team's theory behind the value of the integration of these systems in the AWS Cloud was that the effort could create a seamless, end-to-end training ecosystem. GIFT and CaSS provide adaptive learning experiences that are tailored to individual Soldiers' needs and competencies. The YetAnalytics LRS captures and stores learning data from these systems in an xAPI format, and provides analytics capabilities to derive insights into the effectiveness of the training and also provides interface endpoints for reading/writing learner data via the XI. The XDT provides the STEEL-R Cloud the capability to add experiential exercise design including the capability for an instructor to customize multiple training variables. All of these systems have been integrated together to provide software hooks for external training applications, for instance scenarios in Unity3D or Virtual Battlespace 4 (VBS4), that assist in providing a realistic virtual environment for Soldiers to apply their skills and competencies learned from GIFT and CaSS.

As the STEEL-R Cloud platform continues to mature past the initial prototype phase, the reader is encouraged to frequently check back with the authors or directly at <https://steel-r.org>.

### ***STEEL-R Cloud Subsystems***

The STEEL-R Cloud Subsystem has maintained stability as compared to the definitions provided in a relevant paper presented in GIFTSym10, defined in Owens et al. (2022). Specifically, the following STEEL-R Cloud Subsystems are still present in the overall architecture and roughly adhere to past definitions, with a caveat that each of the systems are undergoing current integration efforts as part of STEEL-R Cloud:

- Experience Design Tool (XDT)
- xAPI Profiles / eXperience Index (XI)
- Generalized Intelligent Framework for Tutoring (GIFT)
- Competency and Skills System (CaSS)

Additionally, each of the STEEL-R Cloud subsystems listed above has undergone optimizations, authentication, and messaging system updates as the online AWS Cloud implementation of STEEL-R Cloud mimicked the capabilities of the I/ITSEC 2022 STEEL-R floor demonstration. These updates are in the process of being formally delivered to the Simulation and Training Technology Center (STTC) and presented to DOD stakeholders and the general learning community should the project receive authorization to do so.

As part of the STEEL-R Cloud communal integration efforts; GIFT's Java servers were provided with new configurations and a new UserAuthenticationInterface that uses Keycloak's Jetty server adapter. GIFT's Dashboard was also changed to bypass GIFT's usual login page in favor of the login page provided by STEEL-R Cloud's Keycloak Server. This allows users to sign in and interact with other STEEL-R applications and then seamlessly jump into GIFT's webpages without being forced to sign in again with

GIFT Portal credentials. Altogether, the effort evolved to support the STEEL-R Cloud instance of GIFT and led to a number of improvements to GIFT's Server mode logic that provide GIFT developers with more options with how GIFT can be deployed and integrated with other web-based solutions.

In order to centralize and secure the STEEL-R Cloud landing site, the development team also configured middleware communication tools that utilized common network design patterns (Meyer et al., 2022). Extending the capability for communication, the development team added authentication and centralized data store frameworks to allow users to maintain a Single Sign-On (SSO) style of interaction with and between the different subsystems, allowing a single user's account to be associated across the breadth of their experience in STEEL-R Cloud.

### ***New Benefits of the STEEL-R Cloud Implementation***

The STEEL-R Cloud, even in its implementation infancy during the 1<sup>st</sup> Quarter of 2023, has already provided several items of value to program stakeholders. Namely;

- **Cost Effectiveness:** Cloud-based software implementations of the associated suites eliminated the need for expensive hardware, software licenses, and IT (Information Technology) infrastructure for each organization responsible for deployments. This reduced upfront costs and enabled a pay-as-you-go model, where costs are incurred only for the resources used. In AWS, these cost savings are recognized across every phase of a project, additionally including savings due to quicker deployment times (once a software suite is architected to be cloud-compatible), quicker access to resources necessary for system administration, and completely offloading any cost associated with cloud system administration that would be incurred with local-only implementations.
- **Accessibility:** STEEL-R Cloud can be accessed from anywhere with an internet connection, which made working on a project with this scope possible, as developers across different demographics and work locations needed to collaborate and work seamlessly together on it. Working together across disparate organizations proved to be particularly useful, especially in a post-COVID-19 workforce, as many employees work remotely or have multiple offices in different locations. By hosting software in the cloud, developers were able to access the software from any location, using any device, without the need for local software installations.
- **Another significant benefit of implementing STEEL-R Cloud in AWS is enhanced security.** AWS, which was used to secure STEEL-R Cloud instances, offers a range of security features and services that help protect against threats such as malware, hacking, and data breaches. AWS has a shared responsibility model, in which AWS is responsible for the security of the cloud infrastructure, while customers are responsible for securing their own applications and data. However, AWS provided a range of security services, including identity and access management, network security, encryption, and compliance certifications, to help developers of STEEL-R secure the applications and data. Please note that formal programmatic review of data security for STEEL-R Cloud has not occurred yet, so no sensitive data should be entered in by any user experimenting with the STEEL-R Cloud applications.
- **The new capability for SSO in STEEL-R Cloud allows for all of the benefits (and risk) of consolidating multiple user accounts into one.** In past implementations, the subsystems in the STEEL-R Cloud were all independent of each other and required specific, new interfaces to be designed when data was required to move between software suites. In the current implementation, developer teams now have the option of communicating with the central authentication server (running a software program called KeyCloak), to obtain secure access tokens associated with mapped user accounts across subsystems. This allows for a more seamless user experience as

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

different Plan/Prepare/Execute/Assess workflow stages are accessed for different capabilities across STEEL-R Cloud.

- **Reliability:** AWS cloud provides a 99.9999999% uptime claim (Nine 9's of reliability) and features that ensure that software applications remain accessible even during infrastructure failures or disasters. On the STEEL-R Cloud instances and over 5 years of GIFT Cloud operations, the development team can corroborate this claim by AWS. This results in minimal downtime and ensures that critical business operations continue uninterrupted at all times.
- **DOD Programmatic Compatibility:** As mentioned in Goldberg et al. (2021), even a project with the large scope of STEEL-R Cloud can be architected in ways to increase compatibility with existing ecosystems such as STE-IS, and by association its components such as the Training Management Tools (TMT), Training Simulation Software, and One-World Terrain. Through adapting larger program's security, accessibility, and design patterns, the STEEL-R Cloud team has worked to ensure that the integration points of all STEEL-R Cloud subsystems not only maintain flexibility as individual parts, but also architect the entire design to be as modular as possible. Paradigms such as these, along with guidelines from organizations such as ADL with regards to documentation on the TLA form a rich set of foundational guidance with which STEEL-R Cloud has been designed to (Smith, et. al., 2021).

Data supporting the above claims has been tracked on three different levels; namely programmatic (cost/schedule), technical (minimum viable criteria, system modernization feasibility, mapping functionality-to-requirement-to-value), and via AWS (uptime, resource utilization, Cloud Watch information). This data has been deemed Controlled Unclassified Information, and thus will not be formally included as part of this publication. However, the reader is encouraged to contact the organizations/authors listed in this paper if the reader believes they have a Need to Know and proper credentials.

### **Risks of Cloud Technology and GIFT/STEEL-R Cloud**

While using AWS cloud services for software development and deployment has many benefits, there are also some risks and downsides to consider. One significant risk is the possibility of data breaches or cyber-attacks, as storing data in the cloud can make it vulnerable to unauthorized access. It is essential to continue to implement robust security measures, such as secure login keys and access controls as STEEL-R Cloud already has, to mitigate this risk.

Another potential downside is the cost of using cloud services (contrary to the likely cost-savings with properly implemented cloud solutions), as fees can quickly add up depending on usage and storage needs. Additionally, if the software requires high levels of processing power or data transfer, the cost of using cloud services may be higher than maintaining a local infrastructure. As experimentation with STEEL-R Cloud continues, the development team may have additional data that spikes costs upwards during particularly demanding training scenarios with extended times or large numbers of users accessing STEEL-R Cloud.

Furthermore, relying on third-party cloud services such as AWS can also present potential issues with vendor dependency, where it may be challenging to switch to a different provider or move the software back to a local location. While GIFT and STEEL-R Cloud currently experience large savings identified, from decreased labor needs to hardware savings, many commercial or larger-scale applications could see cost increases depending on the relationships of those factors to the number, type, and engagement of users for the hypothetical software.

Modular software development and implementation also have their risks and downsides. One potential risk is the complexity of managing and coordinating multiple interfaces and message formats, which can lead to compatibility issues and delays in development. Additionally, modular software may require more significant upfront planning and design to ensure that the modules are compatible and can work together seamlessly as with STEEL-R Cloud, that had multiple teams working across multiple organizations with high quality contact to the stakeholders.

Finally, the use of the TLA or other standards for software development can also present some risks and downsides (though the teams at ADL work specifically to combat the following hypothetical risk). For example, adherence to strict standards can limit the flexibility and creativity of developers, potentially leading to less innovation and slower development times if not provided/imposed correctly. Additionally, there may be some resistance to adopting new standards and technologies, particularly if they require significant changes to existing systems or workflows.

## **CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

---

Cloud computing has revolutionized the software industry, and with the availability of services like AWS, it has become easier to create and maintain software in a modular way. Moving GIFT to the AWS cloud has provided many documented benefits over the past 5 years such as increases to scalability, accessibility, and security. And, when considering the added functionality of GIFT when used together with the XDT, CaSS, LRS/XI, and the additional middleware components to enable STEEL-R Cloud, the development team can perform new categories of experiments with a wide new audience. With new capabilities being planned and developed every day such as SSO, users only need to log in once to access multiple applications, which reduces the risk of unauthorized access and consolidates the learner experience. Additionally, the use of modular software architecture in the AWS cloud enables compatibility with future DOD efforts such as the TLA, STE-IS, and the TMT.

However, moving operations from a local or network implementation to the cloud also has some risks and downsides, such as the potential for data breaches and a loss of control over data. It is important to take appropriate measures to secure the data and maintain control over access to it using available AWS security measures, mature system administration techniques, and training employees to avoid being targeted by hostile social engineering efforts at all costs. Furthermore, there may be compatibility issues between different software suites when operating together on the AWS cloud if integrated too tightly. And finally, the need for constant internet connectivity is another potential limitation for cloud-based software which the development team has mitigated by having options such as offline data collection be feasible.

### **Implications for GIFT, STEEL-R Cloud, US Army, and DOD, and Future Research**

The move towards cloud-based computing has the potential to revolutionize the way that educational and training software is developed, delivered, and managed. In the case of a cloud platform like STEEL-R Cloud and GIFT, requirements across the system vs. subsystem domain must always be accounted for and kept in-sync as much as possible, as mutual benefit should always be a goal between all levels of development. This understandable shift during and after the COVID-19 pandemic also presents some potential challenges and risks, such as concerns around data security, reliability, and accessibility. To mitigate these risks, future research could continue to focus on the development of best practices for software developers, possible implementations of other secure network integrations such as AWS GovCloud or FedRAMP, as well as ongoing monitoring and evaluation of cloud-based systems to ensure that they are meeting the needs of the military, and the learning community and educators at large.

Another area of potential future research is the development of new and innovative applications for cloud-based educational software. For example, STEEL-R Cloud, heavily utilizing GIFT, could now further investigate the benefits of using cloud-based systems together to support personalized learning or adaptive instruction with the XDT and training applications, or explore how cloud-based analytics tools could be used to track student progress and identify areas where additional support may be needed with heavier integration with the LRS and CaSS. Additionally, GIFT researchers could examine the impact of cloud-based software comparing GIFT to STEEL-R Cloud GIFT on student/Soldier outcomes and engagement, or investigate how different implementation strategies and architectures impact the effectiveness of these systems.

A final and slightly more-technical area for future research is the development of new technologies and approaches to enhance the security, accessibility, and reliability of STEEL-R Cloud. For example, researchers could investigate new encryption and authentication methods to improve data security that could feed “up” into larger paradigms such as the TLA or STE-IS, or explore strategies for improving the availability and accessibility of cloud-based systems for Soldiers in remote or under-resourced areas. Additionally, developers could investigate strategies for reducing latency and improving system performance, such as the completion of the port of GIFT Windows to GIFT Linux.

The authors conclude that nearly equal amounts of research can be done to continue adding to the value of STEEL-R Cloud while, in parallel, performing research and staying updated to address and mitigate risks. Moving forward, future research should focus on improving the interoperability and compatibility of different STEEL-R Cloud software subsystems, as well as exploring the potential of emerging technologies such as cloud-based machine learning and artificial intelligence to enhance the capabilities of GIFT.

## ACKNOWLEDGEMENT

---

This material is based upon work supported by the US Army Combat Capabilities Development Command Soldier Center (DEVCOM SC), Training and Simulation Division (TSD), SFC Paul Ray Smith Simulation and Training Technology Center (STTC) under contract No. W912CG-22-C-0015.

## REFERENCES

---

- Bouchrika, I. (2022). 51 LMS Statistics: 2023 Data, Trends & Predictions. Research.com: Educational Sectors. <https://research.com/education/lms-statistics>. Accessed April 30, 2023.
- Barbierato, E., Campanile, L., Gribaudo, M., Iacono, M., Mastroianni, M., & Nacchia, S. (2021). Performance evaluation for the design of a hybrid cloud based distance synchronous and asynchronous learning architecture. *Simulation Modelling Practice and Theory*, 109, 102303. <https://doi.org/10.1016/j.simpat.2021.102303>
- Goldberg, B., Owens, K., Gupton, K., Hellman, K., Orlando, E., Austin, F., Leavenworth, F., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy US Army TRADOC Proponent Office Synthetic Training. Retrieved May 1, 2023, from <https://gifttutoring.org/attachments/download/4295/21332.pdf>
- Hernandez, M., Blake-Plock, S., Owens, K., Goldberg, B., Robson, R., Welch, T., & Ray, F. (2022). Enhancing the Total Learning Architecture for Experiential Learning. Retrieved May 1, 2023, from [https://gifttutoring.org/attachments/download/4745/2022\\_iitsec\\_22461\\_steelr\\_Enhancing\\_TLA.pdf](https://gifttutoring.org/attachments/download/4745/2022_iitsec_22461_steelr_Enhancing_TLA.pdf)
- Meyer, C., Roberts, N., Kalaf, M., Heylmun, Z., Sinatra, A. (2022). A Review of GIFT Network Interface Paradigms and Potential Future Directions. Retrieved May 1, 2023, from [https://gifttutoring.org/attachments/download/4563/GIFTSym10\\_Proceedings\\_Final\\_052722.pdf](https://gifttutoring.org/attachments/download/4563/GIFTSym10_Proceedings_Final_052722.pdf)



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Owens, K., Goldberg, B., Robson, R., Hoffman, M., Ray, F., Colburn, A., Hernandez, M., Blake-Plock, S., & Casey, C. (2022). Synthetic Training Environment Experiential Learning for Readiness (STEEL-R) Demonstration. Retrieved May 1, 2023, from [https://gifttutoring.org/attachments/download/4643/GIFTSym10\\_Proceedings\\_STEELR\\_demo\\_paper.pdf](https://gifttutoring.org/attachments/download/4643/GIFTSym10_Proceedings_STEELR_demo_paper.pdf)
- Smith, B., Schatz, S., & Turner. (2021). Total Learning Architecture Data Model for Analytics and Adaptation. Retrieved May 1, 2023, from [https://adlnet.gov/assets/uploads/2021%20---%20\(IITSEC\)%20Total%20Learning%20Architecture%20Data%20Model%20for%20Analytics%20and%20Adaptation.pdf](https://adlnet.gov/assets/uploads/2021%20---%20(IITSEC)%20Total%20Learning%20Architecture%20Data%20Model%20for%20Analytics%20and%20Adaptation.pdf)
- Sottolare, R., Goldberg, B., Brawner, K., & Holden, H. (2012). A Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems (CBTS). Retrieved May 3, 2023, from <https://gifttutoring.org/attachments/download/144/12017.pdf>
- Walcutt, J. J., & Schatz, S. (2019). Modernizing Learning: Building the Future Learning Ecosystem. Advanced Distributed Learning Initiative.

## ABOUT THE AUTHORS

---

***Ruth Feliz** is a highly skilled and experienced professional in strategic project management and finance. She has a proven track record of successfully coordinating and overseeing various projects and strategic initiatives across continental and global organizations. She holds two Master's degrees, the last of which was a Global Master's degree in Business Administration with a particular focus on corporate finance and data analytics from Tufts University's Fletcher School of Law and Diplomacy. For over 10 years, she has applied her extensive experience in cross-cultural communications to multiple projects within the profit and non-profit sectors.*

***Michael Kalaf** has over 30 years of Modeling, Simulation and Training leading large scale efforts leveraging cutting edge technology. Mike has worked in the commercial and military aviation, training, and simulation business. Mike has led several programs integrating "state of the art" technology and delivering highly successful technology and business innovation. Mike's formal education includes an earned Mechanical Engineering degree from Rochester Institute of Technology, RIT.*

***Christopher Meyer** received his Bachelor and Master of Science degrees in Computer Science from Kansas State University, also receiving minors in Economics and Modern Languages. Chris also studied abroad in Japan at Chukyo University dedicated to the specialized study of Artificial Intelligence. After completing traditional education phases, Chris was employed at Lockheed Martin for 10 years working together with representatives from the Departments of Defense, Health and Human Services, Energy, and Education to assist in the creation of solutions to solve challenges at a national level before co-founding a 501c3 Non-profit Organization called Synaptic Sparks and supporting the GIFT program.*

***Nicholas Roberts** is a senior software engineer at Dignitas Technologies and the engineering lead for the GIFT project. Nick has been involved in the engineering of GIFT and supported collaboration and research with the intelligent tutoring system (ITS) community for nearly 10 years. Nicholas contributes to the GIFT community by maintaining the GIFT portal ([www.GIFTTutoring.org](http://www.GIFTTutoring.org)) and GIFT Cloud ([cloud.gifttutoring.org](http://cloud.gifttutoring.org)), supporting conferences such as the GIFT Symposium, and participating in technical exchanges with Soldier Center and their contractors. He has also been heavily involved in integrating GIFT into TSS/TMT.*

***Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.*



# Defining and Measuring Competencies under Varying Conditions

Robby Robson<sup>1</sup>, Fritz Ray<sup>1</sup>, Benjamin Goldberg<sup>2</sup>, Shelly Blake-Plock<sup>3</sup>, Cliff Casey<sup>3</sup>, Kevin Owens<sup>4</sup>, and Sazzad Nasir<sup>1</sup>

<sup>1</sup>Eduworks; <sup>2</sup>US Army Combat Capabilities Development Command (DEVCOM) – Soldier Center; <sup>3</sup>Yet Analytics, <sup>4</sup>Applied Research Laboratories: The University of Texas at Austin

## INTRODUCTION

---

Conditions such as terrain, lighting, and time pressure can significantly affect the tools and methods used to perform a skill. For this reason, conditions must be reflected in the definition of skills and competencies and in the tools and models used to measure them in military, workplace, sports and other applications that involve experiential and context-based learning (Kolb, 1984; Rose, 2012). This paper discusses how conditions are factored into competency and skills frameworks, measurement, and learner models used in one such application – the STE Experiential Learning - Readiness (STEEL-R) project.

## BACKGROUND

---

Traditional education takes place in controlled environments that are intended to be predictable and relatively free of distractions. Whereas such environments may be effective for acquiring knowledge and cognitive skills, they are not often designed for learning to apply skills in real-world situations or for acquiring psychomotor or affective skills. Proficiency in such skills is aided by engaging in deliberate practice under a variety of conditions, which we call *Guided Experiential Learning* (Clark, 2005). Guided experiential learning is core to on-the-job, military, athletic, and similar types of training.

In STEEL-R, Army doctrine is translated into multi-level competency and skills frameworks that can contain granular skills such as “controlling breathing while aiming a weapon, high level tasks such as “enter and clear a room,” and transferable team-oriented skills such as “communicates effectively” and “recognizes and considers the capabilities of team members when making assignments.” The ability to apply these skills can be functionally different under different conditions. To ensure readiness, STEEL-R seeks to ensure that Soldiers, teams, squads, etc. are provided with deliberate and focused practice under a representative set of conditions and that they have achieved the level of proficiency in each developmental phase before progressing to the next one. These conditions are reflected in the data streams, algorithms, and models that are used to estimate skill and competency levels. The developmental phases in STEEL-R, designated as *crawl*, *walk*, and *run* in accordance with Army practice, correspond to a progression from fully synthetic to mixed reality to live environments, i.e., a progression in which readiness must be demonstrated at each phase to increase the effectiveness of training (and reduce the risk of physical injury and failure) in the next one.

As described in (Goldberg et al., 2021; Hernandez et al., 2022; Owens et al., 2022; Robson et al., 2022), STEEL-R uses multiple components to implement a data strategy. The data flow in STEEL-R can roughly be described as follows:

1. Activities observed in training scenarios are converted into evaluations of performance on tasks.
2. All activities and evaluations are reported and collected via xAPI (eXperience Application Programming Interface) statements.

3. These are processed into standardized xAPI statements that report performance on tasks and behaviors.
4. These are converted to assertions about skills and competencies in a competency and skills framework.
5. These assertions are used to estimate the skill and competency level of each learner (or group of learners).
6. Skill acquisition is tracked over time and displayed in a dashboard.
7. Commanders set competency-based training goals using the dashboard as an aid.
8. STEEL-R provides a list of scenarios that could be used to achieve these training goals.

Competency estimates are made for each of three *developmental stages* - crawl, walk, and run - that roughly correspond to purely synthetic, mixed reality, and live training. Competency levels in STEEL-R are Novice, Practiced, Proficient and Expert. Each trainee, which could be an individual Soldier or a team, squad, platoon, etc., can hold each skill or competency in a framework at each level within each developmental stage.

### Introduction Conditions

Guided experiential learning requires focused practice under a variety of conditions. STEEL-R views conditions as modifiers on performance and on competencies and skills. It evaluates performance of tasks under different conditions, notes conditions in assertions, and estimates skill and competency levels separately under each condition. When reporting data via xAPI, physical and environmental conditions can also be translated into measures of stress and difficulty. Thus, a Soldier might be estimated to be Expert at a given skill under daylight conditions but Novice at that same skill under nighttime conditions, and to be considered Proficient overall, STEEL-R may require that a Soldier be Proficient under multiple relevant conditions. The exact way in which this is done is explained in the next section.

## INCORPORATING CONDITIONS INTO STEEL-R COMPONENTS AND COMPUTATIONS

---

As is described in more detail in the STEEL-R data strategy, STEEL-R involves a series of components and computations in addition to the systems and environments in which training takes place. This section provides details of how conditions are incorporated into the current version of STEEL-R.

### xAPI Profiles

In STEEL-R, Soldiers engage in pre-configured training scenarios in which they must accomplish a set of tasks and exhibit a set of behaviors. Granular event-based activity from these is captured and evaluated by the Generalized Intelligent Framework for Tutoring (GIFT) (US Army Research Laboratory, 2022). Using formulas pre-programmed into its Domain Knowledge File (DKF) (GIFT, 2021; US Army Research Laboratory, 2022), GIFT evaluates whether performance was above, below, or at expectations. These evaluations, and data about the conditions under which the tasks were performed and behaviors measured, is emitted in the form of xAPI statements (IEEEb, 2023) to a “noisy” Learning Record Store (LRS) (ADL, 2021; Yet Analytics, 2021). These xAPI statements conform to GIFT’s xAPI Profile and are then processed through an LRS pipeline and stored in a second, “transactional” LRS in a manner that is governed by a second xAPI profile (the STEEL-R profile). The STEEL-R xAPI Profile governs what data are available in the transactional LRS and the form in which they are expressed. The result is standardized xAPI statements that express what tasks were performed and behaviors were exhibited in the training systems, how successfully these were completed or demonstrated, what events were associated with them and used to

evaluate them, when these events occurred, the conditions under which these events occurred, and the results of GIFT's evaluation. The xAPI statements in the transactional LRS report contextual parameters (such as illumination and terrain), stressors and factors that raise difficulty (such as the presence of civilians), and a measure of stress and difficulty on a scale of 1 to 10.

## The CaSS Decoder

xAPI statements in STEEL-R are consumed by the Competency and Skills System (CaSS), which continually polls the transactional LRS. They are then translated into *assertions* about one or more competencies or skills in a framework. This translation takes place in a *decoder*. The decoder uses a lookup table to associate each task or behavior in a scenario with competencies and skills. This association can be one-to-many. For each competency or skill associated with a task or behavior, and for each possible performance outcome (e.g., at, below, or above expectations), the lookup table also assigns a number between -1 and 1 that represents how strong of an indicator that outcome is for each associated competency or skill, with positive numbers representing evidence that the competency or skill has been demonstrated and negative numbers representing evidence that the competency or skill is lacking. Conditions reported by xAPI statements are converted into conditions used for competency estimation, described below, and both the strength of the evidence and the conditions are incorporated into an assertion that says that a competency or skill has or has not been demonstrated under given conditions.

As an example, if the skill is determining the direction of travel from the position of the sun, and the task is to navigate without a compass or GPS, successful performance on the task might be considered to be strong evidence of the skill, so performance above expectations might be assigned a value close to 1, performance at expectations may be assigned a slightly smaller value, and performance below expectations might be assigned a moderate negative value. On the other hand if the task measured in a scenario is ordering a squad to move in the correct direction during a "react to contact" scenario, performing at or above expectations may be considered as very weak (or no) evidence of the squad leader's land navigation skills, so the numbers assigned would be small and performance below expectations would not be considered negative evidence at all. In both cases, the lighting and terrain conditions would be reported, so that in the model estimates competencies and skills, performance under the condition of darkness would be ignored, while the model would require that the skill be demonstrated under different terrain conditions before concluding that the skill had been mastered.

## Competency and Skill Estimation

STEEL-R uses assertions generated by the decoder to estimate the state of each competency and skill for each Soldier and echelon. Currently, the model used by STEEL-R examines all assertions made about each competency and skill in a framework and computes two scores: an *evidence* score and a *practice* score. The evidence score accumulates the strengths of the relevant assertions, weighted by a Pareto skill-decay function (Averell & Heathcote, 2011), to produce the equivalent of a test score that takes longitudinal performance into account. The practice score counts how often a competency or skill has been practiced, taking spacing (Robson et al., 2022) as well as decay into account, so that rapidly repeated practice and practice after too long of a pause does not contribute much.

The model used by STEEL-R compares these scores to pre-configured thresholds to determine whether a Soldier or team of Soldiers has achieved a state of Novice, Practiced, Proficient, or Expert. The model also considers the hierarchical structure competencies and skills as reflected in frameworks and can place requirements on sub-skills. These requirements are called *rollup rules*. Thus, to be considered Proficient or Expert in marksmanship, a rollup rule may require that a Soldier have achieved a state of Proficient on breathing control. The full model supports rollup rules that also involve the evidence and practice scores

on sub-skills, but in the actual implementation, rollup rules are simplified by (a) not using this latter type of rule and (b) requiring that an equal or greater state be achieved on all sub-skills of a competency or skill before that state can be achieved on the competency or skill in question.

In the first version of STEEL-R, competency and skill states were estimated using evidence and practice scores that ignored developmental phases (crawl, walk, run) and conditions. In the current version, evidence and practice scores are computed separately for each developmental phase and each applicable condition. Given a condition and a phase, the evidence and practice scores are computed using only assertions that were generated by events that took place under that condition and in a training environment associated with that phase. This permits conditions to be included in the model but exponentially increases the amount of data that is needed if a competency and skill must be evaluated under many conditions. To address this, the current STEEL-R model uses only binary values for conditions, and the decoder reduces the evidence reported via GIFT and xAPI to discrete measures of stress (low and high), difficulty (low and high), lighting conditions (day and night), and time pressure (low and high). Moreover, competency states are only estimated for each relevant condition under which a competency can be demonstrated or applied. For example, marksmanship is computed under day and night conditions but the ability to use Army terminology is not estimated under different conditions.

## The Navigator and the Experience Index

The “Navigator” in STEEL-R provides training recommendations based on a commander’s requirements (“commander’s intent”). This intent includes goals and objectives for training plus the conditions under which training should take place. Recommendations consist of training scenarios drawn from a database called the *experience index* or XI. The XI stores available scenarios for each available training environment together with metadata that indicates which competencies and skills are practiced and required for each scenario. These metadata are expressed using the IEEE draft standard P2881 (IEEE, 2023a).

The Navigator examines current estimates of a Soldier’s (or echelon’s) competency and skill state under the conditions specified by the commander and looks for training scenarios that can move trainees from their current state to the desired state. In keeping with the simplification of the rollup rules discussed above, the Navigator assumes that if the goal is P on a competency C under a condition c, then P must have already been achieved under condition c on all sub-competencies and sub-skills of C, and similarly for T. If these prerequisites are not present, then the Navigator recommends scenarios that can be used to acquire them. If there are no appropriate scenarios, the Navigator points the user to an *eXperience Design Tool* (XDT) that can be used to design an appropriate scenario.

The Navigator is the key place where pedagogy and conditions come into play. As a software application, the Navigator operates on:

- A pool of available scenarios and related metadata maintained in the XI,
- Competency and skills estimates from CaSS
- Frameworks stored in CaSS, and
- *Commander’s Intent* that specifies:
  1. A trainee (an individual or echelon),
  2. One or more skills or competencies
  3. A developmental stage
  4. A physical or environmental condition
  5. A stress level and a difficulty level

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- The level at which the trainee should hold the skill or competency within the specified developmental stage under the specified conditions (4 and 5).

In response, the Navigator either displays a list of potential scenarios or advises the user to create a new one (Figure 1).

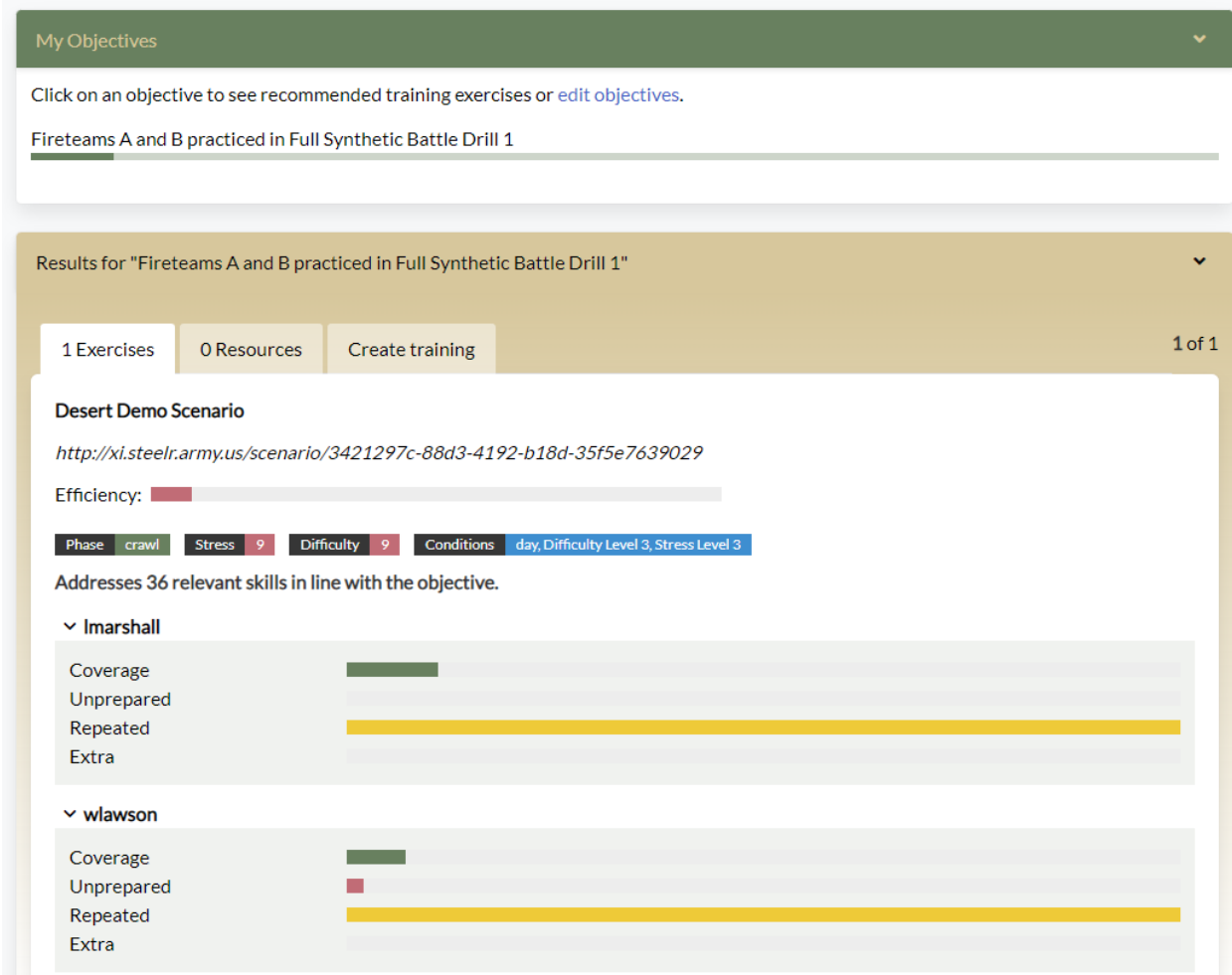


Figure 1. STEEL-R Navigator User Interface

## The eXperience Design Tool (XDT)

The STEEL-R XDT (Figure 2) defines contextual experiences that instantiate experiential learning in a synthetic learning environment. The content of these experiences is based on either actual recorded live experiences whose effectiveness has been demonstrated or experiences that are directly applicable to an anticipated real-world occupational environment. XDT produces an abstracted machine-readable Experiential Training Support Package (XTSP) that is based on existing paper-based training support packages. XTSPs can be consumed and instantiated in suitable synthetic or live education or training environments. XTSPs specify episodic, static, or dynamic experience events that require targeted teams and/or roles to perform targeted tasks under dynamic conditions required by a competency or performance standard. The XDT is a complex tool whose details are described in Mishra et al. (2023) in this volume.

The XDT is essential to the inclusion of conditions in STEEL-R. It is used to design new experiences that require trainees to practice under specified conditions and includes those conditions in the metadata it passes to the XI when it informs the XI when new experiences and experience events have been created. This metadata enables the Navigator and other applications to search for XTSPs for use in training.

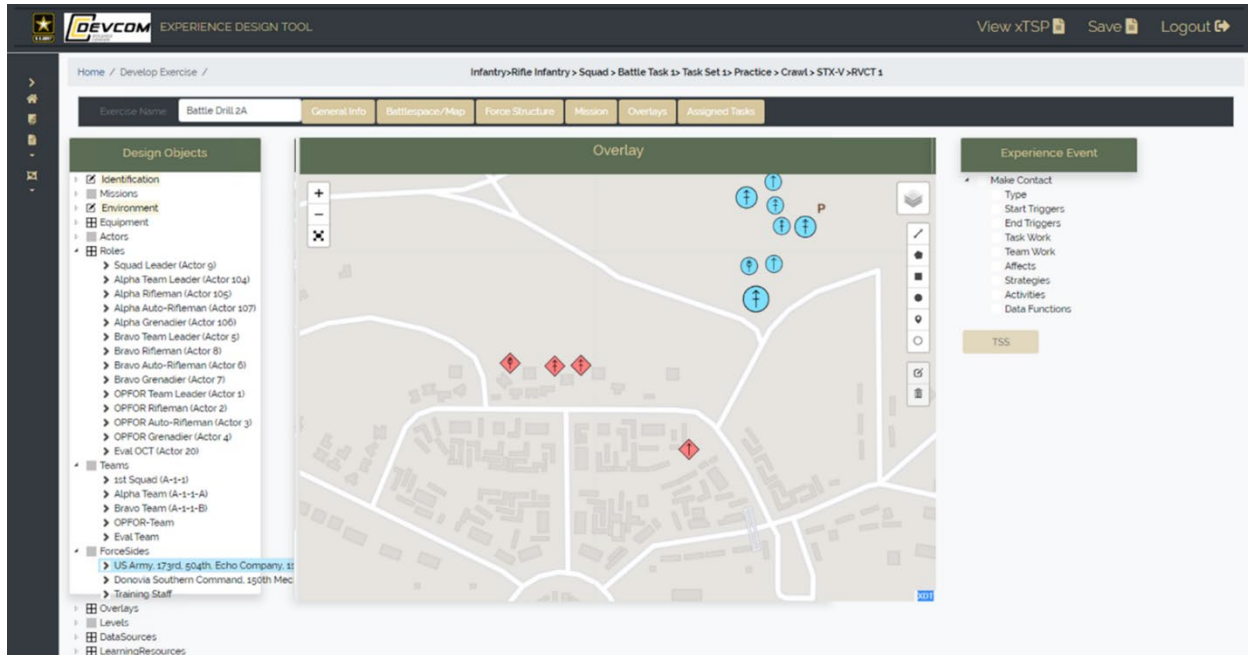


Figure 2. STEEL-R XDT User Interface

## CONCLUSIONS AND FUTURE RESEARCH

This paper has explained how conditions were introduced into STEEL-R in a practical and computational fashion. From a competency modeling perspective, the methods used can be described as splitting each competency and skill that can be performed and evaluated under different conditions into a separate set of competencies and skills, one for each condition, but without changing the sub-skills involved. In practice, skills are more nuanced. A skill such as marksmanship may require different tools and subskills under different conditions (e.g., nighttime and daytime), which is not fully reflected in this approach. Similarly, each meaningful *level* of competency or skill (e.g., Practiced, Proficient, Expert) may require different sub-skills and different levels on each. This is also not fully reflected.

In addition, there are multiple ways to rank alternatives when the Navigator recommends scenarios. In some cases, the objective might be to quickly train a given set of competencies and skills, accepting the outcome that learners without the proper prerequisites may fail and be eliminated from the mission. In other cases, the objective might be to find the fastest path to being rated as proficient in one particular skill, and in other cases the goal may be to achieve readiness across all competencies and skills in a framework. Each of these cases will lead to a different objective function against which scenarios are ranked.

Another consideration is spaced repetition. To boost the practice score, learners should engage in targeted practice at optimized intervals. If the Navigator recommends training scenarios in which Soldiers repeatedly practice the same skills under the same conditions, the effect on their scores may be diminished, and they may not achieve the thresholds required to achieve a proficient level. Currently, the Navigator does not take this into account.



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

The above considerations point out the need to integrate learning theory and strategies into a ranking mechanism, and to consider pathways instead of single scenarios in the Navigator. This is a fertile area of research that should be pursued next.

Finally, STEEL-R is designed to allow thresholds to be determined by machine learning algorithms and to incorporate Artificial Intelligence (AI) into the Navigator. This has not yet been done and is among the next steps that will be taken.

## REFERENCES

---

- ADL. (2021). *Experience API (xAPI) standard*. ADL Initiative. <https://adlnet.gov/projects/xapi/>
- Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of mathematical psychology*, 55(1), 25-35.
- Clark, R. E. (2005). *Guided experiential learning: Training design and evaluation*. Center for Cognitive Technology, Rossier School of Education, University of Southern California.
- GIFT. (2021). *Domain knowledge file 2021-2*. GIFT Tutoring Portal. [https://www.gifttutoring.org/projects/gift/wiki/Domain\\_Knowledge\\_File\\_2021-2](https://www.gifttutoring.org/projects/gift/wiki/Domain_Knowledge_File_2021-2)
- Goldberg, B., Owens, K., Gupton, K., Hellman, K., Fort Leavenworth, K. S., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). *Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy*. Interservice/Industry Training, Simulation, and Education Conference (IITSEC), Orlando, FL. <https://www.gifttutoring.org/attachments/download/4295/21332.pdf>
- Hernandez, M., Blake-Plock, S., Owens, K., Goldberg, B., Robson, R., Welch, T., & Ray, F. (2022). *Enhancing the Total Learning Architecture for Experiential Learning*. IITSEC, Orlando, FL. [https://www.gifttutoring.org/attachments/download/4745/2022\\_iiitsec\\_22461\\_steelr\\_Enhancing\\_TLA.pdf](https://www.gifttutoring.org/attachments/download/4745/2022_iiitsec_22461_steelr_Enhancing_TLA.pdf)
- IEEE. (2023a). *Standard for Learning Metadata (DRAFT)* (No. P2881). <https://standards.ieee.org/ieee/2881/10248/>
- IEEE. (2023b). *Standard for Learning Technology - JavaScript Object Notation (JSON) Data Model Format and Representational State Transfer (RESTful) Web Service for Learner Experience Data Tracking and Access* (9274.1.1-2023).
- Kolb, D. A. (1984). *The process of experiential learning* (Vol. 1). Prentice-Hall.
- Mishra, S., Owens, K., & Goldberg, B. (2023). The Evolution of an Experience Design Tool: From Live Exercises to Synthetic Experiential Learning with the Experience Design Tool and Integration with GIFT. *Proceedings of the Eleventh Annual GIFT Users Symposium*. GIFTSYM11, Orlando, FL.
- Owens, K., Goldberg, B., Ray, F., Mishra, S., & Blake-Plock, S. (2022). *Demonstration of a Competency-Based Experiential Learning Model for the US Army Synthetic Training Environment (STE)*. iFEST 2022, Alexandria, VA.
- Robson, R., Ray, F., Hernandez, M., Blake-Plock, S., Casey, Cliff, Hoyt, W., Owens, K., Hoffman, M., & Goldberg, B. (2022). *Mining Artificially Generated Data to Estimate Competency*. Educational Data Mining, Durham, UK.
- Rose, D. E. (2012). Context-based learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. (pp. 799–802). Springer. [https://doi.org/10.1007/978-1-4419-1428-6\\_1872](https://doi.org/10.1007/978-1-4419-1428-6_1872)
- US Army Research Laboratory. (2022). *GIFT Portal*. <https://www.gifttutoring.org/projects/gift/wiki/Overview>
- Yet Analytics. (2021). *SQL LRS*. SQL LRS. <https://www.sqllrs.com/>

## ABOUT THE AUTHORS

---

**Dr. Robby Robson** is a researcher, entrepreneur, and standards professional who serves as the Principal Investigator on the STEEL-R project. After receiving his doctorate from Stanford, Robby spent fifteen years as a research mathematician in the areas of semi-algebraic geometry and computational number theory. He became interested in Learning Technology and technical standards after co-creating one of the first online learning systems and online Calculus courses in 1996. Following a brief stint in the Silicon Valley, Robby co-founded Eduworks Corporation, became its CEO in 2001, and retired as CEO in 2022. As its chief science officer, he is currently developing the next generation of competency management, talent analytics, and skills-based talent pipeline solutions.

**Fritz Ray** is the Chief Technology Officer of Eduworks Corporation with twenty years of experience developing proprietary and open-source software with applications to education, training, aviation, finance, and law. He led development of the Competency and Skills System (CaSS), is the primary designer and developer of the STEEL-R Navigator and oversaw the technical implementation of the STEEL-R data strategy.

**Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.

**Shelly Blake-Plock** is President and CEO at Yet Analytics. He has served as principal investigator on projects related to xAPI and the Total Learning Architecture, including the ADL-sponsored Data and Training Analytics Simulated Input Modeler (DATASIM). Active in the global standards community, he is an officer of the IEEE Learning Technology Standards Committee and chairs the P9274.4.2 working group on xAPI Cybersecurity as well as the LTSC Technical Advisory Group on xAPI.

**Cliff Casey** is CTO at Yet Analytics. He is the principal designer and developer of the STEEL-R eXperience Index and, together with Shelly Blake-Plock, is responsible for the STEEL-R xAPI strategy and xAPI profiles.

**Kevin Owens** is an Engineering Scientist at the Applied Research Laboratories: The University of Texas at Austin. After retiring from the US Navy, accumulating years of experience as an instructor, curriculum designer, and operational trainer, Kevin earned a BS in Workforce Education and Development from Southern Illinois University and a MS in Instructional Design, Development, and Evaluation from Syracuse University. Kevin has since spent over 20 years in industry and academia conducting DoD-related Learning Engineering projects. He is currently working on Science and Technology projects for the US Army's Synthetic Training Environment and researching a Competency-Based Experiential Learning model.

**Dr. Sazzad Nasir** is a senior machine learning engineer at Eduworks Corporation who has expertise in machine learning, predictive analytics, natural language processing, and computational modeling. Before joining Eduworks, he was a faculty member at Northwestern University, where he taught and led research in cognitive and neuroscience, and managed and supervised doctoral research projects. Dr. Nasir has a PhD in theoretical physics from the University of Cambridge in the UK and published in the fields of mathematical physics and neuroscience prior to joining the Eduworks team.

# The Evolution of an Experience Design Tool: From Live Exercises to Synthetic Experiential Learning with the Experience Design Tool and Integration with GIFT

Sanjay Mishra<sup>1</sup>, Kevin Owens<sup>2</sup>, and Benjamin Goldberg<sup>3</sup>

<sup>1</sup>Oak Ridge Institute for Science and Education; <sup>2</sup>Applied Research Laboratories, The University of Texas at Austin;

<sup>3</sup>US Army Combat Capabilities Development Command (DEVCOM) – Soldier Center

## INTRODUCTION

---

This paper discusses the transformation of an Exercise Design Tool (EDT), originally designed for generating plans for live small-unit exercises, to the Experience Design Tool (XDT), which imports existing training missions (scenarios) used in real operational or live training environments, and reproduce them for use in synthetic environments for experiential learning. The XDT tool also interfaces with the Generalized Intelligent Framework for Tutoring (GIFT) (Sottolare et al., 2012; Sottolare et al., 2017) as a real-time assessment and scenario modification tool, and then outputs to an Experience Application Programming Interface (xAPI) data store, which is used for long-term storage and ultimately as evidence for competence assertions. XDT is also capable of importing training “recommendations” generated by a training “navigator” function built into a competency-management system. The purpose of XDT is to capture and generate a common curriculum that shares real “experiences,” in a standard JavaScript Object Notional (JSON) format, referred to as an eXperiential Training Support Package (XTSP). An XTSP translates any standardized mission scenario information, and occupational task performance measurement functions, into a standard format that a training application could use within the Army’s future Synthetic Training Environment (STE). The paper highlights the capabilities, and lessons learned during the XDT’s development process to ensure XDT’s scalability and usability in a rapidly iterative learning engineering development operation that supported the US Army DEVCOM STE Experiential Learning for Readiness (STEEL-R) research project (Goldberg et al., 2021). Additionally, the paper showcases XDT’s real time integration with other services that were engineered as part of the STEEL-R project. We will also showcase some of XDT’s main capabilities, its product, and discuss its versatility and flexibility.

## XDT CONCEPT, ENGINEERING AND OPERATION

---

### Transition from Live-Training Based EDT to XDT

As noted above, XDT evolved from an EDT product that was engineered to support live exercises as a part of the US Army’s *Squad Overmatch* science and technology project (Johnston et al., 2018). This project developed a variable stress-condition based series of classroom, synthetic, and live “lane-based” events intended to foster teamwork, advanced situational awareness, and combat resilience. An integrated training approach (ITA), including knowledge and skills training for improving advanced situational awareness (ASA), resilience, tactical combat casualty care (TC3), medical skills and decision-making, and team performance was developed to help Army small-unit teams make better decisions. (Milham et al., 2016).

The EDT developed for this project supported trainers in rapidly designed training scenarios with specific measured events that associated with a specific physical training location. It was an online system where trainers could log in, design scenarios, select pre-existing events associated with selected tasks, which associated with selected Mission Essential Task Lists (METL). See Figure 1 below.

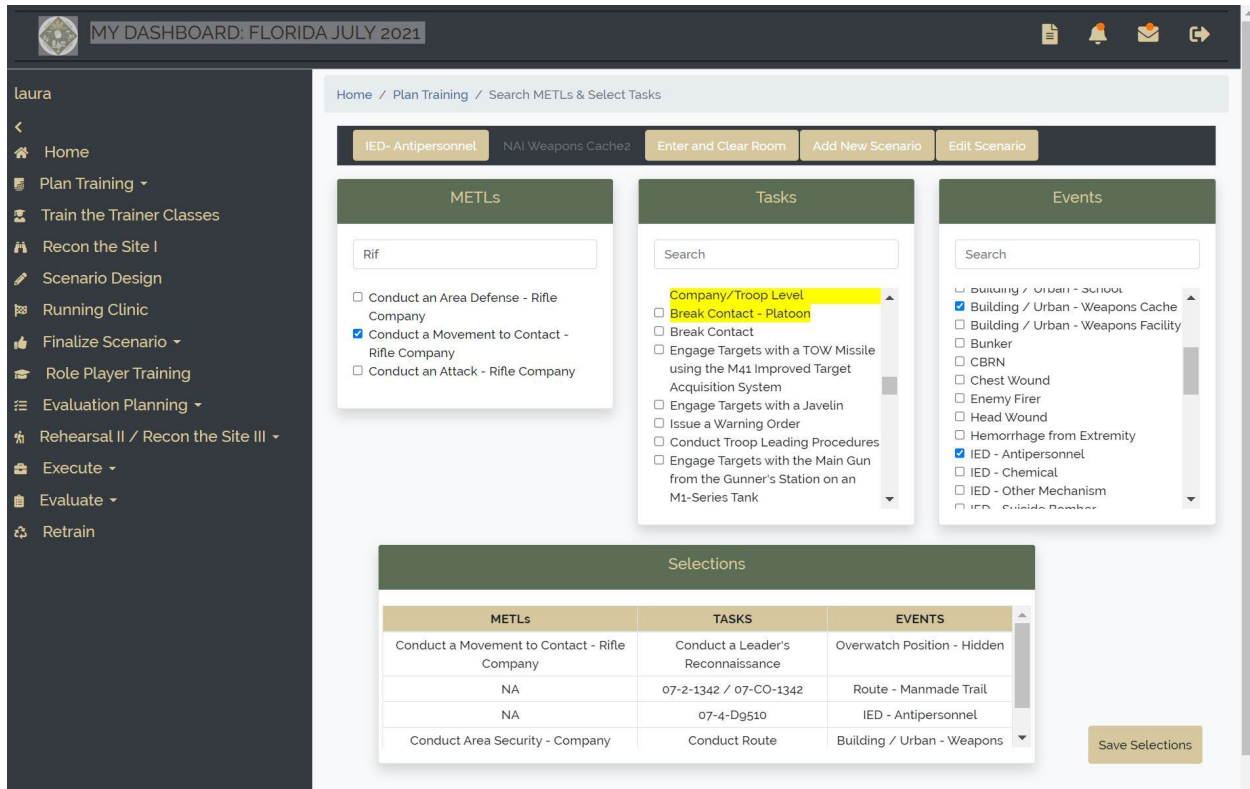


Figure 1. Scenario Design Using the EDT

XDT evolved from the EDT capabilities, which used much of its existing structure but added the ability to dynamically produce what are referred to as *Experience Events* (xEvents). These function similarly to the EDT events but use more syntactic distinct terminology, and most importantly, included the ability to associate standard measures, criteria and data sources to the selected tasks for a pre-selected synthetic or live training environment, mission and levels of difficulty that are associated within a competency framework (aligned with the US Army's current METL based proficiency framework). XDT also provides the capability of producing experiences tailored to targeted learner-actors, for a targeted team or role, within a selected mission-scenario, that can then be directly applied to a STE hosted synthetic training application. At the same time, XDT will configure a real-time assessment and training tools incorporated in services like provided by GIFT.

XDT was also designed to create multiple novel mission-oriented conditional variables that made the experiential learning distinct for better learning encoding. Variables may include terrain parameters, environmental parameters, and specific task targets or opposing force capability-parameters. This capability was also based on EDT capability, with the intention of modifying the difficulty and stress to match a learner's competence and ability. Each experience's xEvent incorporates one or many task performance prompts known as "triggers" that activate other activities and data streams also specified in the XTSP. Triggers may happen in a pre-defined series or can be triggered based on the actors' activities or can be force-activated manually from GIFT. xEvents could be conceived as competency-based "pop test-questions," designed to measure a target learner-actor's ability to perform a specific task, related to a given occupational team, role, and within specified conditions of difficulty (e.g., volatility, uncertainty, complexity, or ambiguity) and stress (cognitive load, physical load, environmental load, etc.). By tracking these xEvents and their outcomes in GIFT, the premise is we can more accurately track and determine a learner-actor's state of competence over time with the raw data they produce from each xEvent.

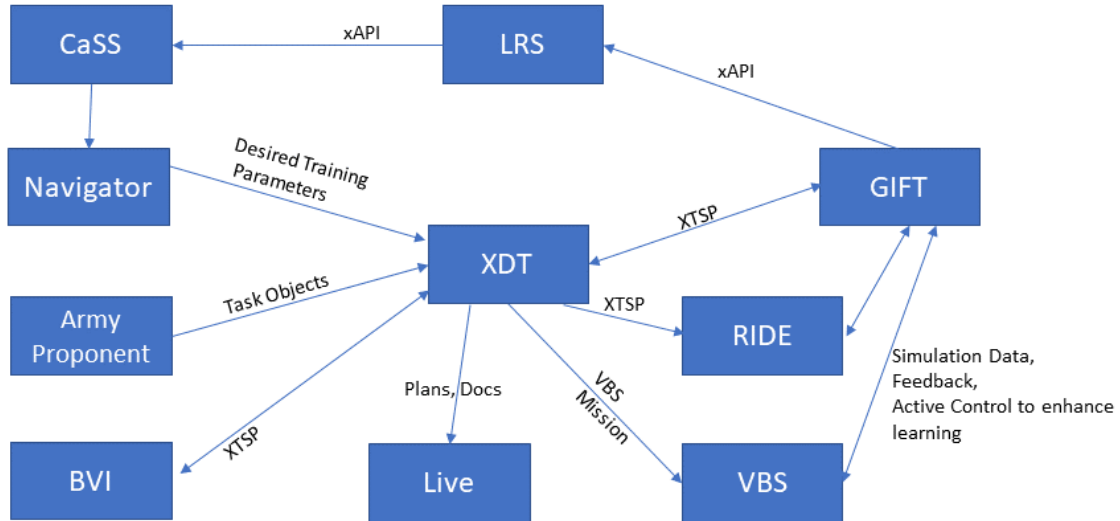
XDT is also designed to simplify the trainer, teacher, or instructor's job in setting up a synthetic training environment. This happens because an XDT not only automates the setup of a selected training environment but configures GIFT to collect and generate objective performance data produced by sensors focused on the learner-actor or the training environment itself. XDT also assists designers in not only testing if the appropriate transactional data is produced correctly but also supports evaluating if the target competency-task is being optimally measured, and subsequently defining appropriately defined criteria for each resulting performance level within an xEvent's conditions.

XDT (utilizing GIFT capabilities) also defines strategies that provide actor interventions (e.g., feedback) or exercise adaptations (e.g., changes in environment or artificial intelligent entity behaviors) to further develop the learning actor's experience and task competence. A premise of STEEL-R is that long-term exposure to these dynamic conditions and interactions, over numerous experiential opportunities, enables learner-actors to rapidly develop tacit-knowledge, which enhances Soldier's ability to automatically respond (or anticipate) similar violent, volatile, uncertain, complex, or ambiguous combat events while performing in a future live training or real-world combat environment. Based on the theory of experiential learning (Kolb, 1984), the resulting experiences, and subsequent data-informed performance feedback, stimulates reflection by the actor (or their leaders), which creates long-term experiential learning and expertise beyond what can be achieved in a didactic classroom learning environment. This process also allows learner-actors to develop ideas for constructing new performance strategies, to improve their previous evidence-based performance outcomes in future xEvents related to the same tasks and competencies. STEEL-R refers to this learning model as Competency-Based Experiential Learning (Owens & Goldberg, 2022).

## Experience Design Tool

The XDT is projected to be used during the US Army's training "prepare phase", before a designed XTSP is executed. During this phase the XTSP is produced in the XDT and then iteratively tested and improved using the selected STE training application and its Training Management Tools (TMT) (e.g., GIFT). At the end of the prepare phase, the actual learner-actors (teams or individuals) arrive at the training site and are assigned to sides, teams and respective roles based on unit organization or other objectives. Each learner-actor is also allocated a specific synthetic training and/or sensor device(s). This allows GIFT to track, assess and report data per mapping laid out in the Domain Knowledge File (DKF), between the learner-actor and their assigned training device and/or sensor. This also allows GIFT to track all activities of a learner-actor's synthetic entity within a simulation or other sources of data produced by the training application associated with a specific learner-actor. This preparation phase also enables the designated training leader the opportunity to modify or add extra parameters to the XTSP based on the specific competency attributes of a given learner-actor's profile. Once this has been done the training can commence into the Army's training *execution phase*.

XDT was engineered to play a central role in the STEEL-R architecture. Figure 2 shows the interfacing and usage of XDT in the overall STEEL-R training architecture.



**Figure 2. XDT's Role in the STEEL-R Architecture**

Specific activities that transpire between XDT and other STEEL-R services:

1. The request for a new training scenario or modification of an existing training scenario comes to the XDT from the Navigator to focus development of a XTSP on the unit, team or individual training goals and needs. The XDT user can always over-ride the inputs provided by the navigator as needed but this automated input helps in making the design process easier and with less steps.
2. The XDT user then selects or creates the training events and their triggers, targeted actors, tasks, and strategies using the web-based tools provided by the XDT.
3. The resulting training is saved as an XTSP and can then be shared with other applications (e.g., BVI (Battlespace Visualization and Interaction)) and/or GIFT to refine its design. This will be an iterative process a user can do between various tabs in a commercial browser since each application is presumed to be coming from a common web-server in the STE-TMT.
4. This XTSP is then tested using the training application and devices it defines as needed for measuring the selected tasks in the xEvents. This is another iterative process that continues until a satisfactory XTSP design is completed.

This design process of a XTSP can be used to configure any synthetic environment so long as they are able to parse and translate the XTSP elements it delineates.

We will now describe the specific operations a user of XDT can perform



### *XDT Capabilities for XTSP Creation and Edits*

The XDT first imports information from the STEEL-R Navigator service and/or allows the user to manually define the context and requirements of a new experience with its *Experience Context* form, shown in Figure 3.

The screenshot shows the 'Experience Context' form in the XDT application. The form is titled 'Experience Context' and contains various dropdown menus and checkboxes for defining the context of an XTSP. The form is divided into several sections with colored backgrounds: green for Occupation and Organization details, light blue for Unit Mission and Tasks, light orange for Competence and Training details, and light red for Classification and Distribution. The left sidebar shows the user 'SSG Meers' and navigation options like 'Home', 'Design New Exercise', 'Edit Existing Exercise', and 'Import Data'. The top right has 'View xTSP', 'Save', and 'Logout' buttons.

**Figure 3. Defining the XTSP Experience Context**

The information from this form is used to query and filter authoritative data needed to fill-out the XTSP model shown along the left side of the XDT. This data is also used to select the tasks that will be available to include within an xEvent. The designer must first create a new (or use an existing) xEvent as shown in Figure 4 below. From there the specific tasks that will be prompted and measured with data in the selected xEvent (now shown on the right side of the XDT) is/are selected as shown in Figure 5 below. Once an xEvent is created, and the tasks are defined, the user can define the experiential sequence of stimuli using the XDT 2D map tool shown in Figure 6 below.

The 2D map tool allows for easy placement of not only targeted mission/organizational sides, teams, and roles, but also overlays and triggers, as well as graphic based criteria for automated assessment. The map loaded will be based on the mission associated map and environment that is delineated in the selected real mission operational order. The user can also place any additional map-objects within the xEvent terrain as a means of modifying the conditions and difficulty or stress of the xEvent.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

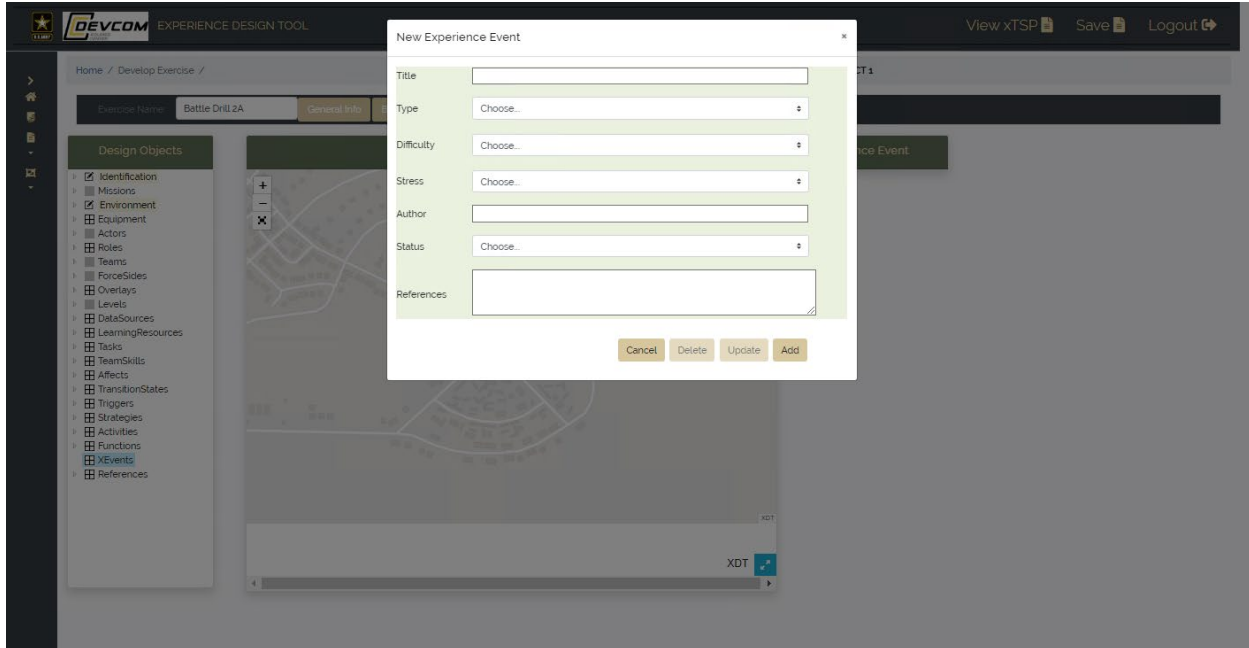


Figure 4. Creating a New xEvent

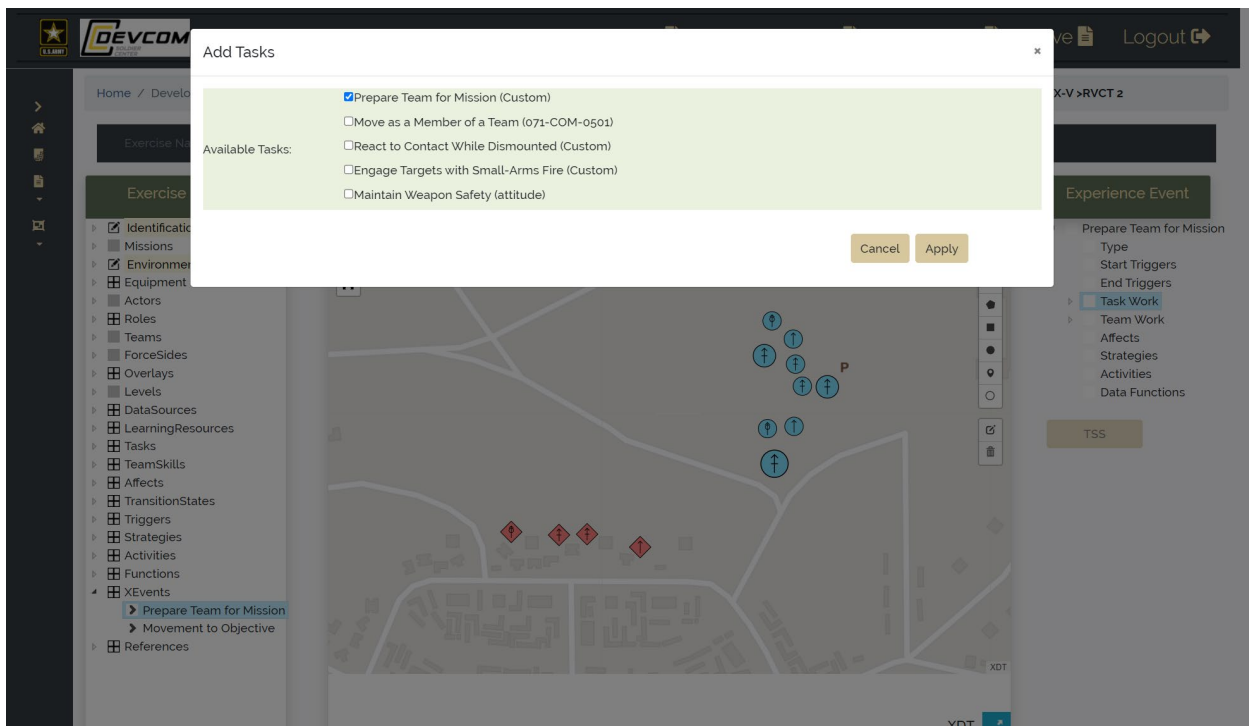


Figure 5. Selecting or Deselecting xEvent Tasks – Note Selectable xEvent Parameters on Right Side of XDT



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

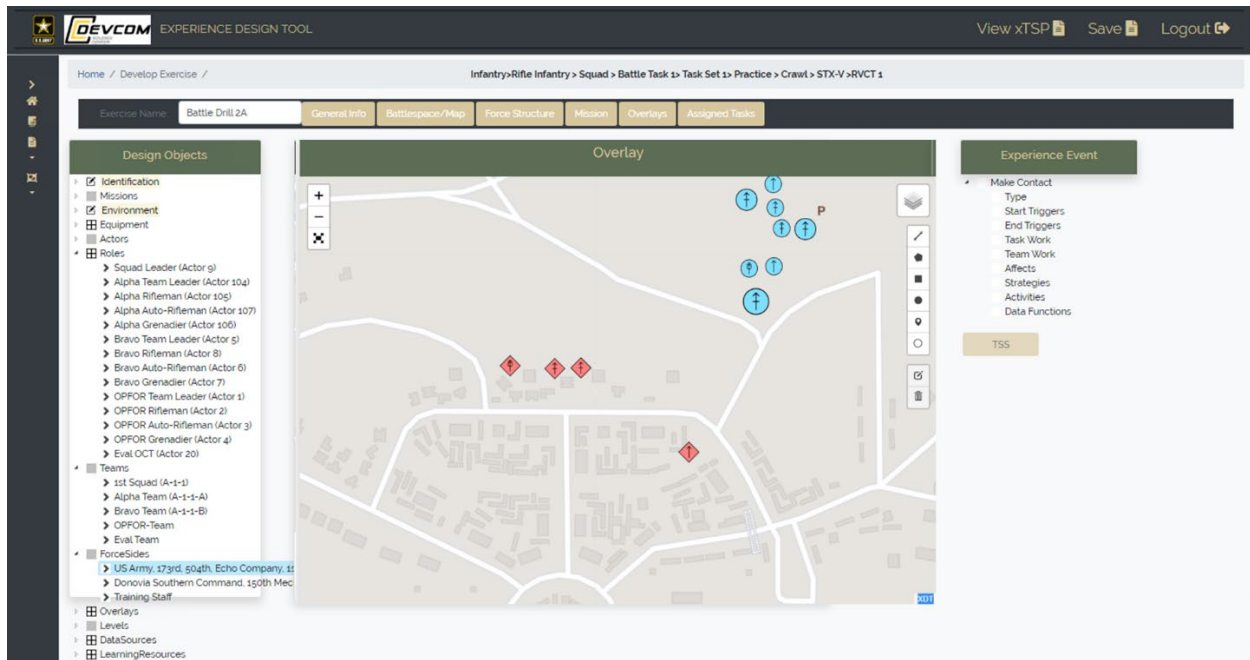


Figure 6. XDT 2D Map with Drag and Drop Capabilities

Once mission elements are dragged and dropped onto the map, positioned, and their properties edited by using associated pop-up menus, they can also be viewed in their 3D representation using an experimental Unity WebGL based 3D viewer as shown in Figure 7 below.

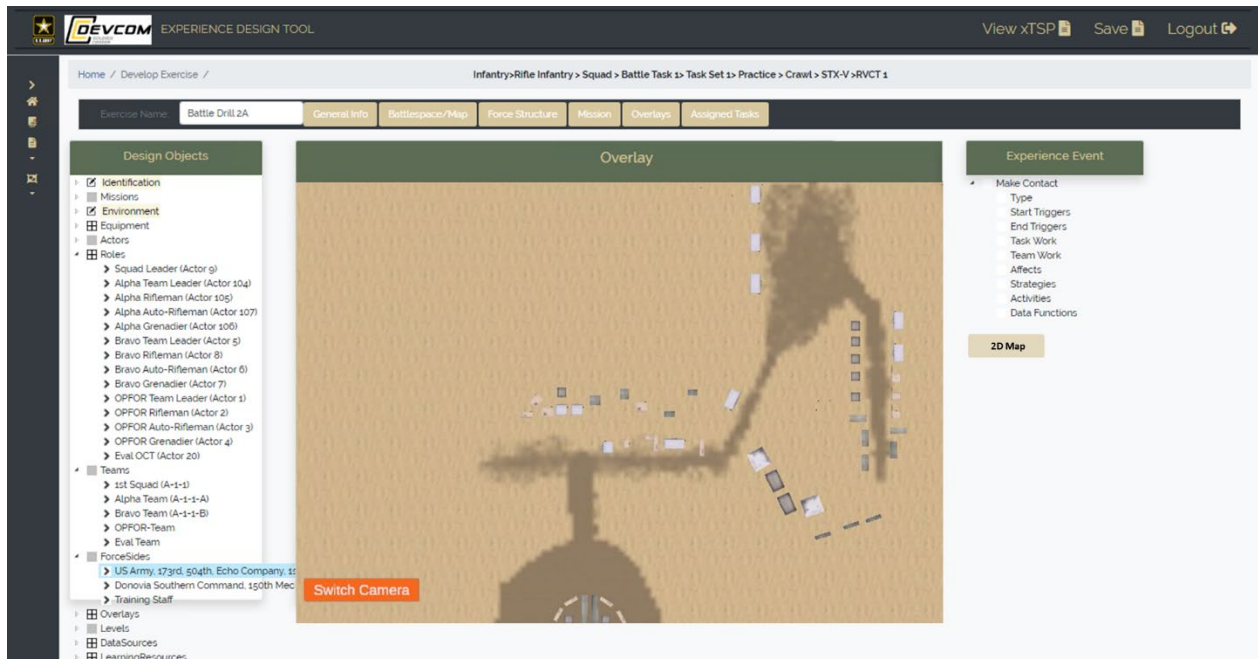


Figure 7. XDT 3D Map

The resulting XTSP experience, and its various xEvents are then saved as an XTSP JSON format that is stored in a central XTSP repository. A link to the XTSP and its respective xEvents is also listed in an

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

Experience Index, that is used by the Navigator service to help training planners determine what experiences are available for given competencies, units, missions, and other training goals or needs. As noted earlier, the xEvent can also be viewed in other external applications for similar viewing capabilities. We will now discuss the two different training environments we used to test the ability of XDT to interface with different training applications that may be used in the STE.

### ***Rapid Integration & Development Environment (RIDE)***

RIDE is a product developed by the University of Southern California Institute of Creative Technologies for the US Army to support synthetic training research. The version of RIDE we received provided a basic Unity-based rendering engine with functional support to produce scenarios; however, it required the designer to do so using the Unity native C# programming language.

Our first task was to make the base-version of RIDE easier to use by a Soldier using XDT, to complete the design process we were researching, and to interface it with GIFT. The challenge was that the baseline RIDE product depended on a user having to directly program the game engine to create or change any scenario. It also did not have any way to communicate with GIFT in a standardized way. We also wanted to use RIDE in a web-environment so that it could be used remotely via the WebGL plug-in.

First we made RIDE capable of being configurable over the internet or an intranet using JavaScript. JavaScript is the language of web pages, is widely used, and is likely to be developed and supported over the foreseeable future. It is also a standard and multiple implementations exist in different languages. We used a C# implementation of JavaScript and integrated it into RIDE.

Next we made it so the user can interact with RIDE to test experience events and other activities using a command prompt or by connecting to it over a TCP socket connection. Upon connection, the user is given a JavaScript prompt at which they can type in commands and query the status of various objects in the RIDE simulation. The same commands and interfaces are available to any other external application that has to interact with the RIDE simulation engine.

Finally, to facilitate the fast interchange of binary data with external applications - for example with GIFT - we also built a Google protobuf formatted interface. This protobuf interface can carry commands to and from external applications along with embedded information in a binary format. We used this to test another unique design we engineered which was to use RIDE to conduct much of the measuring of a learner-actor's activities in the synthetic environment – as opposed to GIFT having to query for that data. This new interface specification enabled transferring data in a binary format that was more efficient and faster as compared to transferring the same data in a textual format as is typically done between web applications.

With these changes, a scenario can now be developed with XDT without requiring the aid of a skilled programmer. The control, measurement and assessment interfaces to GIFT were also built-in and available from the XTSP. The overall result was RIDE becoming a more malleable research tool.

### ***Virtual Battlespace 3***

Virtual Battlespace 3 (VBS3) is a product produced by Bohemia Interactive – the same company that produces the commercial Arma series of first-person shooter games. VBS is used for the Army's Games for Training program that is used to provide simulated training in various Army tasks, such as gunnery, mission command, and other collective-level training. VBS3 is also used at various institutional training commands (e.g., US Military Academy) on military studies.

One of our late challenges in STEEL-R was to determine if we could import the VBS3 SQF compressed mission files (which essentially setup a VBS3 based mission) into XDT, and translate them into an XTSP. From there we can add xEvents to the VBS3 mission and not only save the VBS3 mission as an XTSP but translate it back into a VBS3 SQF and SQM file format.

We were able to engineer a solution to accomplish this challenge that we found also capable of being downloaded to a local server or client machine with VBS installed, and ran on the local VBS installations as shown in Figure 8 below.

The data from the execution of these VBS missions can still be tracked by GIFT using its existing capability to track the VBS3 Distributed Interactive Simulation (DIS) packets. Furthermore, we developed a way for a cloud based version of GIFT to capture and display the outcomes of this VBS experience. This capability will be expanded upon in future research.

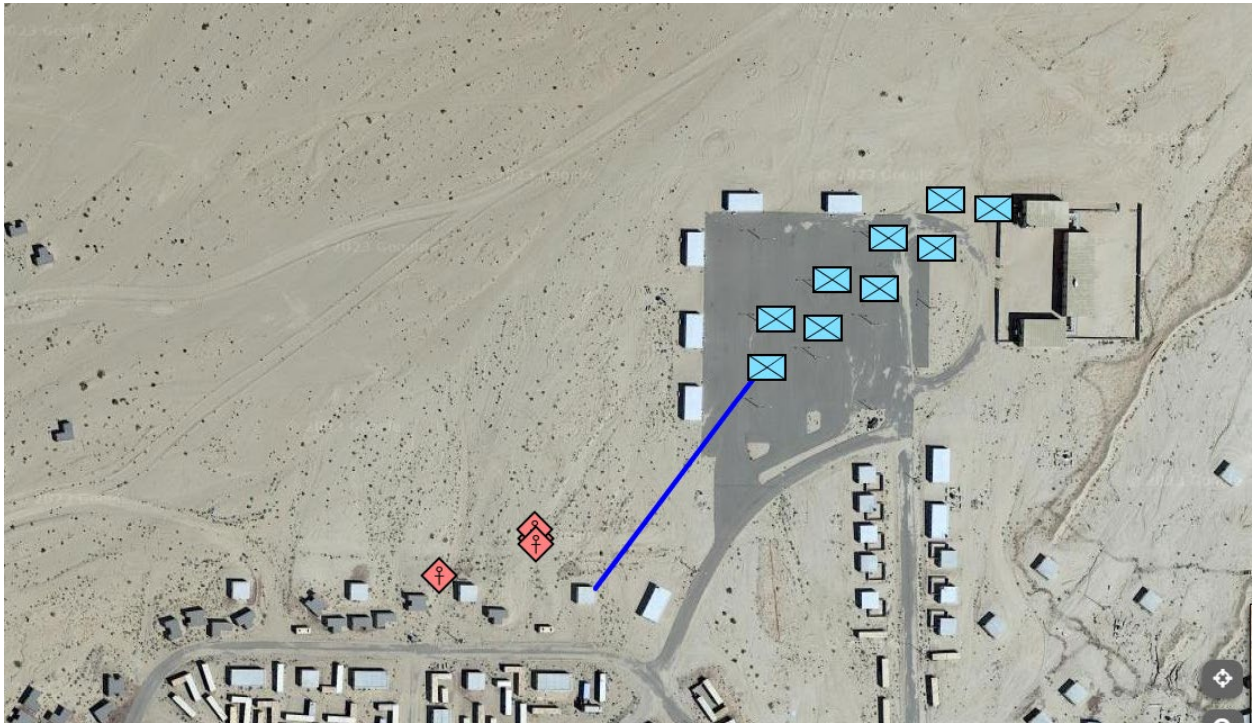
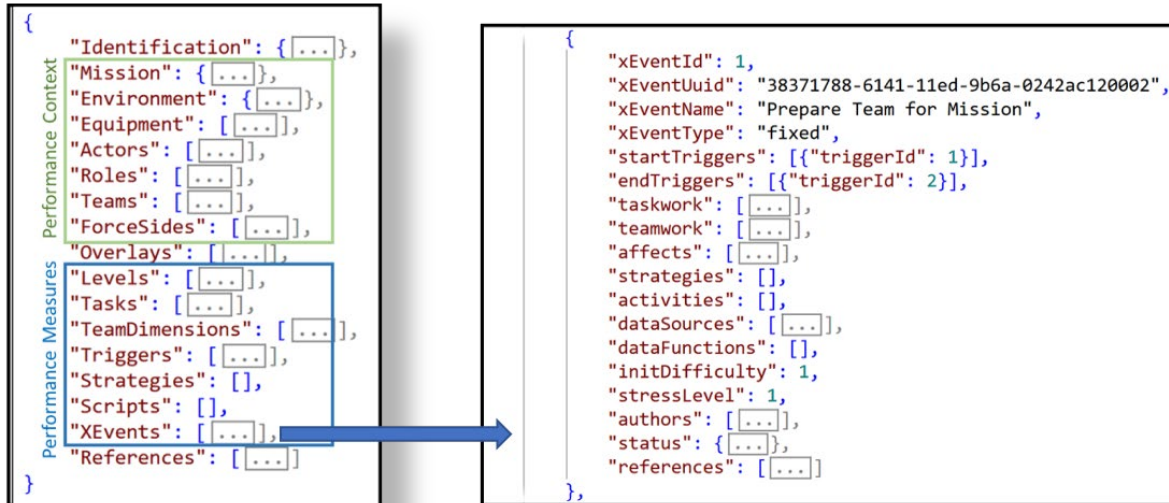


Figure 8. An XTSP Experience Running in local VBS and being visualized on GIFT in the STEEL-R cloud

### Experiential Training Support Package (XTSP)

A key decision made early in the engineering of the XDT, that significantly enhanced its applicability and versatility, was the creation of the standardized XTSP. The XTSP captures the requirements of the existing Army Warfighter Training Support Package (US Dept of Army, 2017). An example of an XTSP is shown on left side of Figure 9 below. In addition to incorporating the fundamental training elements produced by an EDT, the XTSP also delineates the xEvents (discussed earlier) that are triggered manually or automatically in a training experience. An example of an xEvent is shown on right side of Figure 9.



**Figure 9. Experiential Training Support Package and Experience Event JSON Elements**

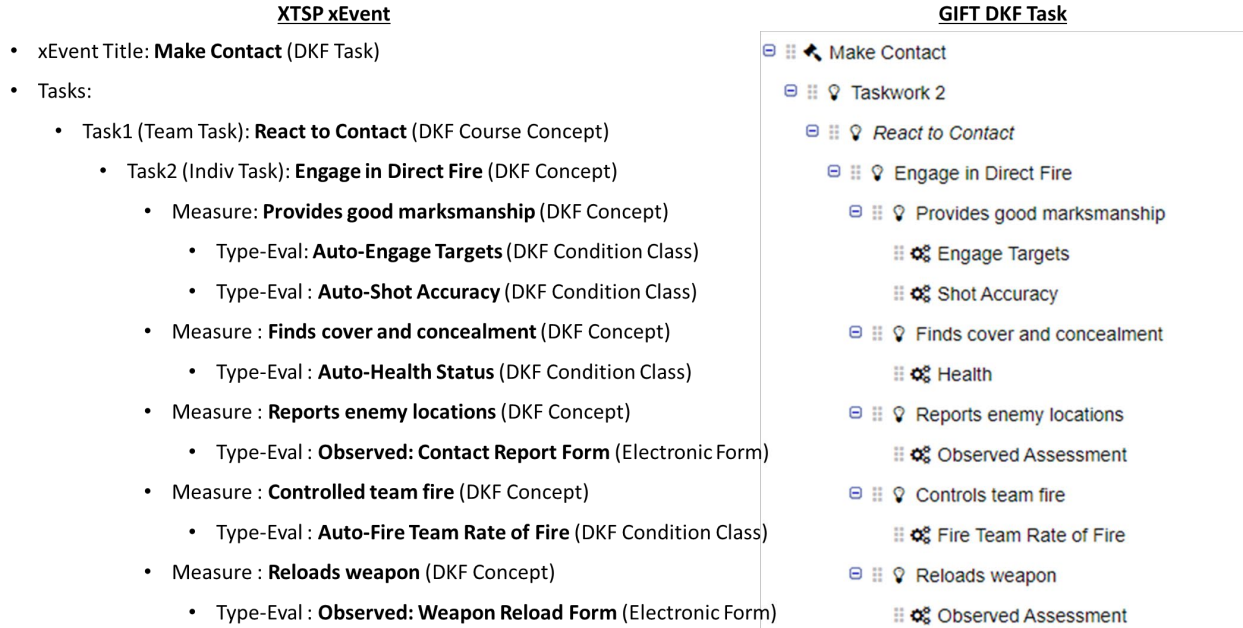
By formatting XTSP in an open JSON format, it enables any application to create, read and modify it. Different applications have been developed and refined over the years that are very good at the tasks for which they were designed for; however, now they can take in a US Army standardized XTSP, use their capabilities to enhance the training events and then save the output back within an XTSP format. This XTSP can then be used to transfer these synthetic experiences or can be further edited to produce an evolved experience.

One system that was used to test and demonstrate the XTSP’s ability to be modified by other systems was the US Army’s Battlefield Visualization and Interaction (BVI) tool. AnXTSP created with the XDT, was then read in to BVI which used its 3D presentation capabilities to evaluate and modify the XTSP contents, and then export the changes back as an XTSP. The resulting XTSP could then be further worked on by the XDT or another application.

As noted earlier, the XTSP also provides GIFT with training experience difficulty and stress points, that GIFT uses to calculate an xEvent’s overall difficulty or stress level reported with the assessment results.

The XTSP was also designed to be imported into GIFT and supply it with many of the elements needed by the GIFT DKF, with regards to actor organizations, roles, tasks, and measurement structure that GIFT currently uses for assessments and strategies for a given xEvent. The common structure and differences between the XTSP syntax and the existing GIFT DKF syntax is shown in Figure 10 below.





**Figure 10. Comparison of XTSP and GIFT DKF Structure / Syntax**

One of the challenges faced with the current DKF structure was its inability to support multiple nested tasks. For example, Figure 10 shows a team task, that requires its own set of measures, and its sub-tasks that may be of a different type, and associated with different roles on the team. In the DKF, team and role assignment are done as part of the condition-class declaration process but while multiple concepts can be rolled-up and collectively evaluated for a summative outcome, multiple tasks cannot, except at the course concept level. Therefore, we engineered a translation that used the DKF “Task” element to represent a distinct xEvent, the course concept or objective to represent the targeted tasks, and then we used DKF concepts to represent both tasks, and measures. Another capability we added was to expand the GIFT DKF “Observed Assessment” condition-class to point to a manual electronic form that a trainer, teacher or evaluator would use to evaluate a specific measure, only using the raw data playback capability STEEL-R provides in its Game-Master interface, so that the measurement would be inspectable and as objectively determined as possible.

Finally, another capability provided within the XTSP is the ability to define a standard profile for reporting the results of an experience, and experience event, its tasks, conditions, measures, and assessment outcomes using xAPI. These “micro-profiles” are embedded within each task object, as they have been pre-defined by a US Army center of excellence, and this is shared with GIFT to support its xEvent and task-based xAPI outputs.

## **XDT KEY CONTRIBUTIONS**

---

The XDT produced several key contributions to both the US Army’s ability to train and assess its personnel so that they can perform more effectively in live exercises or real-combat conditions. These capabilities may also provide capability to new future models for all education and training.

## The Development of the XTSP Standard

The development of XTSP was a key contribution to this effort. XTSP enables the import of live operations or live training scenarios or mission scenarios created in an external system, and places it in a standardized, machine-readable, and user-readable format that can be shared and reused among different synthetic or live training environments for specific competency targets. The XTSP subsequent export for use in other synthetic training environments not only makes it capable of being “agnostic” to any proprietary capability a training application provides, but it makes the setup and design for a given training environment, as well as evaluation and tutoring services like GIFT, easier for the typical Soldier who will need to use the STE capabilities for training. For example, scenarios created for RIDE can be run on VBS3 as well. XTSP also simplifies and enables interoperability of other adaptive learning and assessment systems like GIFT.

If XTSP is not used as a common interchange format, then each system would have to build an interface for every other system. This also enables vendors to produce their own proprietary format which will prevent the ability to re-use experiences across different environments. For STE, as long as each training application or training assessment and learning support system (like GIFT) has the ability to import and export XTSPs, they will be able to work functionally and seamlessly with other systems during collective training within the STE. Mathematically, if XTSP is used:  $N * (N-1)/2$  complexity is reduced to  $(N+1)$  complexity, where N is the number of systems to interoperate.

## The Enhancements of the RIDE

The benefit of RIDE is its open architecture, whereas most proprietary simulators and simulation engines lock down their internal data. By adding JavaScript based scripting to RIDE and enabling it to run XDT generated scenarios, we made it possible for researchers not skilled in C# programming (Unity’s native programming language) and/or ordinary users to create and execute synthetic scenarios over a web-service network using RIDE, as well as research data streaming resources, as we did by interfacing triggered “active data sources” within RIDE with GIFT. These capabilities for enhancing the data collection and assessment capabilities may one-day be resident in all simulators and synthetic training engines or even in gaming environments.

## LESSONS LEARNED

---

### Use an open, well specified, easy to use language for interfaces

XTSP is at its core a language that is specified in JSON format to make it usable by other programs. By making the implementation of XDT separate from its interface language, XTSP, and by defining the XTSP well we significantly increased the value of our efforts. Other tools can now use XTSP as an interface and leverage each other’s and XDT’s capabilities.

### Target users and not developers

We wanted to make it easy to create and run training scenarios developed in XDT. To enable our training scenario to run in RIDE we added JavaScript support to RIDE. This allows non-technical users of XDT to create and run training scenarios supported by XDT without any developer support.

## Software development process insights

We used mature, well-developed, and widely used technologies and concentrated our development efforts on addressing requirements for the XDT. The use of such tools allowed development to proceed faster. Most issues encountered had already been encountered by other users in the past, and their solutions were readily available.

## CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

The use of XTSP makes it possible for different training applications to work together. It also simplifies the task of interfacing one application to another. A training scenario generated using one training application can to a large extent use the capabilities of other training applications. For example, a standalone training application created in VBS, can be imported by XDT, and then instrumented to use GIFT for assessment and control.

We have used XDT and XTSP to run sample training scenarios. Wider and continued usage will expose further requirements and possible deficiencies in the system. These can then be addressed to create a more robust and versatile tool. We expect XTSP to be enhanced with more fields as more data elements are needed.

We need to do further work in making XDT interoperable with VBS and RIDE. We want to make it possible to easily import missions developed in VBS Mission Editor to XTSP, and also have a database of common 3D models. This database of 3D models will help pick a relevant 3D model in each simulation environment to use while running a XTSP.

While JSON is good for exchange of textual data it is not efficient in transferring large volumes of binary data or when data needs to be exchanged fast. Such needs arise while transferring binary 3D model data or when transmitting simulation state data. Modern protocols like protobuf are efficient in this regard. It would be worthwhile to investigate the specification of such protocols in XTSP. At run time both the synthetic training environment as well as GIFT could automatically be set up using the specified protocols and ports thus making the system easy to use and reducing manual setup time.

## REFERENCES

- Goldberg, B., Owens, K., Hellman, K., Robson, R., Blake-Plock, S., Hoffman, M. & Gupton, K. (2021). Forging proficiency and readiness through an experiential learning for readiness strategy. In Proceedings of the 2021 Interservice/Industry Training Simulation and Education Conference (I/ITSEC). Orlando, FL.
- Johnston, J., Burke, S., Milham, L., Ross, W., & Salas, E. (2018). Challenges and propositions for developing effective team training with adaptive tutors. In Johnston, J., Sottolare, R., Sinatra, A.M., and Burke, C.S. (Eds). *Building Intelligent Tutoring Systems for Teams: What Matters*. Research on Managing Groups and Teams (Vol 19, pp. 75-97). Emerald Publishing Limited.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Prentice-Hall. Englewood Cliffs, NJ.
- Milham, L., Phillips, H., Ross, W., Townsend, L., Riddle, D., Smith, K., Butler, P., Wolf, R., Irizarry, D., Hackett, M., & Johnston, J. (2016). Squad-level training for tactical combat casualty care: Instructional approach and technology assessment. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 1-16.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Owens, K.P., & Goldberg, B.S. (2022). Competency-Based Experiential-Expertise. In Sinatra, A.M., Graesser, A.C., Hu, X., Goldberg, B., Hampton, A.J., & Johnston, J.H. (Eds.). *Design Recommendations for Intelligent Tutoring Systems, Volume 9, Competency-Based Scenario Design*, pp. 19 – 29.
- Sottolare, R., Brawner, K., Goldberg, B. & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). US Army Research Laboratory.
- Sottolare, 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, 1-19.
- U.S. Department of the Army. (2017). Army Training and Leader Development: TP 350-1, PP 1-18.  
[https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/pdf/web/ARN18487\\_R350\\_1\\_Admin\\_FINAL.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN18487_R350_1_Admin_FINAL.pdf)

## ABOUT THE AUTHORS

---

**Sanjay Mishra** is a Researcher at Oak Ridge Institute for Science and Education (ORISE). Sanjay has conceived, designed, and delivered products for use in consumer, industry, military and space sectors for more than 30 years. Sanjay specializes in taking an idea from mathematical and physical theory all the way to its realization as a finished product comprising firmware and software running on packaged hardware. He has designed and developed hardware, firmware, device drivers, RF and wired communication systems, and system software. He holds a Bachelor's and Master's in Electrical Engineering.

**Kevin Owens** is an Engineering Scientist at the Applied Research Laboratories: The University of Texas at Austin. After retiring from the US Navy, accumulating years of experience as an instructor, curriculum designer, and operational trainer, Kevin earned a BS in Workforce Education and Development from Southern Illinois University and a MS in Instructional Design, Development, and Evaluation from Syracuse University. Kevin has since spent over 20-years in industry and academia conducting DoD-related Learning Engineering projects. He is currently working on Science and Technology projects for the US Army's Synthetic Training Environment, and researching a Competency-Based Experiential Learning model.

**Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.



# The Value Proposition of GIFT for the Field of Learning Engineering

Shelly Blake-Plock<sup>1</sup>, Kevin Owens<sup>2</sup>, and Jim Goodell<sup>3</sup>

<sup>1</sup>Yet Analytics, Inc.; <sup>2</sup>Applied Research Laboratories: The University of Texas at Austin; <sup>3</sup>Quality Information Partners

## INTRODUCTION

---

In order to show the value proposition of the Generalized Intelligent Framework for Tutoring (GIFT)<sup>2</sup> for the field of Learning Engineering (Blake-Plock et al., 2020; Goodell & Kolodner, 2023), this paper examines its use in the Synthetic Training Environment (STE) Experiential Learning for Readiness (STEEL-R) project (Owens et al., 2022). STEEL-R is a complex applied research project in the space of synthetic training and standards-based data design methodologies which leverages Experience Application Programming Interface Profiles (xAPI Profiles)<sup>3</sup> to link GIFT’s capabilities with a competency assertion engine.

Learning Engineering is “a process and practice that applies the learning sciences using human-centered engineering design methodologies and data-informed decision making to support learners and their development” (Goodell & Kolodner, 2023). The STEEL-R project aligns to the learning standards portfolio of the Institute of Electrical and Electronics Engineers (IEEE) Learning Technology Standards Committee (LTSC)<sup>4</sup> — a portfolio of data standards leveraged in the field of Learning Engineering when implementing the common data components of the Advanced Distributed Learning Initiative’s (ADL’s) Total Learning Architecture (TLA)<sup>5</sup>.

The LTSC portfolio contains the data standards which are key to TLA implementation — including standards for event-based data and data profile modeling<sup>6</sup>, learning activity metadata<sup>7</sup>, shareable competencies<sup>8</sup>, and learner records<sup>9</sup>. This portfolio supports the technical work of the IEEE IC Industry

---

<sup>2</sup> <https://gifttutoring.org/projects/gift/wiki/Overview>

<sup>3</sup> “The human and/or machine-readable documentation of application-specific concepts, statement patterns, extensions, and statement templates used when implementing xAPI in a particular context.” <https://github.com/adlnet/xapi-profiles>

<sup>4</sup> <https://sagroups.ieee.org/ltsc/>

<sup>5</sup> <https://adlnet.gov/projects/tla/>

<sup>6</sup> <https://standards.ieee.org/ieee/9274.1.1/7321/>

<sup>7</sup> <https://standards.ieee.org/ieee/2881/10248/>

<sup>8</sup> <https://standards.ieee.org/ieee/1484.20.3/10749/>

<sup>9</sup> <https://standards.ieee.org/ieee/2997/10632/>

Consortium on Learning Engineering<sup>10</sup> as well as to Learning Engineering professionals and learning technology stakeholders globally. If GIFT were to be considered valuable to Learning Engineering as a bridge between synthetic training environments and standards-based distributed learning data architectures such as the TLA, it would have to be able to monitor or accept event-driven learning experience data from synthetic training scenarios — and then translate those events into the xAPI data format following the principles defined in one or more xAPI Profile statement templates. This data would be validated by a series of conformant xAPI Learning Record Stores<sup>11</sup> in order to maintain data integrity and be useful to standards-compliant business systems within the TLA. In the use case represented by this paper, that business system was the Competency and Skills System (CaSS)<sup>12</sup> whose assertions leveraged the xAPI data as evidence of competency-aligned learning activity.

As part of the STEEL-R project, GIFT was instrumented to produce xAPI data following the model of xAPI Profiles suggested in prior GIFT literature (Blake-Plock et al., 2020; Blake-Plock et al., 2021). Alignment with xAPI Profiles ensures that GIFT's data is conformant with all aspects of expected event-based data flow through the TLA. These events provide evidence for the assertion of competencies. The STEEL-R project itself and its place within competency-based experiential learning has been addressed in research papers published throughout the three-year duration of the project (Goldberg et al., 2021; Hernandez et al., 2022; Owens, Goldberg, Ray et al., 2022; Owens, Goldberg, Robson et al., 2022).

This paper will demonstrate that GIFT or any other adaptive learning service<sup>13</sup>, can be leveraged as a bridge — or standards-based data broker — between synthetic-based experiential learning and the TLA. It is therefore argued that GIFT is of value to Learning Engineering projects — especially those that rely on capturing the rich experiential data that occur in synthetic environments. Implemented as an event-based data broker, GIFT can make the data that represents those experiences available to the data validation, storage, and consumer components of standards-based data architectures. This research will be impactful for the design of future Learning Engineering projects as it clearly demonstrates how GIFT can be leveraged as this form of standards-based data broker in complex experiential learning environments.

## CORE PRINCIPLES

---

### Learning Engineering

Learning Engineering (LE) is a term first coined in the 1960's<sup>14</sup> that more recently has expanded to describe a field that has inspired the development of communities of practice across organizations including at the IEEE, Carnegie Mellon University, Arizona State University, and among other global

---

<sup>10</sup> <https://sagroups.ieee.org/icle/>

<sup>11</sup> <https://github.com/adlnet/xAPI-Spec/blob/master/xAPI-About.md#def-learning-record-store>

<sup>12</sup> <https://github.com/cassproject/CASS>

<sup>13</sup> [https://www.gifftutoring.org/projects/gift/wiki/Adaptive\\_Learning\\_Service\\_API\\_2023-1](https://www.gifftutoring.org/projects/gift/wiki/Adaptive_Learning_Service_API_2023-1)

<sup>14</sup> <https://www.cmu.edu/simon/>

partners and research institutions. The IEEE formally defined LE in 2018 as “a process and practice that (1) applies the learning sciences, (2) using human-centered engineering design methodologies, and (3) data-informed decision-making to support learners and their development” (Goodell & Kolodner, 2023). LE synthesizes various principles of learning science, instructional systems design (ISD), software and systems engineering, data science, and other associated disciplines to develop methods of designing, building, and measuring human learning experiences, environments, and products for the purpose of solving modern learning challenges. Learning Engineering goes beyond just delivering learning experiences and concerns itself with all of the factors that help or hinder learning by teams or individuals.

LE provides a super-positional perspective across a broad range of specialized disciplines and expertise including learning science, neuroscience, engineering psychology, cognitive science, data science, information science, computer science, human factors engineering, information technology, and the creative arts. In the case of STEEL-R, each discipline — as represented by members of the project — provided the perspectives and skills needed to provide the functional solutions within the project’s complex human learning model. The common factor among approaches within STEEL-R — including learning science, cognitive psychology, mathematics, and software engineering — was GIFT itself. In the same way that GIFT can be implemented as central to the exchange of data between synthetic training environments and the TLA, this research suggests that it also offers key value to LE projects at the intersection of those paradigms.

### **Competency-Based Experiential Learning**

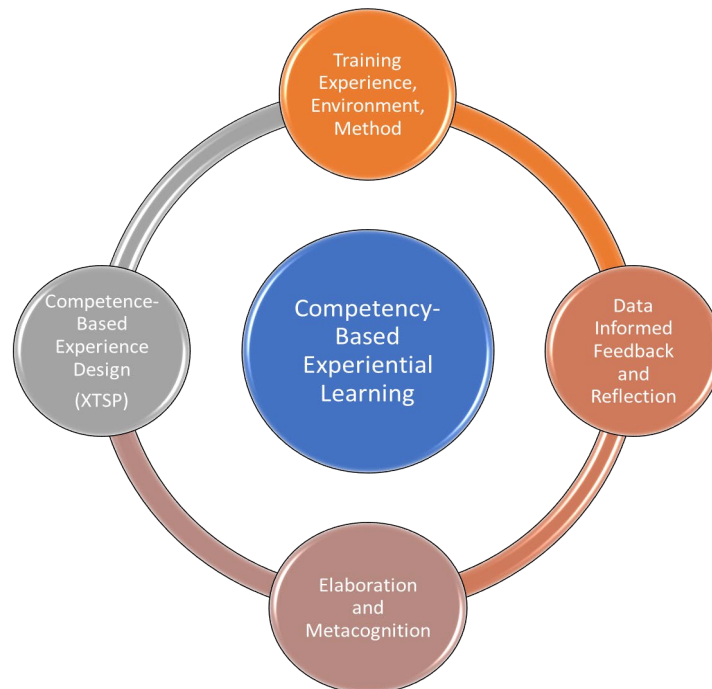
The Experiential Learning Model (ELM) (Kolb, 1984), supports a system based on reflection and repetition. There are many forms of experiential learning; however, in the case of STEEL-R, what was needed was a form of experiential learning that helped the Army produce Soldier and Team level competence (predictive performance) for specific violent, volatile, uncertain, complex, and ambiguous sets of conditions that may occur in combat. A key premise of STEEL-R is that new warfighters not only lack experience in applying the declarative and procedural knowledge they learn in traditional institutional education and training, but are missing the more critical tacit knowledge they can only develop through experience. Tacit knowledge (i.e., experience) is what provides the critical “pre-attentive” abilities:

- to be situationally aware
- to note anomalies
- to focus more on parts of the environment that could be mission critical
- to build resilience in combat related trauma
- to make key split-second decisions

Warfighters need to be more cognitively effective to accomplish many of the US Army’s mission-essential tasks in real operational conditions in the live combat environment. In its experiential learning approach, STEEL-R employed a long-term concept of task-based competence to ensure all forms of knowledge and skills were present and recently measured. This is to help commanders assess unit readiness for a given mission. This version of experiential learning is referred to as competence-based experiential learning (CBEL) (Owens, Goldberg, Hernandez et al., 2022). See Figure 1.

CBEL applied in STEEL-R is described as follows:

- an actor/learner has an imperative to learn, improve, or validate ability to perform a task and thus is prompted to perform within a concrete experience — this includes varying conditions, context, novelty, and difficulty.
- depending on the actor’s competence level — thus their training phase and the type of synthetic training modality used — a learner/actor may receive adaptive instruction and/or feedback during or after each experience as an individual or as a member of a team.
- As the results of the task equate to success, and causal factors from the tasks’ various measures become apparent through data-informed feedback, the actor is able to inspect the raw data and reflect on what they themselves or the team did well compared to an official norm or unofficial (peer) criteria.
- as part of this reflection, actors or team leaders elaborate on their current models and practices by imagining ways they could have performed better in the experience — essentially instantiating a form of metacognitive strategy (using self-developed or indicative best practices from others) to improve in future iterations of performing the subject task.
- this then motivates them to participate in further experiential learning either within a real-team setting or by engaging with artificial intelligent entities that emulate their teammates, opposing forces, or other challenge agents as the tasks require.



**Figure 1. Competency-Based Experiential Learning Model**

For the conceptual research-challenge of this project, the interdisciplinary STEEL-R team benefitted from the LE approach in employing learning science, technology, and data in a systematic way. The research team knew that the development of tacit knowledge was something that a synthetic training environment

could provide. However, it was necessary to engineer a process to create those experiences and manage them over time.

Because time is another challenge in operational training and is a requirement in making experiential learning efficient, a premise of STEEL-R was that it would have to facilitate an ability to render multiple experiences and opportunities to learn in the same timeframe that a single live-exercise would require. Further, it necessarily would produce the data required for inspecting and assessing readiness. Note that a key challenge for STEEL-R was not only the tracking of data to evaluate levels of performance in an adaptive learning service like GIFT, but the tracking of the experiences themselves. This included the amount of experience as rendered in event-based data as well as the conditions or novelty of those experiences — qualifiers and contextualizers of what the activity that comprised those events meant. Our solution was to employ the xAPI data model to provide a solution in achieving those goals — both from the perspective of brute activity tracking as well as through more nuanced and business-centric data profiling.

An engineering challenge was that while the traditional Experiential Learning Model (ELM) is focused primarily on academic individual student experiences, one of the significant challenges for STEEL-R was to engineer an ELM into an operational training context that included teams — and to do this while integrating the experiential learning to support a competency-based structure. STEEL-R defines competence as a long-term state that should only be calculated objectively through data and objective math models. So it was necessary to engineer a solution that could provide for the requirements of the math models while also fulfilling the requirement of CBEL to incorporate experiential learning for teams that collectively performed as a single entity. This required an expansion of the ELM that superimposed a “team-mind” — as well as the tracking of the cross-training and shared mental models that enabled — beyond just the individual task-based knowledge and skills. This approach required a more ethnographic learning mode to include new evaluation and feedback techniques that supported the CBEL model.

## **Applied Research Operations**

STEEL-R team members managed the research project by employing elements of what is called Development Operations (DEVOPS) — or Agile methods as leveraged in the software and information technology industries. These methods were exercised in an applied research context. Because the purpose of applied research is to create practical applications of accumulated theories, knowledge, methods, and techniques to achieve practical, usable, and operational prototypes that can then be later more formally developed in actual business operations, STEEL-R employed an iterative and cyclic process for rapid prototyping. LE methods were deployed throughout the process — such as using data from both reliability testing and psychometrics to refine the development and research lifecycle of the learning technologies and processes present throughout the project.

GIFT itself proved central to this LE-flavored applied research approach due to its pivotal role as the data broker bridging the gap between the observation of synthetic activity and the processing of that activity according to the TLA’s business logic. In the same way that testing and psychometrics conditioned the development and research lifecycle, so too did modular changes and iterations to the xAPI Profile applied to GIFT’s output data allow for insight into the variety of ways that the data model could influence the

whole host of technical and learning insight challenges — from economy of data and efficiency of throughput to the alignment of learning experiences with the evidentiary requirements of business systems. In this way, GIFT was the core technology allowing for the super-positional perspective of LE to all aspects regarding the influence of data on the outcomes of the project.

### **Bringing it All Together**

STEEL-R itself was made possible by the coalescing of key attributes from LE, CBEL, and Applied Research Operations. As such, it would be a fertile ground for examining principal components of its technological ecosystem and, in the case of identifying GIFT as a super-positional broker between data inputs and outputs across the STEEL-R ecosystem, could be a value-add to LE projects wrought more broadly.

## **METHODOLOGY**

---

### **Integrating GIFT with the Total Learning Architecture**

Many of GIFT's existing modular functions were originally designed over a decade ago for supporting a learner-centric computer-based tutoring system (CBTS) and related research (Sottolare, et al., 2017), and they have been expanded upon since. Therefore, GIFT today offers many modular services and functions relevant to sophisticated learning formats, including:

- andragogical on-demand learning applications
- up-skilling platforms
- human/artificial intelligence (AI) interactions
- CBEL experiences

An original purpose of STEEL-R was to research the question of whether it was possible to incorporate a form of Experiential Learning Theory (ELT) (Kolb, 1984) into the Army's new Synthetic Training Environment (STE) program. Key research questions were:

- Could GIFT's modularity and unique position within a learning experience help to incorporate ELT practices into training?
- Could GIFT help to make sense of the data artifacts created in synthetic training?

If GIFT could incorporate ELT practices and help to make sense of the data created during synthetic training, it would need to provide or support a means of maintaining sense — both through interoperable semantics and orthogonal syntax — between the domain of the synthetic learning experience and whatever system would need to make use of the data created there.

The ADL's TLA is a research and development project that has been created in collaboration with multiple stakeholders from across the defense community, standards bodies, industry, and academia. It includes a set of technical specifications and policy guidelines that define a standards-based approach to integrating multiple learning technologies and services through a comprehensive data architecture. The

standards themselves that support data flow through the TLA are maintained by the IEEE LTSC. As such, the TLA served as a baseline architectural reference in the design of the STEEL-R ecosystem and a guide to the standardized flow of data through the system components.

So, where does GIFT fit into this?

GIFT was established as the ideal data broker for a TLA implementation that would connect synthetic training to automated competency assertions and real-time analytics and visualization for after action review (see Goldberg et al., 2021; Hernandez et al., 2022;).

In the GIFT-as-data-broker paradigm, GIFT collected raw data based on the engagement within a GIFT-enabled scenario. That raw data was processed and emitted as xAPI statements according to the rules of an xAPI Profile designed for GIFT. Those statements were validated by a Noisy SQL Learner Record Store (LRS) and this xAPI data was filtered through LRSPipe — an open source forwarder-filter capable of sending xAPI statements according to the patterns in an xAPI Profile. Essentially, the second xAPI Profile governs the business logic of the data entering into the transactional layer of the TLA where it is validated by a Transactional SQL LRS. This filtered transactional data is leveraged as evidence by the CaSS competency assertion engine.

By instrumenting GIFT to emit event-based activity data modeled on xAPI Profile statement templates, the implementation saw synthetic activity collected as inputs to GIFT and saw GIFT emit outputs into the TLA business stream that was conformant with TLA specifications. The result was the ability to observe behavior that occurred in a synthetic training scenario and have the activity of the learners in that scenario immediately be translated into proficiency levels in real-time. By the time a training scenario was finished, instructors and commanders had a clear picture of the competency of their teams.

### **A Note on Process**

Regarding the research itself, first a general concept of operations was defined over many iterations of presentations and feedback from stakeholders. At the same time, a functional architecture for how key elements of the project would sync came together over several rounds of exploring existing capabilities and capabilities in the TLA software ecosystem that would have to be modified and adapted to serve the needs of the research. This architecture had to include an ontology of terminology, competence, and task structures, as well as an overall philosophy of how learning would occur, how data would be employed, and how learners would be evaluated. And the software ecosystem would grow to include new capabilities built during the course of the project for the purpose of solving problems within the project — such as the design and development of an Experience Index that would offer the benefit of standardized metadata describing the learning experiences. When the research team did not have the tool to do a particular job, a new one was invented and built.

Additionally, in alignment with the LE process, the relationship between GIFT and CaSS was leveraged to iteratively determine the xAPI data instrumentation specifications that would allow event-based activity to pass into the transactional layer of the TLA where it could be used as evidence. These iterations were implemented and investigated over the course of the project and helped to determine the creation of new capabilities for taking on the challenge — including the design of products such as the Experience Index as well as the Experiential Design Tool described below.

For the purpose of project management, each of the architectural elements were broken-up into a set of increments established as epics — or commonly themed project elements and objectives — providing the scope and feature sets of minimum viable products (MVP) that would be researched and prototyped. Each epic was assigned an “Epic Lead” who would work with researchers across the various disciplines needed for the epic in order to determine which features to focus on and what MVPs would be possible to demonstrate at the end of an increment (typically 3-4 months long). Epic membership varied in that some members with certain skills and experience were assigned to multiple epics while others were focused on only a few. This membership process required all STEEL-R team members to learn about multiple disciplines outside their own specific expertise — making the project a great experiential learning process on its own.

Partly, the goal here was to explore LE methods in-the-wild, but also to examine how these iterative processes could be supported through the iteration of GIFT’s own model for emitting event-based activity data. In other words, as iterations — whether conceptual or tactical — forced different ways to think about outcomes, the xAPI Profile of GIFT and the related business filtering of its data flow through the TLA could be rapidly iterated. Rather than building a technology and waiting for a summative testing event to determine whether it worked, GIFT lent itself — through the flexibility of xAPI Profiles — to the LE process. GIFT brokered data, but it also brokered the ability to rapidly prototype and iterate upon what the data offered — resulting in the completion of ongoing cycles of creation, implementation, and investigation.

## **FINDINGS**

---

### **GIFT in Support of Learning Engineering**

GIFT provided a stable platform for developing many of the functionalities conceived to be needed as part of the STEEL-R architecture. For example, at onset, GIFT was able to provide multiple features to support the challenge of collecting and storing performance data. And once GIFT was instrumented with an xAPI Profile to emit standard patterns of xAPI data statements, it allowed for direct integration with the TLA. It would be this ability to bridge the gap between learning experiences and events occurring in synthetic environments and data standards originally designed to support distributed learning that would offer the greatest value to future LE projects. In the same way that the new generation of metadata standards for learning experiences<sup>15</sup> describe the experiences themselves no matter the domain in which they reside, GIFT could act to bring any aspect of synthetic experience both into the laboratory of learning science research and into the business streams of operational human capital and readiness systems.

Many of GIFT’s basic capabilities and modules were either employed as is or expanded upon in order to integrate within the larger experience. This included planning, preparing, execution, and evaluation processes, mitigation of which cut down on increment development time significantly. From a LE perspective, the efficiencies offered by GIFT — even before the inclusion of xAPI Profiles — cannot be overstated. For example, during the course of research, thousands of lines of event data were collected

---

<sup>15</sup> <https://adlnet.gov/news/2021/05/28/P2881-and-the-Harmonization-of-Learning-Metadata/>



from the training system as structured data — including Distributed Interactive Simulation<sup>16</sup> (DIS) messages. Researchers were able to reduce the data into specific task and measurement “bins” — pre-defined by an experiential design tool designed during the course of the project (more on this below). Researchers could then evaluate the data against set measures and variable criteria in real-time.

Once the xAPI Profile was in place, the results of all event-based activity and evaluations could be reported out to dashboards, could be accessed as evidentiary data by the competency engine, and were stored as immutable data. An xAPI Profile was established iteratively for GIFT while a second xAPI Profile was designed specifically for the purpose of establishing the business logic for the presentation of evidentiary data at the transactional layer of the TLA. The ability to rapidly iterate on the design and implementation of the xAPI Profiles meant that the researchers were able over time to hone in on exactly the data and the functions of the data that were necessary for the project. GIFT was the hub around which all of this rapid development occurred. From the LE perspective, the modularity of these capabilities align to the needs of evolving design patterns present in LE projects.

### **GIFT Research Supporting New Innovation**

In addition to the Experience Index and the xAPI Profiles described above, GIFT’s position within the STEEL-R project both inspired and necessitated other new innovations. This included new ways to design experiential scenarios that not only produced problems and engagements in which selected taskwork, teamwork, or affective competence could be tested and improved through the ELM, but that would support the creation of pre-defined relevant data streams that GIFT could push into the TLA.

This new data model was described as an Experiential Training Support Package (XTSP) (Owens, Goldberg,, Hernandez et al., 2022), based on a similar — but less descriptive and completely manual — product used in the US Army. The XTSP data model not only includes and expands upon the existing GIFT domain knowledge file<sup>17</sup>, but helps capture and define the synthetic concrete “experience events” that actors essentially would apply in their declarative and procedural knowledge (or learn how to). This is from where they would build their tacit knowledge. The XTSP also defines the data sources, the novel environmental conditions that the unit and/or actors require, and the difficulty of the experience. Additionally, the XTSP contains the idea of xAPI “micro-profiles” that are pre-established for the specific experience event being measured and reported.

The LE of the XTSP led to the development of an Experiential Design Tool (XDT) for authoring and iteratively testing experiential learning outcomes before publishing for use. The XDT works from inputs on the experiential learning phase of what active experimentation is needed and the specific synthetic learning environment being used. It then enables the designer to make incremental changes to match and elaborate the new experience to a learner’s competence and past experience. The intent of the XTSP is to improve both the trainer and the learner experience by automating the setup of a synthetic learning environment — as well as the GIFT performance domain and assessment structure. This way, the trainers and actors can focus on preparing and executing the experiential learning process. Further, the XTSP

---

<sup>16</sup> <https://standards.ieee.org/ieee/1278.2/6202/>

<sup>17</sup> [https://www.gifttutoring.org/projects/gift/wiki/Domain\\_Knowledge\\_File\\_2023-1](https://www.gifttutoring.org/projects/gift/wiki/Domain_Knowledge_File_2023-1)

could provide another key learning data standard empowering GIFT to help broker data for future LE projects.

## **CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

---

### **Learning Engineering Conclusion**

*As a platform, GIFT bridges synthetic training and distributed learning data standards.*

The findings noted above provided many valuable lessons that demonstrated how GIFT provided a significant tool for future LE. Partly, this is due to GIFT's advantageous position as a platform that, when instrumented with new forms of configuration content and xAPI, can bridge the gap between synthetic training and the business systems dependent upon data aligned to distributed learning data standards. This is a central value to LE projects at the intersection of what often are separated technological ecosystems — with training on the one hand and more formal education on the other.

### **Learning Engineering Implication**

*GIFT can be a fundamental part of the solution to scaling synthetic training.*

As a platform, GIFT and its capabilities as a data broker between systems and paradigms lends its use and value beyond learning scientists and more traditionally considered learning professionals to include the broader mix of material-based and software engineers, mathematicians, designers, and data standards professionals — all of whom associate as peers within the construct of a LE team. Because GIFT is accessible and of value across these fields and because a LE project will be facilitated by expertise across these fields, it should not escape the notice of those considering the scaling of technologies and processes in order to fully leverage new domains, that GIFT and LE complement one another so well. It is noted that a project such as STEEL-R is an example of the linkage between the capability, competence, and coherence necessary to scale complex technical learning objectives such as the implementation of synthetic training.

### **Need for More Research**

One of the more difficult demands in this field of research is the need for a common vernacular and ontological authority for describing LE processes and associated meanings. It is challenging to rely on what occasionally delves into illusively descriptive language when attempting to align technological capabilities and LE processes with the terminology of a specific domain such as Army training.

Interestingly, this need for an ontology may be the basis for a new Grand Challenge whereby the intent is to enumerate LE tools and processes via a categorization of methods and capabilities intended to scale standards-based learning solutions. In the same way that this project has recast GIFT as a broker of data and learning experiences between synthetic training environments and TLA-compatible business systems, an aspect of this Grand Challenge could be to reimagine the role of the intelligent tutor to include AI-

enabled LE activities — including some aspect of automation in the standardization of scalable learning data.

## REFERENCES

---

- Blake-Plock, S., Goodell, J., Kurzweil, D., Kessler, A., Olsen, J. (Eds.) (2020). IEEE IC Industry Consortium on Learning Engineering: Proceedings of the 2019 Conference on Learning Engineering. The Institute of Electrical and Electronics Engineers, Inc. New York, NY.
- Blake-Plock, S., Hoyt, W., & Casey, C. (2021). Instrumenting GIFT with xAPI: a use case for IEEE P9274.3.x standards activity and implications for the broader field of ITS and AIS. Proceedings of the Ninth Annual GIFT Users Symposium. US Army Combat Capabilities Development Command - Soldier Center. Orlando, FL.
- Blake-Plock, S., Hoyt, W., Casey, C., & Zapata-Rivera, D. (2020). Data Analytics and Visualization for xAPI Learning Data: Considerations for a GIFT Strategy. In Sinatra, A.M., Graesser, A.C., Hu, X., Goldberg, B., and Hampton, A.J. (Eds.). Design Recommendations for Intelligent Tutoring Systems: Volume 8 - Data Visualization. US Army Combat Capabilities Development Command - Soldier Center. Orlando, FL.
- Goldberg, B., Owens, K., Gupton, K., Hellman, K., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy. I/ITSEC. National Training and Simulation Association. Orlando, FL.
- Goodell, J., & Kolodner, J. (Eds.) (2023). Learning Engineering Toolkit: Evidence Based Practices from the Learning Sciences, Instructional Design, and Beyond. Routledge. New York.
- Hernandez, M., Blake-Plock, S., Owens, K., Goldberg, B., Robson, R., Welch, T., & Ray, F. (2022). Enhancing the Total Learning Architecture for Experiential Learning. I/ITSEC. National Training and Simulation Association. Orlando, FL.
- Kolb, D. A. (1984). Experiential learning: Experience as the source of learning and development (Vol. 1). Prentice-Hall. Englewood Cliffs, NJ.
- Owens, K., Goldberg, B., Hernandez, M., Blake-Plock, S. (2022). Introduction to Competency-Based Experiential Learning. I/ITSEC. National Training and Simulation Association. Orlando, FL.
- Owens, K., Goldberg, B., Ray, F., Mishra, S., Blake-Plock, S. (2022). Multi-Session Demonstration of a Competency-Based Experiential Learning Model for the US Army Synthetic Training Environment (STE). iFEST 2022. National Training and Simulation Association. Alexandria, VA.
- Owens, K., Goldberg, B., Robson, R., Hoffman, M., Ray, F., Colburn, A., Hernandez, M., Blake-Plock, S., Casey, C. (2022). Synthetic Training Environment Experiential Learning for Readiness (STEEL-R). Proceedings of the Tenth Annual GIFT Users Symposium. US Army Combat Capabilities Development Command - Soldier Center. Orlando, FL.
- Sottolare, R.A., Brawner, K.W., Sinatra, A.M., Johnston, J.H. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Gift Tutoring. <https://gifttutoring.org/documents/31>

## ABOUT THE AUTHORS

---

*Shelly Blake-Plock is President and CEO at Yet Analytics. He has served as principal investigator on projects related to xAPI and the Total Learning Architecture, including the ADL-sponsored Data and Training Analytics Simulated Input Modeler (DATASIM). Active in the global standards community, he is an officer of the IEEE Learning Technology Standards Committee and chairs the P9274.4.2 working group on xAPI Cybersecurity as well as the LTSC Technical Advisory Group on xAPI.*

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

**Kevin Owens** is an Engineering Scientist at the Applied Research Laboratories: The University of Texas at Austin. After retiring from the US Navy, accumulating years of experience as an instructor, curriculum designer, and operational trainer, Kevin earned a BS in Workforce Education and Development from Southern Illinois University and a MS in Instructional Design, Development, and Evaluation from Syracuse University. Kevin has since spent over 20-years in industry and academia conducting DoD-related Learning Engineering projects. He is currently working on Science and Technology projects for the US Army's Synthetic Training Environment, and researching a Competency-Based Experiential Learning model.

**Jim Goodell** is editor and co-author of Learning Engineering Toolkit and is Director of Innovation at Quality Information Partners, where he helps lead development of the US Department of Education sponsored Common Education Data Standards and where he supports the US Chamber of Commerce Foundation's T3 Innovation Network and co-led development of the Learning and Employment Record (LER) Wrapper Specification.. He is Chair of the IEEE Learning Technology Standards Committee, chaired the workgroup 1484.20.3-2022 Standard Data Model for Sharable Competency Definitions and Adaptive Instructional Systems Interoperability, and serves on the ICICLE Steering Committee.



**THEME VI:  
AUTHORING TOOLS AND  
INTERFACES**



# A Comparative Analysis Interface to Streamline After-Action Review in Experiential Learning Environments

Caleb Vatrall<sup>1</sup>, Naveeduddin Mohammed<sup>1</sup>, Gautam Biswas<sup>1</sup>, Nicholas Roberts<sup>2</sup>, and Benjamin Goldberg<sup>3</sup>

<sup>1</sup>Vanderbilt University; <sup>2</sup>Dignitas Technologies; <sup>3</sup>US Army Combat Capabilities Development Command (DEVCOM) - Soldier Center<sup>3</sup>

## INTRODUCTION

---

Experiential learning is a philosophy of education that emphasizes the importance of firsthand experiences and learning by doing (Andresen et al., 2020). In contrast to traditional lecture-based instruction and abstract problem solving, experiential learning encourages students to engage in hands-on activities, problem-solving tasks, and real-world applications of concepts and practices to gain a deeper understanding of the material. An important component of effective experiential learning, especially in complex domains, is the debriefing/after-action review (AAR) process that occurs after an exercise (Tokarieva et al., 2019).

During AAR, participants engage in a structured discussion about what went well and areas for improvement. This discussion is guided by an expert instructor and may involve the use of comparative examples and contrasting cases to provide concrete examples of how trainees can improve their skills and strategies in the training environment (Hanoun et al., 2018). Two major types of performance comparison are commonly employed during AAR. The first involves comparing an individual's or team's performance to an expert model, while the second involves comparing current performance to past performance. While the benefits of comparative methods have been shown in prior work (Schwartz et al., 2011; Sidney et al., 2015), adoption of these methods remains challenging due to issues surrounding data management and user interfaces that can provide relevant information that is actionable (Chromik & Butz, 2021; Vatrall et al., 2022a). In this paper, we describe our ongoing work in developing a structured framework for AAR reflection and evaluation through the use of comparative analysis. Our work aims to bridge the gap between theory and practice by creating a dashboard that allows for the direct comparison of session performance data to both expert models and past session performance. This dashboard is being developed as an extension to the Generalized Intelligent Framework for Tutoring (GIFT), which allows for a wide range of tutoring and training applications in various domains (Sottolare et al, 2017).

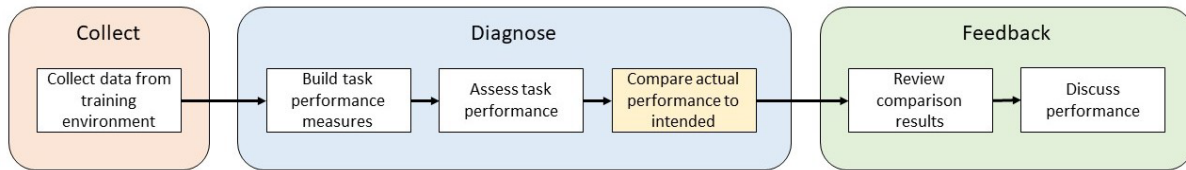
In the following sections, we will describe the framework for comparative analysis in more detail, including the design and implementation of the comparison interface in GIFT, our example case study using video analysis of Soldiers undergoing dismounted battle drill training, and the path forward for continued development work. We believe that our comparative analysis tool will be invaluable to trainees and instructors alike to aid in AAR and promote better learning outcomes.

## COMPARATIVE ANALYSIS IN AAR

---

A crucial element in experiential learning and training is proper assessment and feedback processes, which play a significant role in determining the outcome of the learning experience (Gegenfurtner et al., 2014; Ravert, 2002). Typically, feedback is given in the form of an AAR that takes place shortly after the training session, involving both instructors and trainees (Hanoun et al., 2018). This post-training review method is commonly used to avoid disrupting the flow of events during the training session. During the AAR, trainees and instructors examine the training events, interpret the sequence of events that occurred, analyze trainee responses, identify the connections between these responses and expected performance, and brainstorm about ways to enhance their overall performance in the future. The underlying idea of the AAR process is

that by scrutinizing trainee performance and the associated events, trainees gain valuable insights into self-reflection, comparisons with peers and experts, and a better understanding of the elements that constitute effective performance. In addition, they can get very specific information on how they may improve their performance from instructors and experts. Hanoun et al. (2018) proposed a theoretical model of AAR, which breaks down the process into three phases: (1) collection, (2) diagnosis, and (3) feedback, as depicted in Figure 1.



**Figure 1. The three-phase AAR framework adapted from Hanoun et al. (2018)**

Of particular interest in this work is the end component of the diagnosis phase, i.e., comparison of actual performance to the intended performance. This type of performance comparison is a critical component of effective AAR because it allows trainees to understand both how they have grown and areas which they can still improve performance. Two major types of performance comparison commonly employed in AAR are expert comparison and comparison against past performance. Expert comparison involves evaluating a trainee's or a team of trainees' performance against a benchmark set by experts in the field. This type of comparison is useful for identifying best practices and “gold standards”, and specific areas for improvement to come closer to achieving these standards. Expert comparison helps trainees to align their performance with these “gold standards” and adopt practices that have been proven effective and successful by experts in the field. Comparison against past performance involves evaluating a trainee's or team's current performance against its own previous performance. Such comparisons are important components of the reflection process (Visscher & Coe, 2003). They help trainees and their instructors to identify trends and patterns in their performance and assess the effectiveness of their current training efforts. By comparing current performance to past performance, trainees and their instructors can evaluate the impact of their training program and make informed decisions for future training improvements.

Because of the unique benefits of each type of performance comparison during AAR, both expert and past performance comparison can be combined to provide a structured framework for AAR reflection. By combining both types of comparison, organizations can develop a comprehensive understanding of their performance and make informed decisions for future training needs. However, in previous work (Vatral et al., 2022a), we highlighted some of the challenges associated in effective performance comparison for AAR. Critical among these challenges was the automation of comparative analysis, without which it would likely be missed or ignored by trainees and instructors because of the time-consuming nature of reviewing the full past performance record. Building on the theory presented in that paper, in this work we show how comparative analysis can be streamlined through a new AAR dashboard, which automates performance comparison and highlights specific comparative metrics and recommendations for trainees and instructors.

## **THE GIFT FRAMEWORK FOR COMPARATIVE ANALYSIS IN AAR**

To streamline the implementation of our comparative analysis AAR dashboard, we leveraged the extensible tools available in the Generalized Intelligent Framework for Tutoring (GIFT) (Sottolare et al, 2017). GIFT is an open-source software system developed by US Army, designed to support a variety of intelligent computer-assisted training and tutoring. GIFT uses data from multiple sources, including learner interactions with the system and other assessments, to dynamically adjust the instructional content and



delivery to optimize learning outcomes. GIFT is designed with a modular architecture for intelligent tutoring system (ITS) development, including several tools and modules that allow developers to create a system tailored to their specific needs. One such tool is the Game Master dashboard, which allows users to manage and monitor the game state and flow of events during and after training sessions. One of the key functionalities of Game Master is session playback, which allows instructors and trainees to review details and performance of a previous training session on a dynamic timeline to streamline debriefing and AAR. Since review of past sessions for the purposes of AAR is one of the primary functions of the proposed comparative analysis dashboard, we designed the prototype dashboard as an extension of the Game Master interface.

In Game Master, when the user selects the performance comparison button, a new modal window opens with a dropdown that allows the user to select from either the previously defined expert models or the past sessions. Once selected, Game Master will search all assessments which implement the new comparison functionality, supplying either the selected expert model or selected past session and the currently active Game Master playback session as parameters. This dropdown modal dialog is shown in Figure 2.

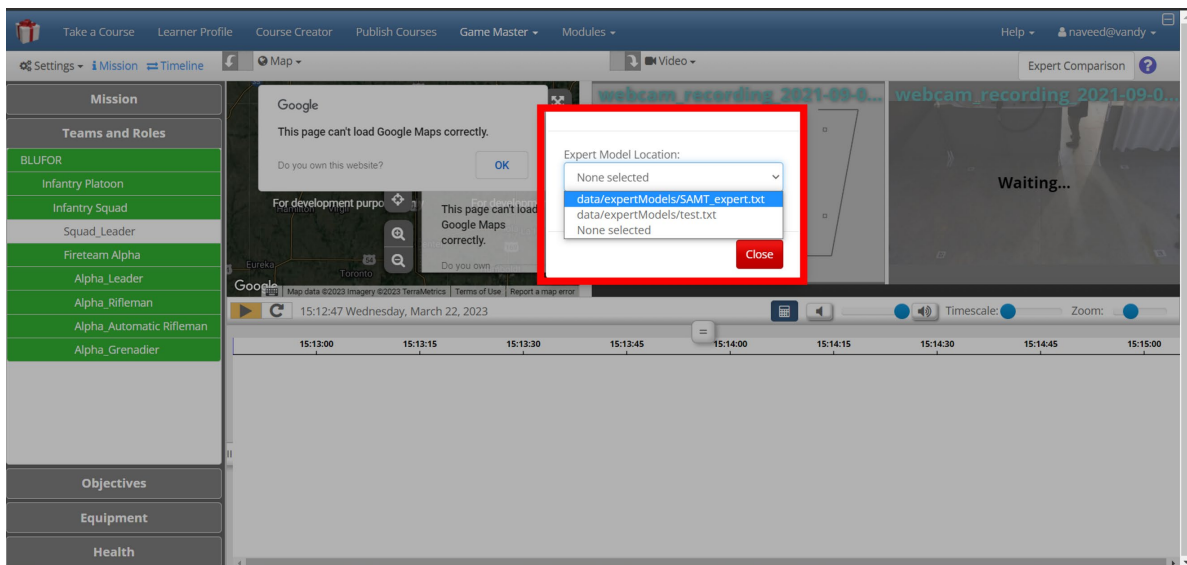


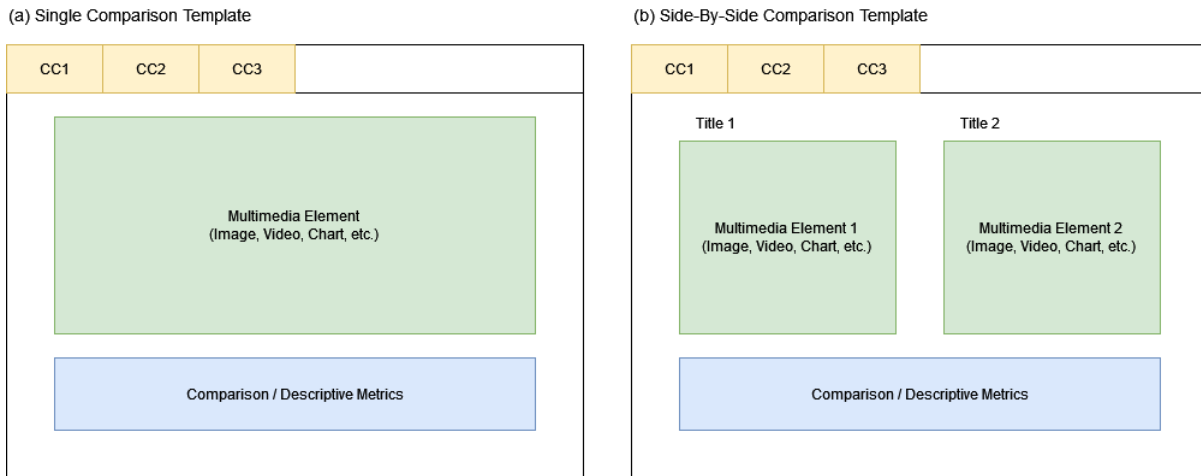
Figure 2. Dropdown modal dialog for selection of the expert model

When designing the comparison dashboard, we wanted to be careful to be faithful to the extensibility that is built into the core of GIFT, ensuring that any assessments built into GIFT could be easily extended to include comparative analysis capabilities. Since the primary framework for assessment in GIFT is the condition class, we designed the performance comparison as a new interface, which condition classes simply implement to define their comparative analysis capabilities. The interface has two simple methods, each with similar function signatures described below, which need to be implemented in order to register a condition class for comparative analysis.

First, the expert comparison method defines the logic required to compare the currently selected Game Master session against the selected expert model. The function takes two parameters: the current Game Master session data and a JSON object representing the defined expert model. The rest of the logic of the comparison is left up to the decision of the designers of the condition class, allowing as much flexibility as possible. To again provide maximal flexibility for assessment designers, the expert models, which are passed to the condition class for comparison, are user definable by simply extending a JSON text file stored in the GIFT data folder. Each expert model file follows a simple structure, as a list of JSON objects. Each

object in the list represents a single condition class and must define one object literal called *ConditionClassName*, corresponding to the name of the condition class that should be sent the given data when performing the expert comparison. All the other stored data in the object is up to the designer of the condition class assessments, providing maximal flexibility. Back in Game Master, when the expert comparison is requested, Game Master reads the selected expert model file, and passes the relevant JSON objects to each condition class based on their *ConditionClassName*. The dashboard’s expert model selection dropdown is populated by each JSON file in the specific GIFT data folder location, allowing users to easily define multiple expert models and swap between them at runtime.

Second, the past performance comparison method similarly defines the logic required to compare the currently selected Game Master session against another selected past session. The function again takes two parameters: the current Game Master session data and the past session data. To once again allow for maximum flexibility in assessment design, the rest of the comparison logic is again left up to the decision of the condition class designers. Past performance comparison is currently not fully implemented, but future work will complete development on this feature.



**Figure 3. Block diagram of the implemented comparison dashboard templates**

Finally, the two defined methods in the new interface have the same return object signature, based on a new flexible XML schema, which defines what should be displayed back to the user in Game Master for the comparison and is managed by GIFT’s AbstractSchemaHandler. Since each condition class and each assessment within the condition classes will likely want to display different information to the end user for the comparison, we designed the XML schema with a strong balance between flexibility and consistency of the user interface. The schema provides several templates from which the assessment designer can select to provide their comparison. As a starting point, we implemented two templates: single-comparison and side-by-side-comparison. In the single comparison template (see Figure 3a), the comparison involves a single multimedia element (picture, video, chart, etc.) to be displayed at the top and a list of text elements to be displayed underneath the media. This could be used, for example, in cases where a single chart or graphic is generated by the condition class to show the comparison, along with several text descriptions or other numeric metrics. The XML schema in this case receives one path to the multimedia element, usually stored in the output session folder, and a list of strings defining the text to be displayed. In the side-by-side comparison template (see Figure 3b), on the other hand, the comparison displays two multimedia elements at the top, each with a specified title, and a list of text elements to be displayed underneath the media. This could be used, for example, in cases where the assessment designer wants the user to directly compare an

image of the expert’s performance to an image of the trainees’ performance. Figure 4 shows an example of an XML element using the schema for the side-by-side comparison template.

```

<xAnalysis>
  <experienceAnalysisModel name="testModel">
    <expertModel>
      <sideBySideComparisonModel>
        <expertData> "Location of the expert data file" </expertData>
        <traineeData> "Location of the trainee data file" </traineeData>
        <textData id="overlap">
          <textValue>"Textual data related to comparison." </textValue>
        </textData>
        <textData id="similarity">
          <textValue>"Textual data related to comparison." </textValue>
        </textData>
      </sideBySideComparisonModel>
    </expertModel>
  </experienceAnalysisModel>
</xAnalysis>

```

Figure 4. Example of the return data XML schema for the side-by-side comparison template.

## CASE STUDY: VIDEO PROCESSING EAE CONDITION CLASS

---

To demonstrate the implementation and use of the new GIFT comparative analysis tool, we now present a case study of Soldiers training on dismounted battle drills. Data was gathered from a case study of two infantry fire teams who took part in a study at Fort Campbell over the course of two days. The teams were trained on the Enter and Clear a Room (ECR) dismounted battle drill, which involved entering a room and neutralizing all enemy combatants, while navigating and controlling other entities, such as civilians and obstacles in the room. Soldiers were trained to enter the room rapidly, following paths of least resistance to reach their points of domination and to neutralize enemies in their sector of fire. Once all combatants were neutralized and civilians secured, team members searched each entity to remove weapons before exiting the room. Soldiers were also required to be aware of changing task conditions that could impact their behavior, such as new combatants entering the room or the presence of explosive devices. The scenario was conducted using the Squad Advanced Marksmanship Trainer (SAM-T), which used a Virtual Battle Space 3 (VBS3)-generated scenario projected onto three screens in a U-shaped arena. The SAM-T system logged a variety of data, but for this study we focus on analysis of the video collected from overhead cameras.

Extending the video processing external assessment engine (EAE) previously developed for GIFT (Vatral et al., 2021; Vatral et al., 2022b), we implemented the interface methods required for the newly developed Game Master comparison dashboard. Since the video processing condition class primarily assesses psychomotor skills through movement patterns in this battle drill domain, we implemented the comparative analysis in a similar manner, focusing on Soldier movement patterns. We implemented the *side-by-side* comparison template defined by the Game Master comparison engine to display the map-view motion paths of the selected expert model next to the map-view motion paths of the Soldiers in the selected drill. This allows trainees and instructors to directly visually compare how the motion paths compare between the two. Within the expert model files, our JSON objects contain the required *ConditionClassName* field, as well as one other field defining the expert’s motion path. In addition to the visualization, we also implemented a similarity comparison metric based on the dynamic time warping distance between the expert and Soldier

motion paths. This similarity metric is displayed below the visualization as a text element. An example result of our comparison implementation for this ECR domain is shown in Figure 5.

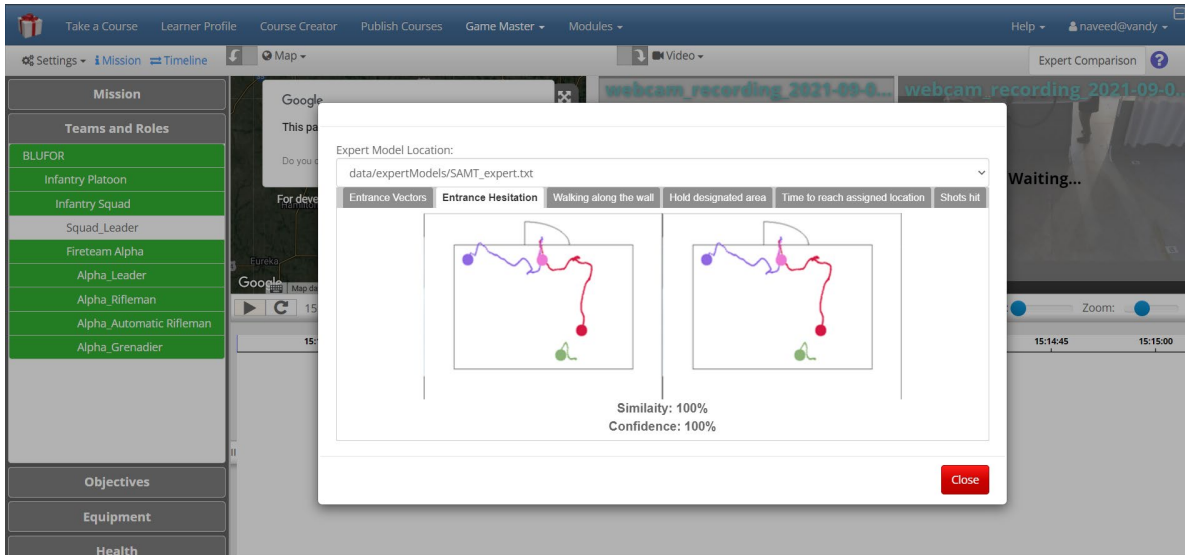


Figure 5. Example of the side-by-side comparison template implemented for the ECR case-study domain

## CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

In this paper, we have presented an ongoing project to develop a comparative analysis interface for AAR in GIFT. By allowing trainees to compare their performance to expert models and past performance data, the interface has the potential to promote self-reflection, identify areas for improvement, and ultimately improve learning outcomes of trainees in various experiential learning domains. We presented a theoretical framework for the use of comparative analysis in AAR processes and showed how our comparative analysis dashboard was implemented in GIFT by extending the Game Master dashboard. We also demonstrated the ease of implementing new comparative analysis into existing condition classes by showing an implementation case study using the video processing EAE condition class in GIFT.

While the preliminary case study results are promising, our work on developing the comparative analysis interface is ongoing, and there are several promising areas for future work. One important direction for future work is to test the usability and effectiveness of the comparative analysis interface with a range of stakeholders, including instructors, trainees, and domain experts. This will involve conducting user studies and collecting feedback on the interface design, as well as assessing the impact of the interface on learning outcomes. Another important area for future work is to extend the comparative analysis framework to include additional types of performance comparison, beyond the expert model currently supported. In the immediate next steps, we plan to implement past-performance comparison, which would allow instructors and trainees to directly compare current performance to past performance and identify trends and gaps. In addition, it may be useful to incorporate a comparison to peer performance, where trainees compare their performance to that of other trainees in the same exercise. We will also explore this peer comparison as another area for future work.

Finally, we believe that the comparative analysis interface has the potential to be a valuable tool for a wide range of educational and training contexts beyond the infantry battle drills examined here. For example, the interface could be applied to other military training exercises or to medical simulations. While the tools

developed here are designed to be extensible to such domains easily by implementing the comparison interface in condition classes, future work will also focus on exploring these other domains in more depth including user studies about the interface and its impact on learning outcomes.

We believe that the comparative analysis interface has the potential to be an invaluable tool for a wide range of educational and training contexts, and we look forward to continuing to develop and refine the interface in the future.

## REFERENCES

---

- Andresen, L., Boud, D., & Cohen, R. (2020). Experience-based learning. In *Understanding adult education and training* (pp. 225-239). Routledge.
- Chromik, M., & Butz, A. (2021). Human-xai interaction: A review and design principles for explanation user interfaces. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18* (pp. 619-640). Springer International Publishing.
- Gegenfurtner, A., Quesada-Pallarès, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*(6), 1097-1114.
- Hanoun, S., & Nahavandi, S. (2018). Current and future methodologies of after action review in simulation-based training. *2018 Annual IEEE International Systems Conference (SysCon)*, 1–6.
- Ravert, P. (2002). An integrative review of computer-based simulation in the education process. *CIN: Computers, Informatics, Nursing, 20*(5), 203-208.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of educational psychology, 103*(4), 759.
- Sidney, P. G., Hattikudur, S., & Alibali, M. W. (2015). How do contrasting cases and self-explanation promote learning? Evidence from fraction division. *Learning and Instruction, 40*, 29-38.
- Sottolare, 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*, 1-19.
- Tokarieva, A. V., Volkova, N. P., & Harkusha, I. V. (2019). Educational digital games: models and implementation.
- Vatral, C., Mohammed, N., Biswas, G., & Goldberg, B. S. (2021). GIFT external assessment engine for analyzing individual and team performance for dismounted battle drills. In *Proceedings of the Ninth Annual GIFT Users Symposium (GIFTSym9)* (p. 107). US Army DEVCOM–Soldier Center.
- Vatral, C., Mohammed, N., & Biswas, G. (2022a). Promoting Explainable Feedback in Simulation-Based Training through Contrasting Case Exemplars. In *Virtual Workshop Proceedings: Advances and Opportunities in Team Tutoring* (p. 11).
- Vatral, C., Biswas, G., Mohammed, N., & Goldberg, B. S. (2022b). Automated Assessment of Team Performance Using Multimodal Bayesian Learning Analytics. In *Proceedings of the 2022 Interservice/Industry Training, Simulation and Education Conference (IITSEC)*. National Training and Simulation Association.
- Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. *School effectiveness and school improvement, 14*(3), 321-349.

## ABOUT THE AUTHORS

---

*Caleb Vatral is a PhD student and research assistant at Vanderbilt University in the Department of Computer Science. Working at the Institute for Software Integrated Systems, their research focuses on combining theoretical foundations in distributed cognition and competency-based education with multimodal data-driven approaches to support*

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

*cognitive modeling and system design in simulation-based training and experiential learning environments. Prior to attending Vanderbilt, they received the B.S. degree in computer science and math from Eastern Nazarene College.*

***Naveeduddin Mohammed** is a Senior Research Engineer with the Institute for Software Integrated Systems at Vanderbilt University. Naveed received the M.S. degree in Computer and Information Sciences from University of Colorado. He is a full stack developer, and his work focuses on designing, developing, and maintaining frameworks for open-ended computer-based learning environments and metacognitive tutors.*

***Dr. Gautam Biswas** is a Cornelius Vanderbilt Professor of Engineering and Professor of Computer Science and Computer Engineering at Vanderbilt University. He conducts research in Intelligent Systems with primary interests in monitoring, control, and fault adaptivity of complex cyber physical systems, as well as developing intelligent open-ended learning environments that adapt to students' learning performance and behaviors. He has developed innovative multimodal analytics for studying students' learning behaviors in a variety of simulation and augmented reality-based training environments. He has over 600 refereed publications, and his research is supported by funding from the Army, NASA, and NSF.*

***Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.*

***Nicholas Roberts** is a senior software engineer at Dignitas Technologies and the engineering lead for the GIFT project. Nick has been involved in the engineering of GIFT and supported collaboration and research with the intelligent tutoring system (ITS) community for nearly 10 years. Nicholas contributes to the GIFT community by maintaining the GIFT portal ([www.GIFTTutoring.org](http://www.GIFTTutoring.org)) and GIFT Cloud ([cloud.gifittutoring.org](http://cloud.gifittutoring.org)), supporting conferences such as the GIFT Symposium, and technical exchanges with Soldier Center and their contractors. He has also been heavily involved in integrating GIFT into TSS/TMT.*



# ChatGPT in the Generalized Intelligent Framework for Tutoring

Faruk Ahmed<sup>1</sup>, Keith Shubeck<sup>1</sup>, and Xiangen Hu<sup>1,2</sup>

<sup>1</sup>The University of Memphis; <sup>2</sup>Central China Normal University

## INTRODUCTION

---

The Generalized Intelligent Framework for Tutoring (GIFT) has a rich collection of course objects (Sottolare et al., 2013, 2017) including Information as Text, Local and YouTube video, PowerPoint, AutoTutor (Graesser et al., 2003, 2012; Nye et al., 2014), and more. Including ChatGPT (*Introducing ChatGPT*, n.d.) from OpenAI in the course objects has a huge potential to automate, plausibly not fully but partially, the authoring process of a course. In this work we explore ChatGPT for authoring and integrating it into course objects in GIFT. Expectation-Misconception Tailored (EMT) conversation style is the tutoring framework of the AutoTutor (Cai et al., 2019). The authoring of a course takes a significant amount of effort, especially the EMT based tutoring system. Several AutoTutor Script Authoring Tools (ASAT) have been developed but it is still necessary to give the tools user friendly improvements (Cai et al., 2019). Integrating ChatGPT API (Application Programming Interface) will reduce the amount of effort to design EMT conversation scripts.

## OpenAI

---

OpenAI is a collection of models which can be used from “content generation to semantic search and classification” (*OpenAI API*, n.d.). The models are trained to understand and generate natural language or code, to generate and edit images with a prompt, to convert voice to text, and to convert text into numbers. Applications requiring chatbots, text completion, comparing texts, or code generation can easily use these models through API calls. This work is intended to use the chat functionality to author a conversation script. One of the models of ChatGPT is trained to follow a conversation (*Introducing ChatGPT*, n.d.). A simple conversational query like “generate a question from probability theory in true false format” generates a question as “True or False: In probability theory, the sum of all possible outcomes for a discrete random variable must equal 1.” (*Introducing ChatGPT*, n.d.).

## AUTOTUTOR

---

AutoTutor (Graesser et al., 2003, 2012; Nye et al., 2014) is an intelligent tutoring system (ITS) that uses conversational tutoring. The flow of a conversation starts from a question. The conversation branches between presenting an ideal answer and a pumping state depending on a student’s correct or incorrect answer. In the pumping state, the conversation encourages students to construct an answer. The conversation enters into a cycle of expectation/misconception if students construct partial answers. This cycle continues providing hints and prompts until students fulfill all the expectations. Conversation units vary in sizes. There are many variables in conversations depending upon the student’s knowledge state. The conversation becomes more complex for multi agent tutoring. Due to these variables, authoring a high quality script for an EMT conversation is a non-trivial task.

## EXPLORING ChatGPT

---

*Following conversation* is the most important feature of ChatGPT in the context of EMT dialogue. It is able to generate a question from a domain and generate follow up questions when asked. It also generates hints and prompts by simply asking it to “provide hints” or “provide prompts” for a question previously generated in the same conversation.

ChatGPT also generates content which is easy to integrate. If we need JSON or HTML format, we may simply request this in the conversation and it generates output in that format. For example, “provide a multiple choice question from Kirchhoffs’s current Law in json format” produced an output shown in Figure 1 (*Introducing ChatGPT*, n.d.).

```
{
  "question": "Which statement correctly describes Kirchhoff's Current Law (KCL)?",
  "options": [
    {
      "id": "a",
      "text": "The algebraic sum of currents entering a junction is equal to the sum of currents leaving the junction."
    },
    {
      "id": "b",
      "text": "The sum of the potential differences (voltages) around any closed loop or mesh in a network is zero."
    },
    {
      "id": "c",
      "text": "The total resistance in a parallel circuit is always less than the smallest resistance value in the circuit."
    },
    {
      "id": "d",
      "text": "The product of current and resistance in a circuit is equal to the voltage across the circuit."
    }
  ],
  "correct_answer": "a"
}
```

Figure 1. Multiple choice question generated in JSON format

## ChatGPT OUTPUT

---

Now the burning question is “how can we make sure that ChatGPT is generating correct text?”. There is no simple answer to this. As an Artificial Intelligence (AI) language model, ChatGPT generates responses based on patterns and statistical correlations learned from large datasets of human language. While ChatGPT can generate impressive and useful responses, it is not perfect, and there is always the possibility of errors or inaccuracies in its output.

One solution is to prevent ChatGPT from giving specific answers to questions. In the conversation we have to explicitly mention “*do not give answer*”, so that the ChatGPT does not include potentially incorrect answers. Instead, it should be allowed to generate *hints* and *prompts* to make students think, in the same way it happens in the *pumping* cycle of EMT.

To help ensure that ChatGPT is producing correct output, here are some best practices:

1. Consider the source: ChatGPT is a machine learning model that has been trained on a diverse range of texts, but it may not have encountered every concept or idea. We have to be cautious when asking questions outside of the model's training data, and we have to keep in mind that ChatGPT may not always be able to provide accurate responses to highly specialized or technical topics.



2. **Verify:** It is good practice to cross-check the information provided by ChatGPT with other reliable sources. We can use Latent Semantic Analysis (LSA) to verify the generated text (Graesser et al., 2007).
3. **Provide context:** The quality of ChatGPT's responses can be improved by providing clear and concise questions or prompts, as well as a relevant context. Providing context helps ChatGPT understand the intended meaning of the question and reduces the likelihood of generating irrelevant or inaccurate responses.
4. **Evaluate the response:** When receiving a response from ChatGPT, it is essential to critically evaluate the information provided. If the response seems dubious, factually incorrect, or inconsistent with common sense or previously established knowledge, it is best to disregard the response and seek other sources for verification.

ChatGPT is a powerful tool for generating human-like language, but it is not a substitute for critical thinking, fact-checking, and subject matter expertise. By using ChatGPT responsibly and critically evaluating its output, we can help ensure that it produces correct and useful information.

The ChatGPT API is integrated in GIFT and AutoTutor as a pilot project. A few parameter adjustments provides powerful means of authoring. Figure 2 shows a question from statistics generated by ChatGPT (*Introducing ChatGPT*, n.d.) and the prompt (e.g., iniPrompt) in the configuration window.

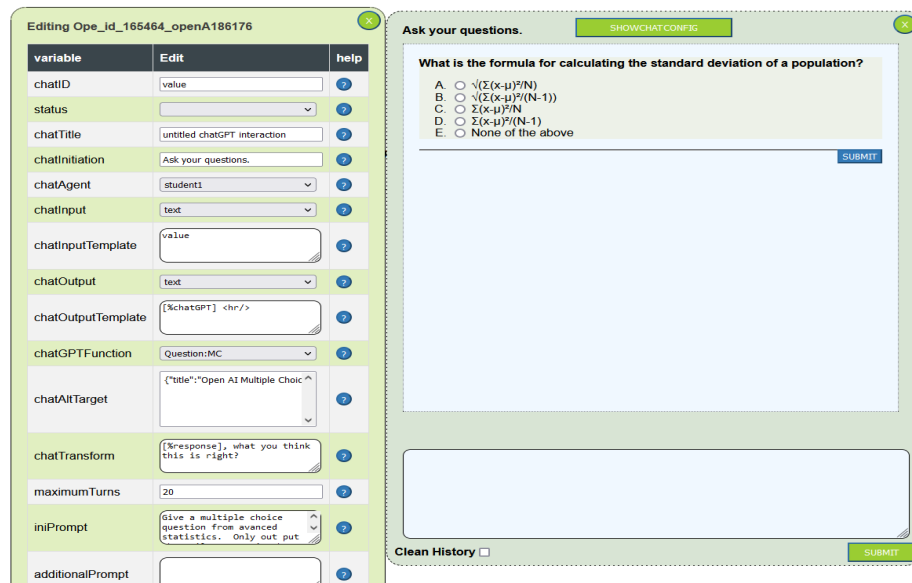


Figure 2. A question from statistics generated by ChatGPT (right) from a given prompt (left e.g., iniPrompt)

## VALIDATION

One way to verify the generated content with general questions is by asking specific questions regarding the generated content. For example, if you ask “generate a question with answers from advanced statistics” it will give you relevant content. Assume it generates this question with answer choices: “In a linear regression model, which of the following assumptions must be satisfied?” It is possible to then extract the

question from the content and ask ChatGPT, “In a linear regression model, which of the following assumptions must be satisfied?”. Then we can measure the distance of that generated content using LSA.

Validating ChatGPT responses automatically can be a challenging task since the output generated by the model can be subjective and context-dependent. However, here are some general steps that you can follow to validate ChatGPT responses automatically:

1. Define a set of evaluation metrics: Choose evaluation metrics that are relevant to task and domain. Some common metrics used for text generation tasks include perplexity, BLEU score, and ROUGE score.
2. Prepare a validation dataset: Collect a dataset of responses generated by ChatGPT that have been manually annotated as correct or incorrect. This dataset should cover a wide range of scenarios and topics that ChatGPT is expected to handle.
3. Use automated evaluation methods: Use automated evaluation methods to evaluate the generated responses against the validation dataset. These methods include using the evaluation metrics defined in step 1, but can also include other methods such as similarity metrics, topic coherence, and sentiment analysis.
4. Fine-tune the model: Use the feedback from the automated evaluation methods to fine-tune ChatGPT. This can involve adjusting the model's architecture, training data, or hyperparameters, which is beyond the scope of this work. OpenAI provides a way to re-generate the answer, and thumbs up and thumbs down to inform about the quality of the produced text.
5. Repeat the process: Continue to validate the ChatGPT responses automatically on a regular basis to ensure the model is performing well and to identify any areas that require improvement.

It is worth noting that while automated evaluation methods can be useful, they are not always sufficient to capture the full range of quality in the generated responses. Therefore, it is important to also incorporate human evaluation to ensure that ChatGPT is generating responses that are natural, coherent, and helpful.

## CONCLUSION

---

In this work we explored the potential of ChatGPT as an authoring tool for conversation design and assessment. We also integrated ChatGPT in GIFT and AutoTutor. AutoTutor is being extended to allow ChatGPT to drive the agent/learner conversation around a problem and learning objective. ChatGPT does well with generating conversational text, however it is possible that the generated text is incorrect or inaccurate. Considering this, it is critical to establish a validation measure.

## REFERENCES

---

- Cai, Z., Hu, X., & Graesser, A. C. (2019). Authoring Conversational Intelligent Tutoring Systems. In Sottolare, R.A. & Schwarz, J. (Eds.), *Adaptive Instructional Systems* (Vol. 11597, pp. 593–603). Springer International Publishing. [https://doi.org/10.1007/978-3-030-22341-0\\_46](https://doi.org/10.1007/978-3-030-22341-0_46)
- Graesser, A. C., D’Mello, S., Hu, X., Cai, Z., Olney, A., & Morgan, B. (2012). *AutoTutor* [Chapter]. Applied Natural Language Processing: Identification, Investigation and Resolution; IGI Global. <https://doi.org/10.4018/978-1-60960-741-8.ch010>
- Graesser, A. C., Jackson, G. T., Matthews, E., Mitchell, H. H., Olney, A., Ventura, M., Chipman, P., Franceschetti, D., Hu, X., Louwse, M. M., & others. (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 25.

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

Graesser, A., Penumatsa, P., Ventura, M., Cai, Z., & Hu, X. (2007). Using LSA in AutoTutor: Learning through mixed-initiative dialogue in natural language. In *Handbook of latent semantic analysis* (pp. 243–262). Lawrence Erlbaum Associates Publishers.

*Introducing ChatGPT*. (n.d.). Retrieved March 19, 2023, from <https://openai.com/blog/chatgpt>

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24(4), 427–469.

*OpenAI API*. (n.d.). Retrieved March 19, 2023, from <https://platform.openai.com>

Sottolare, 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*, 1–19.

Sottolare, R. A., Graesser, A., Hu, X., & Holden, H. (2013). *Design recommendations for intelligent tutoring systems: Volume 1-learner modeling* (Vol. 1). US Army Research Laboratory.

## ABOUT THE AUTHORS

---

**Dr. Faruk Ahmed** is a faculty member in the Department of Electrical and Computer Engineering at The University of Memphis (UofM). Dr. Ahmed received his MS in electrical engineering from UofM and Ph.D. in engineering from UofM. His primary research areas include Assistive Technology Development, Human Computer Interaction (HCI), Machine Learning, Computer Vision, and Adaptive Instructional Systems (AIS).

**Dr. Keith Shubeck** completed his doctoral degree in cognitive psychology at the University of Memphis, currently he is serving a post-doctoral fellow at the same department. He received a Certificate of Cognitive Science and a M.S. in psychology in December 2015. He works with his major professor, Dr. Xiangen Hu, in the Advanced Distributed Learning Partnership Lab at the Institute for Intelligent Systems. His dissertation work compares interactive and vicarious tutoring frameworks in a conversation-based intelligent tutoring system for critical thinking and scientific reasoning.

**Dr. Xiangen Hu** is a professor in the Department of Psychology, Department of Electrical and Computer Engineering and Computer Science Department at The University of Memphis (UofM) and senior researcher at the Institute for Intelligent Systems (IIS) at the UofM and is professor and Dean of the School of Psychology at Central China Normal University (CCNU). Dr. Hu received his MS in applied mathematics from Huazhong University of Science and Technology, MA in social sciences and Ph.D. in Cognitive Sciences from the University of California, Irvine. Dr. Hu is the Director of Advanced Distributed Learning (ADL) Partnership Laboratory at the UofM, and is a senior researcher in the Chinese Ministry of Education's Key Laboratory of Adolescent Cyberpsychology and Behavior.

Dr. Hu's primary research areas include Mathematical Psychology, Research Design and Statistics, and Cognitive Psychology. More specific research interests include General Processing Tree (GPT) models, categorical data analysis, knowledge representation, computerized tutoring, and advanced distributed learning. Dr. Hu has received funding for the above research from the US National Science Foundation (NSF), US Institute of Education Sciences (IES), ADL of the US Department of Defense (DoD), US Army Medical Research Acquisition Activity (USAMRAA), US Army Research Laboratory (ARL), US Office of Naval Research (ONR), UofM, and CCNU.



# Flexible and Live Adaptive Training Tools (FLATT) – Updates and Improvements

Michael Cambata, Mitchell Gorsd, Thomas Lenz, and Nicholas Roberts  
Dignitas Technologies

## INTRODUCTION

---

Intelligent Tutoring Systems (ITSs) and adaptive training solutions promise the ability to provide adaptive computerized or virtual environment training focused on the specific needs of the trainee and their performance. To achieve adaptive training, software changes are often required to adjust the adaptations or the criteria under which adaptations are applied, which can be a lengthy process on the order of weeks or months.

Software exists to read data from ITSs and evaluate user performance. However, the ability to create general-purpose adaptations is limited, and existing solutions can take time to properly implement. The Flexible and Live Adaptive Training Tools (FLATT) was created to resolve this issue by creating a suite of tools to improve ease of use when creating adaptive training solutions for research or training purposes. It allows for streamlined visualization of a training course's rules and allows researchers or instructors to adapt the course in real-time based on preconfigured rules or dynamic operator-driven changes. Trainers and researchers can use this data to increase the effectiveness of training solutions or research situations, by characterizing a participant's state or performance level, and then adapting the virtual training situation.

FLATT connects multiple systems to provide the capability to adapt virtual training exercises in real time based on the trainee's state and actions. These systems include the Generalized Intelligent Framework for Tutoring (GIFT), two virtual training environments (VTEs): Virtual Battle Space (VBS) and the Rapid Integration and Development Environment (RIDE), physiological sensors, and FLATT's own trainee state customizations (TSCs), visual rule authoring user interfaces, and rules engines. A complete prototype was developed, demonstrating multiple complex rules and activities in a single training course (Dignitas Technologies LLC, 2022).

FLATT allows instructors, researchers, and course authors to adapt VTEs in real-time based on trainee behaviors and states. This is achieved via a straightforward user interface (UI) that can be used to create rules for how the VTE responds. The authoring UI provides a simple visual workspace to connect complementary inputs and outputs. To be consistent with GIFT, the FLATT authoring UI shares similarities to GIFT's course authoring flow interface. The FLATT rule authoring capability validates complementary events, states, operators, and sensors and prevents incompatible types from connecting. Built-in validation reduces authoring errors and facilitates ease-of-use.

FLATT integrates with the GIFT ITS, which provides automated assessments of trainee performance to drive VTE customizations. But FLATT's architecture is both ITS- and VTE-agnostic and can be extended to operate with other ITSs in the future.

FLATT's rules have two components: Triggers and Customizations. Triggers specify when and customizations specify how to adapt a VTE scenario. The trigger uses data from sensors, VTE performance, learner state, and/or historical records; the customizations are defined by TSCs, which vary by VTE but include adaptations such as trainee feedback, VTE difficulty modifications, and environmental effects (Dignitas Technologies LLC, 2022).

The primary VTE integration was VBS3, largely for its wide use in Army training as well as research contexts. VBS makes it easy to capture data; the available API (Application Programming Interface) supports real-time adaptations. Similarly, a modified version of the Institute for Creative Technology (ICT)'s RIDE software (Rapid Integration & Development Environment, 2023) was selected as a supplemental VTE and has been integrated with FLATT. This version of RIDE is available for download from the GIFT Portal (Downloads – GIFT, 2022).

FLATT provides the capability to:

- Collect data from numerous sources to assess trainee state
- Observe trainee or participants' biometrics during a live session to monitor performance
- Author rules, triggers, and customizations via an intuitive graphical user interface
- Provide TSCs automatically via authored rules
- Modify rules during course execution in real-time
- Provide unscripted feedback/TSC in real-time
- Generate an After-Action Review (AAR) of trainee or participant performance

FLATT also provides GIFT with features as a service, such as real-time rule tracking status visualization and sensor data graphs. These are useful in monitoring the user and scenario state. GIFT provides pre-existing connection, communication infrastructure, and assessment capabilities with the VTEs.

FLATT provides several benefits, including a flexible user experience authoring tool built on a Service Oriented Architecture (SOA) that enables easy integration with different sensors and training solutions to provide point of need adaptations. Researchers will be able to develop new models of learners and determine effective training adaptations that are appropriate for various learner states. Trainers will be able to increase the effectiveness of training solutions by characterizing a learner's state and then adapting or customizing the training path of each learner or participant (Dignitas Technologies LLC, 2022).

Several proof-of-concept demonstrations were built and include both mounted and dismounted use cases. Mounted use cases involve driving tasks, whereas the dismounted tasks include battle drill 2A, room clearance, as well as a sample scenario with multiple user-driven activities. In the context of dismounted scenarios, the VTE is modified based on how participants respond to surveys (BD2A with an empty building vs. with several Opposing Forces [OPFOR]). In addition, the open-world Rahmadi terrain (a VBS-pre-baked terrain) is used to show features such as breadcrumbs, spawn entities, change of day, and change of weather VTE adaptations. In each case, learner state or performance causes the environment to update (Dignitas Technologies LLC, 2022).

## **FLATT SYSTEM ARCHITECTURE**

---

FLATT operates as a service that is designed to interface with an ITS and extend its capability. While FLATT's architecture is designed to be modular and adaptable to multiple ITSs, its current prototype implementation is integrated with GIFT, in order to make use of GIFT's extensive support and existing functionality (Figure 1) (Dignitas Technologies LLC, 2022).

Information from the ITS, such as assessments, sensor data, and learner states, is sent to FLATT and evaluated via FLATT's rules engine, which can produce a trainee state-driven customization (TSC) to be sent back to the ITS. Through this process, FLATT's rules offer additional functionality on top of the ITS's existing capability.

In order to communicate what rule elements are available in a simulation, the ITS must produce a service definition file, which FLATT uses to ensure compatibility. A service definition is created for every ruleset in FLATT (Dignitas Technologies LLC, 2022).

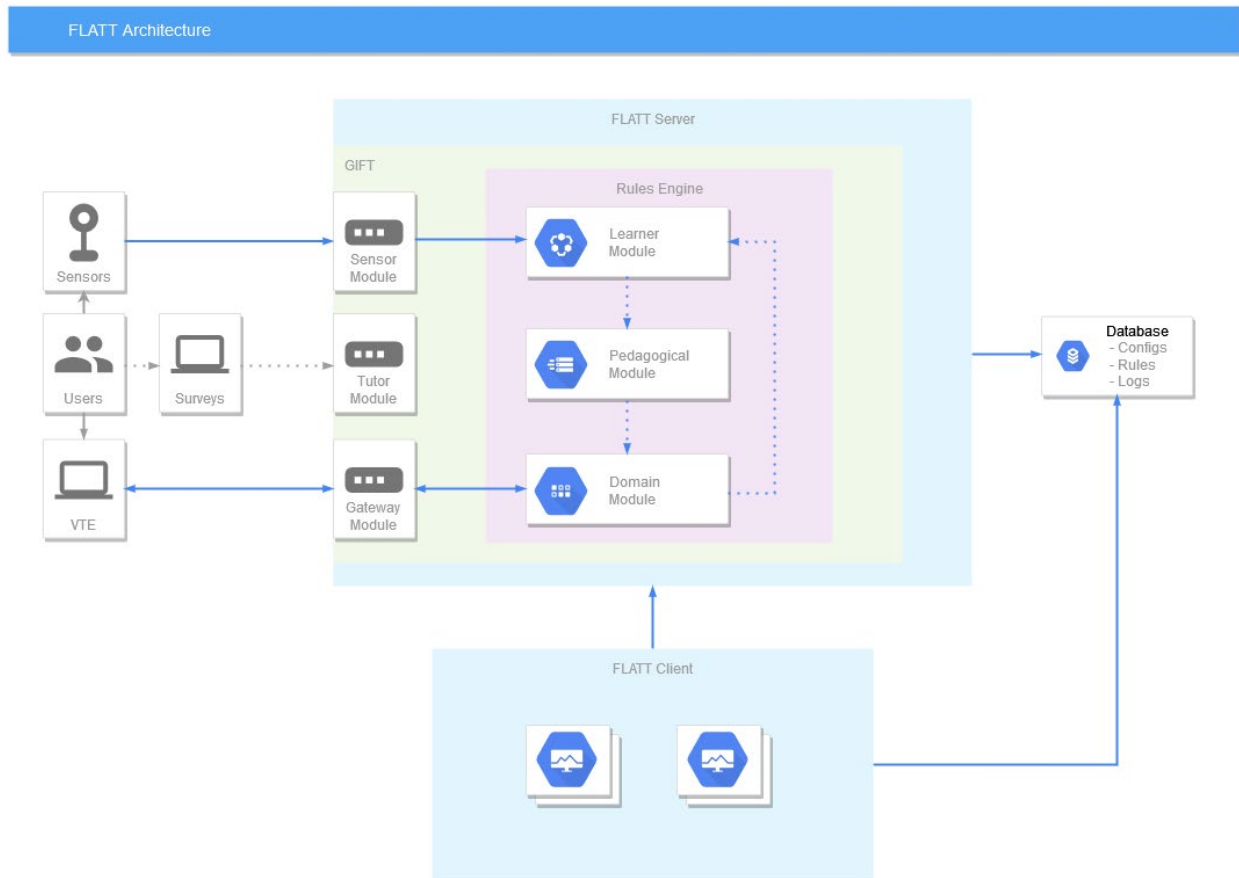


Figure 7. FLATT high level system diagram (Dignitas Technologies LLC, 2022)

## FLATT SYSTEM DESIGN

FLATT was built to be a modular system of dynamic pieces, so each piece can be reused to simplify the user experience. Its rule engine relies on the following interconnected elements to offer full functionality (Dignitas Technologies LLC, 2022).

**Rule** – This is the core of FLATT’s rule engine. Each rule is comprised of a trigger and a customization, which specify when a rule should be triggered, and what customization occurs once those conditions are met. For example, a rule might be *triggered* when a learner has engaged an enemy in combat, which could cause a *customization* to highlight the enemy’s position.

**Rulette** – A rulette is a part of a trigger which can be reused by multiple rules. This is designed to allow rule authors to create complex or detailed behavior as part of a trigger and re-use it in multiple rules. For example, if multiple rule triggers check that a learner’s heart rate is between two values, while also

checking other conditions, a rulette can contain the behavior which checks the heart rate range and can then be easily added to each rule trigger that requires it.

**Trigger** – A trigger is the first half of a rule. Its elements are connected together via logical operators, and ultimately lead to a customization which is sent if the trigger’s final value evaluates as true. In this way, the trigger determines when a customization is sent.

**Customization** – A customization is the second half of a rule. It contains a number of actions, which are sent back to the ITS when a rule’s trigger is evaluated as true. Customizations can also include timed delays between each action. In this way, the customization determines what a rule does when triggered.

**Ruleset** – A ruleset is a collection of related rules and rulettes. Each ruleset requires a service definition file, which determines the possible elements used by triggers and customizations within that ruleset.

**Service Definition** – A service definition is a file which informs a FLATT ruleset what information its rules and rulettes have access to, including the trigger’s events, states, and sensors, and the customization’s actions. In the current prototype, GIFT generates a service definition based on a given course and sends that service definition to FLATT. FLATT also has the functionality to create a ruleset based on a pre-existing service definition.

## **FLATT AUTHORIZING TOOL**

---

Building triggers, customizations, and rulettes is done withing the FLATT Authoring Tool. It uses a robust if/then system to build a flow of logic from a rule trigger to a customization. Formatted in a graph-like context, each action or condition involved in a course is represented by a node on the graph (Figures 2 and 3). Each item can be dragged from a panel on the left onto the chart area and can be quickly connected to other nodes to design a rule.

Because of the many available options, FLATT has features to assist users and mitigate confusion. Each input and output node has a distinct shape, color, and symbol to indicate which nodes it can connect to. And when the user clicks and drags from one such node, invalid points of connection become increasingly transparent, to highlight only the valid connecting nodes (Dignitas Technologies LLC, 2022).



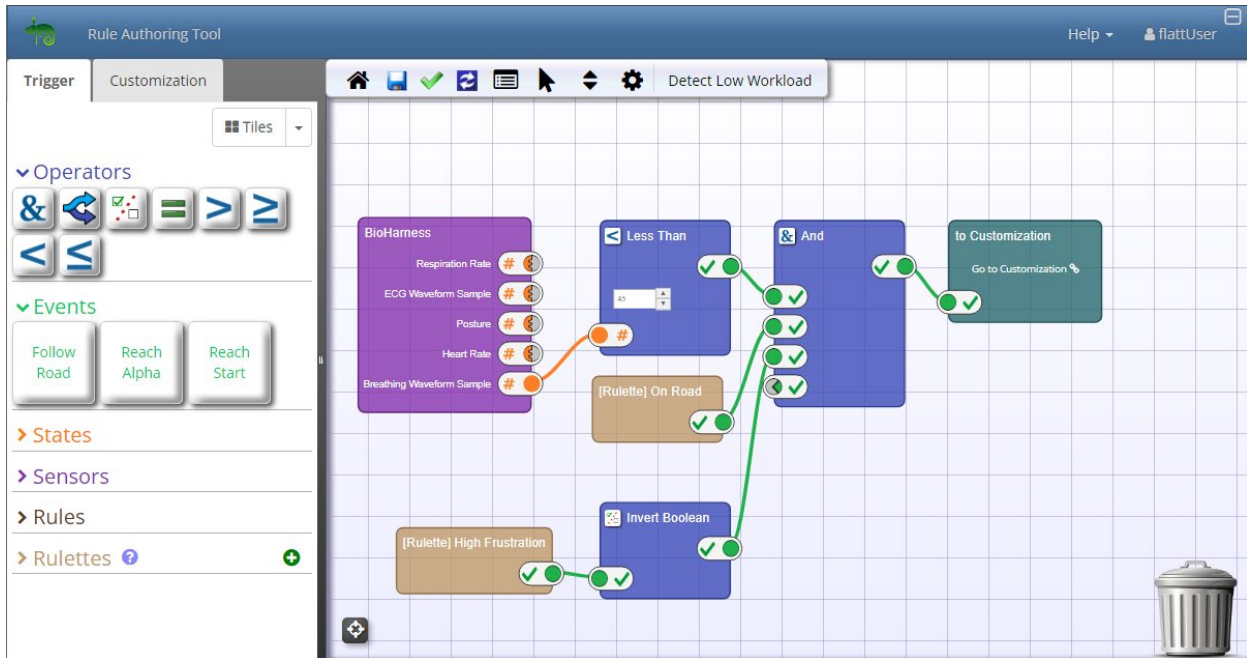


Figure 8. Chart Objects in Rule Trigger (Reprinted with publisher permission from Harrison et al., 2022)

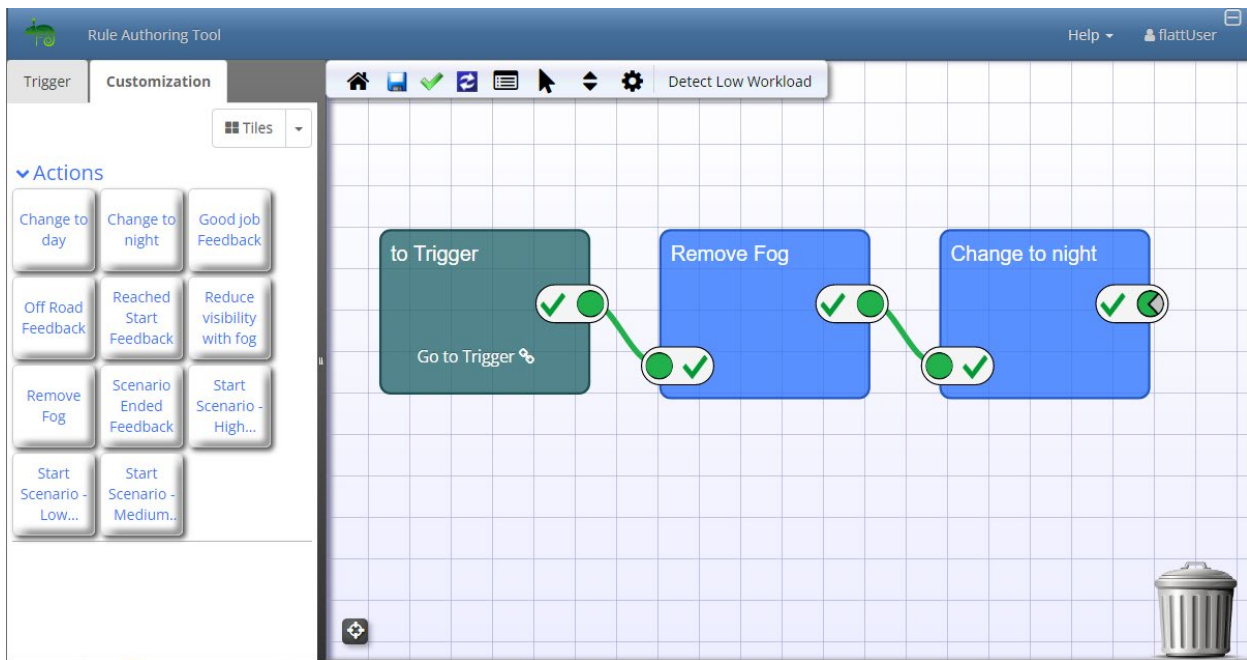


Figure 9. Chart Objects in Rule Customization (Reprinted with publisher permission from Harrison et al., 2022)

The FLATT authoring tool is designed for simplicity and ease of use. The rules chart allows complex behaviors to be laid out in an intuitive way and provide access to all available data sources at a glance.

The FLATT authoring tool also contains a validation feature, which automatically warns the user of invalid rule charts whenever they save their work, or when prompted by pressing a button. This informs the user what issue caused any validation errors, and how to work toward resolving it (Harrison et al., 2022).

## **INTEGRATION BETWEEN FLATT AND GIFT**

---

FLATT integrates with the GIFT ITS, which provides automated assessments of trainee performance to drive VTE customizations. Though FLATT's architecture is both ITS- and VTE-agnostic and can be extended to operate with other ITSs in the future, it already offers a range of improvements to GIFT (Figure 4) (Harrison et al., 2022).

FLATT's Rule Authoring Tool provides an intuitive tool that can be used alongside GIFT's others, to create more complex rules than GIFT might ordinarily allow, or to create them more quickly. Its logical operators offer flexibility, and in a format that can be visualized and understood more easily than GIFT's current interface, which can involve extensive cross-referencing across multiple tabs.

In addition to FLATT's rules engine, the system also provides GIFT with certain features as a service, including a real time rule tracking status visualization and sensor data graphs (Harrison et al., 2022). These are useful in monitoring the user and scenario state. GIFT provides pre-existing connection, communication infrastructure, and assessment capabilities with the VTEs. Using rule tracking and state visualization, researchers and trainers can watch a user's performance in real time. FLATT also provides a Game Master page to display biometric sensor data, including lines that indicate named thresholds, to better visualize numerical value ranges that might be referred to by words like "high" or "low" in FLATT's rules.

FLATT also allows the user to alter rules in real-time through GIFT's Game Master. While a session is running, the user can open an existing rule and lock connecting nodes to "True" or "False," *or* they can alter the rules chart directly, saving the result after the fact. This lets users adjust scenarios while courses are running, and can be used to fine-tune behavior or manually alter logic to apply to a changing situation which is not necessarily represented in the pre-written course. These changes are recorded in the session's log, so that FLATT's After Action Report tool will display when any modifications to existing rules were made during a live session (Dignitas Technologies LLC, 2022).

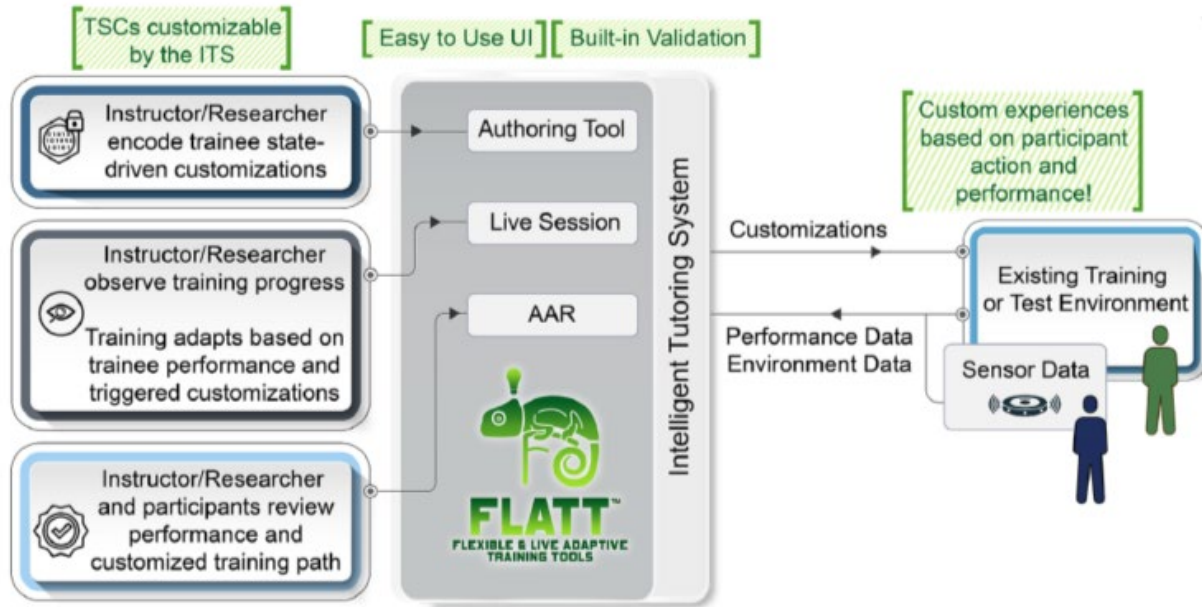


Figure 10. FLATT Capabilities Overview (Dignitas Technologies LLC, 2022)

## AFTER ACTION REVIEW

FLATT gathers data from the trainee, the course itself, and the proctor of the session and uses that data to create a series of display services for After-Action Review (AAR). The logged data includes biometric sensor data, a graphical representation of the rules used in the course, a count of how many times each TSC has been triggered, and any intervention an observer may have performed during the course from the Game Master. The AARs are shown in a linear top-down timeline dashboard, and have several filters presented on a filter panel to organize them based on data, time, and other aspects. All the data is presented in easily digestible methods. The rules are presented in a static rules chart, sensor data is presented in a complete line graph, and the interventions and trigger lists are presented in a simple list format. From the dashboard the user can download a copy of the service definition used, and a data file containing all the AAR data, which can be used for analysis or archival purposes (Dignitas Technologies LLC, 2022).

## EXEMPLAR CAPABILITIES

Multiple proof-of-concept demonstrations were created with FLATT, including both mounted and dismounted use cases.

In one course, the user undergoes a Battle Drill 2A room-clearing scenario in VBS3, with the presence or absence of opposing forces based on participant responses to a survey.

In another set of courses, the user is tasked with driving a vehicle along a road in VBS3, with environmental adaptations being triggered in real-time in response to multiple data points, including biometric values, survey responses, and whether or not the user's vehicle is on- or off-road.

FLATT's most complex demonstration is its final exemplar, which is designed to showcase its capabilities in a variety of different activities. This demonstrated integration with GIFT, FLATT, and VBS3, and contains several "zones," each of which supports a different activity.

The Target Zone contains a target range, and gives customized feedback based on which targets the user hits within a pre-defined time limit. The Breadcrumbs Zone prompts the user to approach an objective via a specific pre-defined path, and displays breadcrumb markers to guide them. If they follow the proper path, they are given feedback noting their success, but if they do not follow the proper path, they are ambushed by opposing forces and given feedback noting their failure. The Weather Zone and Time of Day Zone, contain signs that can be approached to change the simulation's weather and time of day, respectively. In addition, there are two combat encounters, which use FLATT's rule engine to highlight the position of opposing forces when combat begins, and remove the highlights after all OPFOR in the encounter are injured. In a pre-course survey, the user also has the option to answer survey questions and connect the weather and time of day to biometric data, to make things easier or harder to see if the user's biometrics fall within certain values (Dignitas Technologies LLC, 2022).

The exemplar demonstrations showcase various scenarios that FLATT's rules engine is able to support, with the final exemplar in particular demonstrating the complexity that FLATT can allow.

## **BENEFITS OF FLATT**

---

FLATT's prototype offers a set of useful capabilities which expand GIFT's own, and a modular architecture that would allow it to interface with other ITSS in the future. Its graphical, chart-based approach to rules make it easier for researchers or trainers to create and understand scenarios, and its ability to modify rules in real-time streamlines the process of testing and correcting the intended behavior of those scenarios.

## **CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

---

In the future, solutions such as FLATT should be leveraged for research and training contexts. By implementing tools, such as FLATT, researchers will benefit from decreased time and effort to implement unique scenarios for trials and instructors will be better able to tailor virtual training to a specific trainee. GIFT, with its robust set of tools, could benefit from implementing similar strategies to offer course authors and researchers additional easy-to-use tools.

In addition, additional research and development may be applied to this area to identify additional virtual training environments that may benefit from such a capability. Future Department of Defense (DoD) training capabilities would benefit from easy-to-author capabilities by reducing the need for system-specific experts or software engineers to modify training scenarios in an ad hoc fashion. As cross-service training events become more prevalent in a more connected world, leveraging tools with easy-to-use authoring and automated validation could save significant labor in the future.

## **REFERENCES**

---

Harrison, S., Gorsd, M., Hoffman, M., Pollard, K., & Files, B. (2022). *FLATT: A New Real-Time Assessment Engine Powered by GIFT (2022 Edition)*. In Proceedings of the Tenth Annual GIFT Users Symposium (GIFTSym10).

Dignitas Technologies LLC (2022). *Flexible and Live Adaptive Training Tools Phase II SBIR FINAL*

Downloads - GIFT. (2022). GIFT Portal. <https://gifttutoring.org/projects/gift/files>

Rapid Integration & Development Environment. (2023). RIDE; USC Institute for Creative Technologies.  
<https://ride.ict.usc.edu/>

## ABOUT THE AUTHORS

---

***Michael Cambata** is a senior software engineer at Dignitas Technologies. Michael has worked at Dignitas Technologies for nearly 7 years, and has been on the GIFT team since 2020. In 2022, he became the engineering lead for FLATT Phase 2.*

***Mitchell Gorsd** is a developer at Dignitas Technologies hired in 2021 and brought directly onto FLATT, currently working on AFITT, who worked on FLATT for phase 2 and helped write several papers about it.*

***Nicholas Roberts** is a senior software engineer at Dignitas Technologies and the engineering lead for the GIFT project. Nick has been involved in the engineering of GIFT and supported collaboration and research with the intelligent tutoring system (ITS) community for nearly 10 years. Nicholas contributes to the GIFT community by maintaining the GIFT portal ([www.GIFTTutoring.org](http://www.GIFTTutoring.org)) and GIFT Cloud ([cloud.gifttutoring.org](http://cloud.gifttutoring.org)), supporting conferences such as the GIFT Symposium, and participating in technical exchanges with Soldier Center and their contractors. He has also been heavily involved in integrating GIFT into TSS/TMT.*

***Thomas Lenz** is a senior systems engineer and GIFT principal investigator at Dignitas Technologies. Thomas has been involved in various aspects of GIFT's development and testing for several years, recently shifting into a leadership role within Dignitas Technologies. His background includes over 10 years of instructional systems design for live and virtual training systems and systems engineering for Army and Navy use cases.*





**THEME V:  
COLLECTIVE/TEAM BASED  
METHODS**





# Improving Dialogue Classification Models to Support Generalizable Team Communication Analytics in GIFT

Jay Pande<sup>1</sup>, Stephen Paul<sup>1</sup>, Wookhee Min<sup>1</sup>, Randall Spain<sup>2</sup>, and James Lester<sup>1</sup>

<sup>1</sup>North Carolina State University; <sup>2</sup>U.S. Army Combat Capabilities Development Command (DEVCOM) - Soldier Center

## INTRODUCTION

---

Spoken communication between team members can provide deep insights into how a team cooperates and coordinates actions during periods of high and low demand (Meslec et al., 2020). Advances in natural language processing (NLP) are offering new capabilities to support automated analysis of team communication (Jensen et al., 2021; Min et al., 2021; Pande et al., in press). When these capabilities are integrated into synthetic training environments (STEs), instructors and units can obtain timely insights on squad performance through automatic assessment of team communication behaviors. Over the past three years, North Carolina State University and the U.S. Army Combat Capabilities Development Command - Soldier Center, Simulation and Training Technology Center have been collaboratively designing and developing an NLP-driven framework, the Team Communication Analysis Toolkit (TCAT). TCAT analyzes team communication data by assigning descriptive labels to each utterance and computing statistics that summarize the team’s communication patterns. These statistics may uncover important patterns of communication between team members that lead to high or low team performance. TCAT uses the Microsoft Azure speech-to-text cloud service to automatically produce a transcript of team members’ spoken communication. This transcript is passed to our NLP framework that analyzes the dialogue and automatically assigns labels to each utterance to indicate *dialogue act*, the intent of the speech (e.g., PROVIDE INFORMATION, ACKNOWLEDGEMENT, COMMAND), and *information flow*, how information was passed between different levels of the team’s hierarchy (e.g., PROVIDE INFORMATION DOWN when the squad or team leader shares information with team members, REQUEST INFORMATION UP when team members ask for information from the squad or team leader). The framework’s labeling output is available to TCAT’s data analysis interface, which one can use to compute the statistics, and its data visualization interface, which transforms these statistics into informative graphics (e.g., charts of the frequency of each label per speaker). The driving objective of TCAT is to integrate NLP-driven insights about team communication into the Generalized Intelligent Framework for Tutoring (GIFT). This integration will enable more robust adaptive coaching, scaffolding, and assessment in GIFT.

We begin this paper with a description of our latest refinements to TCAT’s NLP framework (Figure 1) to provide more accurate dialogue act recognition and information flow classification. Specifically, we discuss the latest iteration of TCAT’s NLP models (Pande et al., in press), which are based on the T5 (Text-To-Text Transfer Transformer) deep neural architecture (Raffel et al., 2020). We compare the classification results to our previous models (conditional random fields utilizing ELMo embeddings and bidirectional long short-term memory networks) (Min et al., 2021) on the Squad Overmatch dataset (Johnston et al., 2019), highlighting how the T5-based models achieved state-of-the-art results for both dialogue act recognition and information flow classification in the evaluations conducted with cross-validation, a held-out test set, and a related dataset (i.e., domain-transfer task). We also discuss the strengths of the T5 modeling approach.

In addition, we discuss results of a multi-class confusion matrix-based error analysis we completed that identified strengths and weaknesses of the T5 framework’s labeling performance. The analysis highlighted

several labels in dialogue act recognition and information flow classification that the framework often correctly applied to utterances, but also identified labels which the framework incorrectly inferred.

We also describe promising future directions of our current work centering on devising more reliable team communication analysis models based on error analysis results, improving generalization performance of the T5 team communication analysis framework, and exploring how to utilize dialogue analysis results for assessing team performance. Robust team communication analytics will enable researchers and teams to better understand patterns of communication across datasets and deepen our understanding of how team communication relates to team performance, as we work towards creating effective adaptive team training environments. We conclude with a discussion of how TCAT will be integrated into GIFT to provide insights into team communication and coordination behaviors in other domains such as Crew Gunnery and Squad Battle Drills.

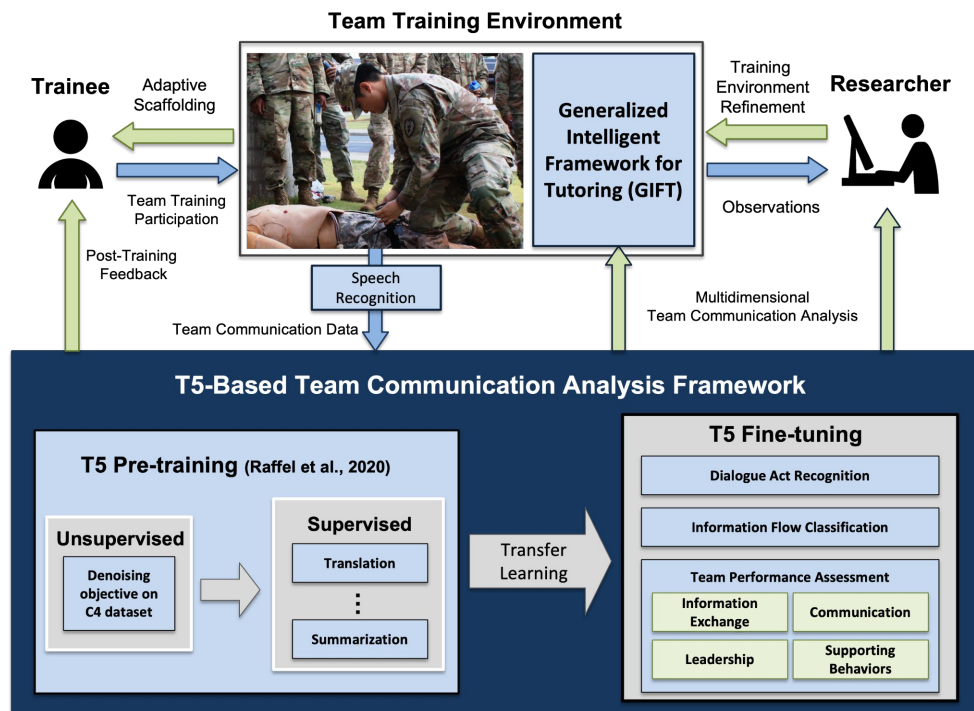


Figure 1. Team Communication Analysis Framework for Creating Adaptive Team Training Environments (Adapted from Pande et al., In Press)

## BACKGROUND

Investigating the spoken communication of team members is an effective way to understand the processes that lead to successful team performance (Marlow et al., 2018). Advances in deep learning-based automatic speech recognition (ASR) and NLP have made it possible to automate transcript generation and perform automated syntactic and semantic analyses of dialogue from the generated transcripts (Nassif et al., 2019, Otter et al., 2020). TCAT (Min et al., 2021, Pande et al., in press) leverages these advances to support natural language analysis of spoken dialogue between team members. The NLP pipeline for TCAT includes two major components: ASR and dialogue act recognition. The current implementation of TCAT uses the Speech-to-Text cloud service from Microsoft Azure for ASR to convert spoken team communication into text for downstream NLP tasks including speaker utterance-level dialogue act recognition. To perform the downstream classification tasks, TCAT employs a hybrid NLP framework that merges the structured prediction capabilities of conditional random fields with contextual sentence embeddings generated by deep

neural networks to classify team communication utterances into nine dialogue acts (Min et al., 2021). Training of the dialogue act recognition model within this framework used labeled transcripts from the Squad Overmatch Mission 3 dataset, which included 6,181 utterances captured from six squads during a 45-minute live training exercise. TCAT provides users with meaningful statistics and data visualization techniques to support team communication analytics.

Since our initial work on TCAT, there have been many advances in NLP-based architectures that have achieved considerably improved performance on a broad range of NLP tasks. Variants of the Transformer architecture introduced by Vaswani and colleagues (2017), such as Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) and Generative Pre-trained Transformer (GPT) (Brown et al., 2020) have become some of the most popular architectures in NLP. When compared to previous deep learning approaches such as recurrent neural networks and convolutional neural networks, the key innovation of the Transformer architecture is the self-attention mechanism. Self-attention enables the model to selectively focus on different parts of the input sequence, based on the relevance of each part to the current processing step. The Transformer architecture includes both an encoder and a decoder, and each of these is made up of multiple layers of self-attention and feedforward neural networks. The encoder's responsibility is to process the sequence of input tokens, and it generates a sequence of hidden states. Next, the decoder generates a token for each position in the output sequence based on both the hidden states and tokens at earlier positions in the sequence. The Transformer architecture can effectively model long sequences by preserving long-range dependencies with the self-attention mechanism, whereas recurrent neural networks and their variants (e.g., LSTMs, GRUs) process each input token sequentially, often forgetting characteristics captured early in the sequence by the time tokens near the end of the sequence are processed. Transformer-based models have exhibited state-of-the-art results on several benchmark datasets for tasks such as machine translation, question answering, and language modeling (Maslej et al., 2023).

The T5 architecture we present in this paper confers several benefits to team communication analysis. First, the text-to-text architecture of T5 allows end-to-end training and inference, whereas our previous models had an underlying multi-step process involving generating sentence embeddings and conducting principal component analysis for dimensionality reduction. Second, the T5 architecture can perform multiple dialogue act labeling tasks simultaneously, which can considerably reduce memory and computation time requirements without sacrificing performance. Third, because the T5-based models were pre-trained with large-scale external language resources, they hold significant potential to generalize well to new datasets by using learned representations of common statistical patterns in language. Fourth, because the fine-tuning process incorporates “task-specific prefixes”, which are text-based descriptors of NLP tasks that precede the main content of the input (Raffel et al., 2020), T5 can infer characteristics of tasks directly from the phrasing of the prompts to effectively train models to perform these tasks.

## **SQUAD OVERMATCH DATASET**

---

In developing and evaluating our new T5-based framework, we utilized a set of coded transcripts from the Squad Overmatch project (Johnston et al., 2019). The transcripts contain utterances of squad members as they completed a live training mission. The data was annotated using a coding scheme consisting of 27 dialogue act labels and 18 information flow labels. The dialogue act labels capture what a speaker intends to communicate in an utterance, such as information about what they see, a suggestion about what to do next, or a question for another squad member. The information flow labels capture how information is exchanged within the chain of command by describing the relationship between the speaker and the target of the utterance. Each utterance was labeled with its corresponding dialogue act label, but utterances which did not convey information from one part of the hierarchy to another did not receive an information flow label. Statistics for interrater reliability for both tasks are strong. For dialogue act labels, Krippendorff's  $\alpha = 0.679$ , raw agreement = 0.715, expected agreement = 0.113. For information flow labels, they are

Krippendorff’s alpha = 0.710, raw agreement = 0.817, expected agreement = 0.335 (D. Traum, personal communication, July 25, 2022).

In our previous work (Min et al., 2021), we streamlined the dialogue act labels into a smaller set where each label was applicable to a broader category of utterances. Through this, we hoped to make the labels easier for stakeholders to understand and easier for our models to correctly infer. The reduced set of labels, which were used for the work discussed in this paper, are ACKNOWLEDGEMENT, ACTION REQUEST, ACTION STATEMENT, COMMAND, ATTENTION, GREETING, PROVIDE INFORMATION, REQUEST INFORMATION, and OTHER. For information flow classification, we continued to use the original set of 18 labels (examples include COMMAND COMING FROM THE SQUAD LEADER, PROVIDE INFORMATION UP THE CHAIN OF COMMAND, and REQUEST INFORMATION FROM DOWN THE CHAIN OF COMMAND; Min et al., 2021). Table 1 shows a few utterances from our dataset and the associated labels for each. We also included an N/A label for information flow classification so that we could still classify utterances which are not covered by the other information flow labels.

**Table 1. Example Utterances and Their Dialogue Act (DA) and Information Flow (IF) Labels**

Speaker	Example Utterances	DA	IF
Squad Leader	Two alpha two say again description of the man not recognized over	ACTION REQUEST	REQUEST INFORMATION FROM DOWN THE CHAIN OF COMMAND
Alpha Team Leader	Two alpha moving	PROVIDE INFORMATION	PROVIDE INFORMATION UP THE CHAIN OF COMMAND
Bravo Team Leader	You see anything let me know	COMMAND	COMMAND COMING FROM THE TEAM LEADER
Squad Leader	Hey we're going to move to the church	COMMAND	COMMAND COMING FROM THE SQUAD LEADER

## T5-BASED NLP FRAMEWORK

As a transformer model, T5 leverages the benefits of the self-attention mechanism described above. An additional strength is that it has undergone multiple stages of training: first, unsupervised learning involving predicting masked tokens (i.e. words or parts of words) in sequences from the Colossal Clean Crawled Corpus (C4) dataset, and second, supervised learning using several NLP benchmark datasets (Raffel et al., 2020). The creators of T5 have prepared multiple models of varying sizes, each of which has been trained in this way. The “small” model is most appropriate for our use case since it has only 60 million parameters (note that this model remains reasonably expressive, with a 32,128-token vocabulary size), and thus can still be trained and used for inference when computational resources are limited.

Because T5 was designed to perform many different NLP tasks, its text-based input format is very flexible. Our encoding of Squad Overmatch utterances leverages this flexibility by putting the speaker role in place of the “task-specific prefix” (Raffel et al., 2020), and the speaker utterance following the colon. An example formatted utterance from the squad leader is *sql: Hey what side of the building is that?* Before fine-tuning with the Squad Overmatch data, we first loaded two pre-trained “small” T5 models with the default architecture. Then, using the formatted Squad Overmatch utterances, one of the models was fine-tuned for dialogue act recognition and the other for information flow classification.

## T5-Based-Classification Results in TCAT

Over the past several months, we have been investigating the classification performance of TCAT’s T5-based NLP architecture. As reported in Pande et al. (in press), we created separate T5 models (i.e., two single-task models) for dialogue act recognition and information flow classification, performed five-fold cross-validation at the squad level on each model to identify the best-performing model hyperparameters for our framework, and examined its generalization performance on a held-out test set. The data used for cross-validation consisted of utterances from five squads (the sixth squad from the same dataset was held out for testing after cross validation). It should be noted that the dataset used for the cross-validation and held-out test evaluations is identical across different models (T5, CRF-ELMo, BiLSTM) for a fair comparison. Each squad served as the test set for one fold, and the remaining four squads were the training set for that fold. The cross-validation results demonstrated that, on both dialogue act recognition and the information flow classification, the NLP-based team communication analysis framework outperforms the majority-class baselines (25.7% and 45.8%, respectively). For dialogue act recognition, the T5 framework achieved an average accuracy of 73.55%. For information flow classification, it achieved an average accuracy of 65.72% (Pande et al., in press).

Following cross-validation, we further explored the framework’s inference capabilities using a held-out test set (the data in this set was not used during cross-validation and thus did not influence the selection of hyperparameter values). As reported in Pande et al. (in press), the dialogue act recognition model achieved 75.92% accuracy on the held-out test set, and the information flow classification model achieved 67.23% accuracy (Table 2). Further analyses showed the framework demonstrated modest improvements in domain transfer performance compared to previous NLP approaches (Pande et al., in press). In sum, results of our recent work demonstrate that the T5 models outperformed the CRF-ELMo models that were previously investigated in Min et al., (2021) in cross validation, inference using a held-out test set, and on a domain transfer task (Pande et al., in press).

**Table 2. Evaluation Results (Accuracy Rates) of Cross-Validation, Held-Out Test Set, and Domain Transfer on Dialogue Act (DA) Recognition and Information Flow (IF) Classification. Highest Accuracy for Each Evaluation Task Is Marked in Bold.**

	CRF-ELMo (Min et al., 2021)		BiLSTM (Min et al., 2021)		T5 (Pande et al., in press)	
	DA	IF	DA	IF	DA	IF
Cross-validation	68.88	58.84	62.07	55.79	<b>73.55</b>	<b>65.72</b>
Held-out test set	69.42	64.92	64.61	61.88	<b>75.92</b>	<b>67.23</b>
Domain transfer	67.35				<b>70.60</b>	

This is evidence that our updated team communication analysis framework, which leverages T5’s natural language understanding capabilities and familiarity with a diverse array of NLP tasks, is less likely to mislabel utterances than prior approaches and thus more trustworthy for stakeholders relying on the framework for insights into team communication. It also shows that this framework can generalize appropriately to new datasets that share some characteristics from the training data.

## Error Analysis Results

In addition to examining the overall performance of the framework, we have also investigated its performance on individual labels. We utilize a multi-class confusion matrix for dialogue act recognition and another for information flow classification. This fine-grained approach yields detailed insight into the strengths and limitations of the model across communication labels. When classifying dialogue acts, the

model showed good performance for ACKNOWLEDGEMENT, COMMAND, and PROVIDE INFORMATION dialogue acts, while it struggled to classify the OTHER dialogue act. OTHER was frequently confused with PROVIDE INFORMATION and COMMAND dialogue acts. When classifying information flow labels, the model exhibited the highest accuracy when correctly identifying utterances of COMMAND MIDDLE. It showed poor accuracy when it frequently misclassified PROVIDE INFORMATION UP instances as PROVIDE INFORMATION DOWN, signaling an error in how it differentiated between the direction of these two labels.

Overall, we identified communication labels, such as ACKNOWLEDGEMENT and COMMAND MIDDLE, that were effectively classified and can be more confidently applied to new and similar datasets. We also detect areas of poor performance that may not have emerged from high-level approaches to classification accuracy, such as misidentifications between the direction of flow of an information flow label and types of dialogue acts that are more frequently confused. This will spur future investigation into the conceptual and operational utility of communication labels used in this context as well as the development of more robust NLP models.

## UPCOMING TCAT DEVELOPMENT PRIORITIES

---

In the upcoming months, our team will focus on integrating several enhancements to TCAT. In particular, we will update the public release version of TCAT to utilize the T5-based framework, replacing the CRF-ELMo framework. The T5-integrated TCAT will allow for more accurate dialogue act predictions, eliminate additional preprocessing steps required in CRF-ELMo, and extend TCAT to additionally support information flow predictions. In addition to dialogue act labels, the inclusion of the information flow labels will further enrich investigation of salient team communication patterns, which could provide additional insights into team coordination activities and performance.

Another development priority is updating TCAT's user interface. At last year's GIFT User Symposium, we provided an overview of the features and research interface (Spain et al., 2022). Users interact with TCAT through a main menu with two tabs. In the File tab, users can import audio and transcript files as well as load NLP models. In the Actions tab, users can generate a transcript from the imported audio files using ASR and use the loaded NLP model to infer dialogue acts of utterances in the imported transcripts. Figure 2 shows the current interface, which includes step-by-step instructions to guide users through the menu options, as well as a mock design for the future interface. We will aim to refine the user interface, menu options, and workflow to make the TCAT software more intuitive for new users.

In addition, the current NLP models in TCAT are specific to the Squad Overmatch data, since the models take as input speaker roles, which are specific to the dataset, as well as utterances. We will extend our models to be more generalizable for the speaker roles using a two-way split (e.g., high and low in the chain of command) or a three-way split (e.g., high, middle, and low). The design of the updated input features for the NLP models will be further explored and evaluated with other datasets including Crew Gunnery and Squad Battle Drills. Finally, we will work towards exploring how TCAT can be integrated with GIFT as an external analysis engine to support team communication analytics.



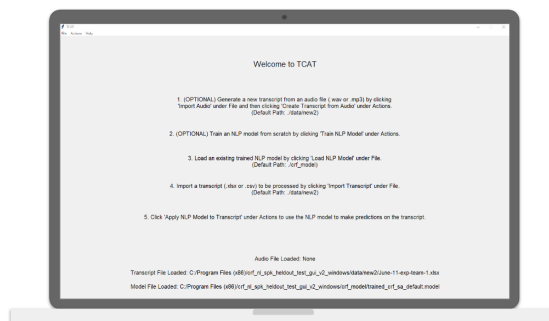


Figure 2. The Menu-based Design of the TCAT User Interface

## CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

---

In our research over the past year, we have implemented and evaluated an updated framework for team communication analysis based on the T5 architecture (Pande et al., in press). The updated framework displays improved performance on dialogue act recognition and information flow classification when evaluated on the Squad Overmatch dataset, compared to the prior CRF-ELMo and bidirectional LSTM approaches (Min et al., 2021). The T5-based framework has greater potential to generalize to new domains and integrate smoothly with GIFT.

As we move forward with our project, we will work to expand TCAT’s capabilities beyond Squad Overmatch data and into other domains such as Crew Gunnery training exercises, as well as support real-time analysis of team communication. We have identified several promising areas for future research, including integrating TCAT with GIFT for end-to-end natural language analysis, which could be achieved by collaborating with the GIFT user community. However, there are technical challenges that need to be addressed, such as optimizing ASR accuracy in a noisy environment, determining how to segment utterances, synchronizing speech and transcript data across multiple speakers, and evaluating how NLP models can be refined to robustly perform team communication analytics for noisy transcripts. Also, it will be important to evaluate the generalization errors of the TCAT NLP models with other communication datasets and improve its generalizability. A promising direction to address these challenges is to investigate alternative cloud-based ASR services and supplementary procedures which improve robustness to noise in audio data, evaluate TCAT’s real-time dialogue classification capacity, and improve generalizability of the NLP architecture. Additionally, it will be important to explore approaches to reliably assessing multidimensional team performance, including information exchange, communication, leadership, and supporting behaviors, based on team communication analysis results (e.g., Spain et al., 2021). Finally, we will refine TCAT’s user interface and data visualization to ensure it contains useful features for team communication analytics, and explore how the software could help with assessing core course concepts and providing feedback to team members through GIFT.

## ACKNOWLEDGEMENTS

---

The research described herein has been sponsored by the U.S. Army DEVCOM, Soldier Center under cooperative agreement W912CG-19-2-0001. 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

---

- Devlin, J. Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics.
- Jensen, E. L., Pugh, S., & K. D'Mello, S. (2021). A deep transfer learning approach to modeling teacher discourse in the classroom. In M. Scheffel, N. Dowell, S. Joksimovic, & G. Siemens (Eds.), *Proceedings of the LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 302–312). Association for Computing Machinery.
- Johnston, J. H., Phillips, H. L., Milham, L. M., Riddle, D. L., Townsend, L. N., DeCostanza, A. H., Patton, D. J., Cox, K. R., & Fitzhugh, S. M. (2019). A team training field research study: Extending a theory of team development. *Frontiers in Psychology, 10*, 1480.
- Marlow, S.L., Lacerenza, C.N., Paoletti, J., Burke, C.S., & Salas, E. (2018). Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes, 144*, 145–170.
- Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). The AI Index 2023 Annual Report. *Institute for Human-Centered AI*.
- Meslec, N., Duel, J., & Soeters, J. (2020). The role of teamwork on team performance in extreme military environments: An empirical study. *Team Performance Management: An International Journal, 3*(5-6), 325–339.
- Min, W., Spain, R., Saville, J. D., Mott, B., Brawner, K., Johnston, J., & Lester, J. (2021). Multidimensional team communication modeling for adaptive team training: A hybrid deep learning and graphical modeling framework. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Proceedings of the 21st International Conference On Artificial Intelligence In Education* (pp. 293–305). Springer, Cham.
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., & Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. *IEEE Access, 7*, 19143-19165.
- Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions On Neural Networks And Learning Systems, 32*(2), 604-624.
- Pande, J., Min, W., Spain, R., Saville, J. D., & Lester, J. (in press). Robust team communication analytics with transformer-based dialogue modeling. *Proceedings of the 23rd International Conference On Artificial Intelligence In Education*.
- Park, K., Sohn, H., Min, W., Mott, B., Glazewski, K., Hmelo-Silver, C., & Lester, J. (2022). Disruptive talk detection in multi-party dialogue within collaborative learning environments with a regularized user-aware network. In O. Lemon, D. Hakkani-Tur, J. J. Li, A. Ashrafzadeh, D. Hernández Garcia, M. Alikhani, D. Vandyke, & O. Dušek (Eds.), *Proceedings of the Twenty-Third Annual Meeting Of The Special Interest Group On Discourse And Dialogue* (pp. 490–499). Association for Computational Linguistics.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research, 21*(1), 5485–5551.
- Spain, R., Min, W., Saville, J. D., Emerson, A., Pande, J., Brawner, K., & Lester, J. C. (2022, May). Leveraging Advances in Natural Language Processing to Support Team Communication Analytics in GIFT. In A. M. Sinatra (Ed.), *Proceedings of the Tenth Annual Gift Users Symposium (GIFTSym10)* (pp. 147–156). US Army Combat Capabilities Development Command–Soldier Center.
- Spain, R., Min, W., Saville, J., Brawner, K., Mott, B., & Lester, J. (2021, May). Automated assessment of teamwork competencies using an evidence-centered design-based natural language processing approach. In B. S.



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

Goldberg (Ed.). *Proceedings of the Ninth Annual GIFT Users Symposium (GIFTSym9)* (pp. 140–149). US Army Combat Capabilities Development Command–Soldier Center.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In U. von Luxburg, I. Guyon, S. Bengio, H. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (pp. 6000–6010). Curran Associates, Inc.

## ABOUT THE AUTHORS

---

**Jay Pande** is a PhD student in Computer Science at North Carolina State University. He received a BS in Computer Science from Duke University in 2020. His research interests lie in the use of speech data to improve educational outcomes for users of adaptive learning environments.

**Stephen Paul** is a Doctoral Candidate student in Industrial/Organizational Psychology at North Carolina State University. He is interested in leveraging technology to improve industrial/organizational psychology research methods and drive real-world impact.

**Dr. Wookhee Min** is a Research Scientist in the Center for Educational Informatics at North Carolina State University. He holds a PhD in Computer Science from North Carolina State University. His research focuses on artificial intelligence centering on user modeling, natural language processing, educational data mining, and multimodal learning analytics.

**Dr. Randall Spain** is a Research Scientist in the Soldier Effectiveness Directorate at U.S. Army DEVCOM, Soldier Center. He holds a PhD in Human Factors Psychology from Old Dominion University. His research focuses on the design and evaluation of advanced training technologies on learning and performance.

**Dr. James Lester** is a Distinguished University 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. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI).



# Measurement Error and the Generalizability of Competency Assessments

Grace Teo, Eric Sikorski, Michael King, and Jennifer Solberg  
Quantum Improvements Consulting

## INTRODUCTION

---

Training in the modern Army cannot merely be a series of discrete activities that check off a set of requirements. Instead, training should be learner-centered, focused on changing the individual and, by extension, transforming teams. To achieve this transformation, we need to understand how training develops individual and team competencies in addition to assessing performance and task results. Competencies are the improvable knowledge, skills, and abilities (KSAs) that an individual possesses and draws upon to perform a task under standardized conditions that resemble real-life situations, assessed against some benchmark of performance (Shavelson, 2010). There are key differences between performance and competencies in terms of understanding and assessing the Soldier. Performance relates more to task execution and outcomes, while competencies are more descriptive of the performer. Performance tends to show what the Soldier can do, while competencies are more revealing of who the Soldier is. Performance assessments are typically task-oriented, while competency assessments are typically Soldier-focused. This competency-based, Soldier-centered approach aligns with the Army's *Human Dimension Concept* (U.S. Army Training and Doctrine Command, Army Capabilities Integration Center, 2016). From a personnel management perspective, there are advantages to including competency and performance assessments. Understanding a Soldier's competencies and their previous performance, training, and experiences can enable the Army to make certain personnel and staffing decisions. These decisions include identifying other tasks and vocations for which the Soldier may be best suited and selecting the most appropriate Soldiers for novel tasks. Competencies and performance assessments help the Army understand its Soldiers and the teams they form.

Designing practical training for Soldiers requires assessments that accurately track changes in competencies over time and across different training exercises. Training systems, such as the Army's Generalized Intelligent Framework for Tutoring (GIFT) (Sottolare et al., 2012) and the Synthetic Training Environment Experiential Learning for Readiness (STEEL-R) framework, aim to do this. These training systems seek to assess student competencies across multiple activities and experiences over multiple occasions. Their central premise is that competency assessment should not be confined to discrete assessment events that only occur at isolated time points but should be conducted consistently to build and update a persistent model of student competencies. However, because competencies are latent constructs that cannot be directly measured but must be inferred from behaviors observed in the assessment, any assessment will always involve some degree of measurement error. This raises two questions: (i) how do we identify and estimate the impact of different sources of measurement error and (ii) how does measurement error affect the generalizability of scores?

## SOURCES OF MEASUREMENT ERROR IN ASSESSMENTS

---

The GIFT approach asserts that the behaviors from various activities and events count toward assessment and should be considered for evidence of competency (Goldberg et al., 2021). The approach acknowledges that any assessment event is merely an instance within a complex "universe" of instances for observing behaviors that can be evidence of competencies (Shavelson, 2010). The sources of measurement error arise because any competency assessment involves sampling from this universe. Shavelson (2010) identifies five types of sampling: (i) criterion sampling, (ii) task-response sampling, (iii) occasion sampling, (iv) rater

facet, and (v) assessment method facet (Shavelson, 2010). We explore these various types of sampling using our work in assessing teamwork competencies as part of the Small Unit Performance Analytics (SUPRA) study as a use case. Early in the SUPRA study, we developed measures of squad team competencies such as leadership, monitoring, communication, coordination, and cooperation (Jeffcott & Mackenzie, 2008) from audio recordings of the squad's communications during training. The SUPRA study design consisted of squads conducting two different battle drills, each performed on three separate occasions. The study yielded multi-task, multi-occasion data, which aligns with GIFT's competency assessment approach.

### ***Criterion Sampling***

The task used in competency assessments should be representative of actual real-life tasks. The possible responses to the assessment task should also map to actual responses in real life (Shavelson, 2010). In SUPRA, the criterion tasks were *Battle Drill 2A* (BD2A: "Conduct a Squad Assault") and *Battle Drill 6* (BD6: "Room Clearing"). These criterion tasks relate to common field and urban operations and are performed in typical infantry squad task environments. The study protocol ensured a level of fidelity between the assessment criterion tasks and real-life operations. For example, in the BD2A, the ratio of opposition forces (OPFOR) to blue or friendly forces (BLUFOR) was maintained throughout the drill to be doctrinally appropriate so the squad would continue to engage just as they would in real life. For both drills, drill environments resembled the actual mission environments, and the mission fragmentary orders (FRAGOs) were crafted to be like those in actual operations such that the possible squad responses within the drill mapped to real life. Even the psychological aspects of the real-life encounter with the OPFOR were preserved. For instance, the OPFOR location was different for each BD2A iteration to maintain the element of surprise. For BD6, the target location within the rooms and the rooms containing targets varied.

### ***Task-response Sampling***

To draw inferences about the Soldier's competency level, behaviors, or responses to tasks indicative of the competency must be identified. For every task within the criterion task, there are infinite possible behaviors or responses, so the behaviors identified to count towards the assessment are only a sample of the universe of conceivable behavioral responses for that task (Shavelson, 2010). Often, the challenge is identifying the omissions of behavior since both the presence and absence of behavioral responses should count towards the assessment. Within each criterion task in SUPRA, there were task-response pairs to which the Observer Controllers (OCs) were attending, comprising a sample of all possible task responses. These task-response pairs can be identified from thorough task analyses of the criterion tasks. For BD6, the sample of task-response pairs includes *not shooting* (good response) or *shooting* (poor response) in the room with the civilian (task). When the squad entered the hallway (task), the OCs were looking to see if they *lingered in the fatal funnel* (poor response) or *promptly took their positions in the room* (good response).

For BD2A, some task-response pairs did not reflect the real-life situation. Simulated rounds were used instead of actual ammunition because human OPFOR were involved. These simulated rounds caused more malfunctions than would be the case with live rounds. Therefore, the proportion of the task-response of having a malfunction occur (task) and *clearing malfunction* (response) could be higher in the SUPRA study compared to that in an actual military operation. However, the higher proportion provided more opportunities to assess the squads' responses during weapon malfunctions. Another exception to having real-life task responses in the assessment were the "resurrections" in BD2A. Each time a BLUFOR Soldier was "shot" by the simulated OPFOR round, they were "resurrected" by the OCs and could rejoin the drill. Although this does not resemble what happens in real life, it was necessary to maintain the appropriate BLUFOR to OPFOR ratio that was doctrinally mandated for the drill. As a result, assessing the squads' responses to a BLUFOR death or injury (task) was impossible, which may be valuable for assessing certain competencies.

**Occasion Sampling**

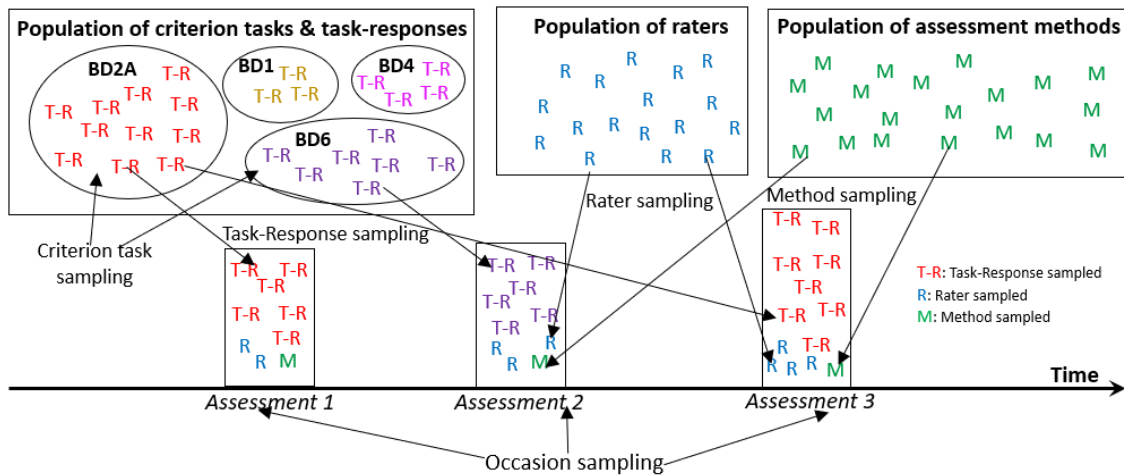
The time in which the assessment event takes place is only a sample of all possible time points or occasions in which assessments can be conducted. All assessments are mere snapshots of the Soldier in time from different perspectives (Shavelson, 2010). This implies that it could be reasonable to link seemingly isolated training events and consider them a series of assessment occasions that can inform the same competencies. In SUPRA, there were multiple occasions to assess the squads’ teamwork competencies. Each squad executed BD2A and BD6 on three consecutive days. BD2A had different OPFOR locations; some were easier for the squad, and others were arguably more difficult. These were counterbalanced, so every squad experienced all difficulty levels of the drill. Some squads conducted BD2A in the rain, which was arguably more difficult. These assessments are not identical but are still samples of the possible occasions to assess teamwork. Analyses showed that although the OCs’ ratings indicated that most squads executed the drills better by their final iteration, some competency measures did not always follow this trend. Practice effects could have contributed to improved performance, as there was little training between the consecutive iterations that could have resulted in improved competencies.

**Raters Facet**

The raters or observers scoring the assessments are a sample from a pool of other possible raters or observers who could have been trained to score the assessment. This facet is particularly influential in assessments that rely heavily on human assessors who should be trained to calibrate their ratings so that the scoring is reliable (Shavelson, 2010). In SUPRA, different sets of OCs scored different squads. The OCs assessing the squads in BD2A were also different from those in BD6. Standardized OC training can help to improve inter-rater reliability and ensure consistency in scoring strategy. The subjective evaluations would then be better calibrated, thereby limiting the raters facet as a source of measurement unreliability.

**Assessment Condition or Methods Facet**

The method used in a particular assessment is only one of many possible assessment methods (Shavelson, 2010). Having multiple possibilities for an assessment method is necessary, especially when resources are constrained or when the best method for assessment may be costly or dangerous. A potential concern with having multiple options for an assessment method is ensuring that the measures from different methods are valid and that the assessments are still tapping the same competency construct. In SUPRA, the squads were assessed in a field exercise with simulated rounds because human OPFOR were used. The study could have been conducted on a training simulator with synthetic teammates and enemies.



**Figure 11. Samplings and facets in assessments (Every assessment conducted on a particular occasion involves task-response pairs from a criterion task, raters, and a method).**

## GENERALIZABILITY THEORY

Given that any assessment is merely a snapshot of the individual or team competencies, how do we estimate the impact of measurement error and determine the generalizability of assessment results? For instance, a possible concern with the SUPRA assessment is whether the results are generalizable beyond BD2A and BD6 since these are only a sample of all possible criterion tasks. Would the assessment scores be different if the assessments were conducted with other criterion tasks, such as BD1 or BD4? Classical Test Theory (CTT) partitions the assessment score variance only into true score variance and variance due to measurement error. There is no further information about what could have contributed most to the measurement error or how this can affect the generalizability of the results. Conversely, Generalizability (G) Theory provides a way to identify and estimate the influence of different sources of measurement error (Brennan, 2010; Shavelson et al., 1989; Webb et al., 1988).

For SUPRA, we assessed the squads’ leadership competency from BD2A indicators such as *SL calls Shift Fire* and *SL calls Lift Fire*, and from BD6 indicators such as *SL conveys plan* and *SL gives verbal directions to enter building*. Sources of variability and measurement error in assessing leadership competency include (i) the squads, (ii) the BD2A and BD6 criterion tasks, (iii) the various task-response pairs, (iv) the three iterations, (v) the OC raters, and (vi) the assessment method. For simplification, only the squad (S), criterion task (T), and occasion (O) facets will be discussed in this illustration. Table 1 compares how an observed competency assessment score from SUPRA is deconstructed in CTT and G theory.

**Table 1. Decomposition of an observed score by CTT and G Theory**

Classical Test Theory (CTT)	Generalizability Theory (G theory)			
Observed score = True Score	Observed score = Grand mean			
+ Error (not broken down by source)	<table border="0"> <tr> <td style="vertical-align: middle;">Potential sources of error</td> <td style="font-size: 2em; vertical-align: middle;">}</td> <td>                     + Squad effect (S)                      + Task effect (T)                      + Occasion effect (O)                      + (S x T) effect                      + (S x O) effect                      + (T x O) effect                      + (S x T x O) effect, error                 </td> </tr> </table>	Potential sources of error	}	+ Squad effect (S) + Task effect (T) + Occasion effect (O) + (S x T) effect + (S x O) effect + (T x O) effect + (S x T x O) effect, error
Potential sources of error	}	+ Squad effect (S) + Task effect (T) + Occasion effect (O) + (S x T) effect + (S x O) effect + (T x O) effect + (S x T x O) effect, error		

Under G theory, a G study is conducted to estimate the variance components to identify the facet that is the greatest source of measurement error. Results of the G study can be used in a decision (D) study to design studies that improve the generalizability of the measurement by increasing observations for the facets that need them (Brennan, 2010; Webb et al., 1988). For instance, if the concern was whether the assessments generalize across the task facet, multiple criterion tasks should be included, as in the SUPRA study design. Considering only the variability in assessments due to squad, task, and occasion, a G study would estimate the variance components from SUPRA (Table 2). Since the interest is in assessing squad teamwork competencies, “Squad” is not a source of measurement error, although it is still a source of variability in assessment scores.

Table 2. Variance components in a G study\* for the SUPRA study design

Source	Variance component	Description of variance component
Squad (S)	$\sigma^2_s$	Averaging over tasks and occasions, this component shows how much squads differed on the competency assessed
Task (T)	$\sigma^2_T$	Averaging over squads and occasions, this component shows how different competency scores are across tasks
Occasion (O)	$\sigma^2_o$	Averaging over squads and tasks, this component shows how different competency scores are across occasions
S x T	$\sigma^2_{ST}$	This component shows how different the squads' relative standings are across tasks
S x O	$\sigma^2_{SO}$	This component shows how different the squads' relative standings are across occasions
T x O	$\sigma^2_{TO}$	This component shows how different the rank orderings of tasks are across occasions
S x T x O, error	$\sigma^2_{STO,e}$	This component shows how different squads' relative standings are across tasks and occasions, and other unsystematic sources of error

\*Webb & Shavelson, 2005

Measurement error impacts decisions differently. The generalizability of assessment scores is not an immutable property of the assessment but depends on the type of decisions the assessment informs. A set of assessment scores may be generalizable for some decisions, but more data and observations may be needed for other decisions. G theory can provide decision-makers a way to quantify the impact of various sources of measurement error on the generalizability of the competency assessment. For *relative* or *norm-referenced* decisions, such as comparing squads to see which is the highest performing, the consistency of squads' rankings relative to each other on the competency is more important than their absolute scores on that competency assessment. For such decisions, assessments need only indicate the relative standing of the squad and the actual score is not required. In contrast, in *absolute* or *criterion-referenced* decisions, the assessment must show if the squad has met an absolute standard regardless of their relative standing. Therefore, an accurate actual competency assessment score is required (Shavelson et al., 1989; Webb et al., 1988).

Suppose a Commander needs to make a relative decision to select the best squad among the 15 squads assessed in SUPRA. A potential concern is that the relative standings of squads may change with the criterion tasks used in the competency assessments. Since the relative decision focuses only on the squads' rankings, the variance components of interest are those that involve the interaction of *squad* and other facets. Variance components that do not involve the *squad* do not enter error variance for this relative decision. These are the variance components of the main effect of *task*, *occasion*, and the interaction effect *task x occasion* (Table 2). The effects of *task* and *occasion* are the same across squads for a fully crossed design, so these facets do not change the squads' relative standings. As long as the variance components that involve squad are small, the decision-maker may still justify generalizing the assessment score across tasks and occasions even if their corresponding variance components are large (Webb et al., 1988). For instance, the variance component for *task* in SUPRA could be large, indicating that the average competency score for all squads in BD2A differs from that in BD6. Such a result does not influence the relative ranking of the squads on that competency because all squads had the same opportunity to be impacted by both tasks (Webb et al., 1988). If the decision to be made were an absolute one, then the assessment may not be as generalizable. Further confirmation about generalizability of the assessments can be obtained from the generalizability coefficients for the different decisions (Brennan, 2000).

## CONCLUSION AND FUTURE DIRECTIONS

---

This paper examines the GIFT/STEEL-R approach to competency assessments. Instead of regarding each isolated assessment event as a definitive measure of Soldier or team competencies, the approach allows multiple events and training experiences to provide behavioral indicators of competency. This view of assessments, however, raises concerns about measurement error and the generalizability of scores. Drawing from our work in the SUPRA study, we discussed the sampling and facets contributing to measurement error in any assessment and proposed the possible application of G theory to address these issues. Further research is needed to articulate how best to apply concepts of G theory within the GIFT/STEEL-R framework.

When assessing teamwork competency, it is important to derive individual-level assessments from team-level assessments and vice versa. We need to examine assessments at both the individual and team levels to understand team dynamics. While a team is greater than the sum of its members, any transformations to the team can only be achieved by changing the individual team members within it. One challenge is that not every team-level competency construct has an equivalent individual-level construct or operates similarly at the individual level (Kozlowski & Chao, 2012; Kozlowski & Klein, 2000). A thorough task analysis is needed to understand the relationship between the team and individual-level competencies. Multilevel researchers propose several models describing possible interactions and dynamics of individuals within the team that warrant different ways to derive the team- and individual-level assessments (e.g., Chan, 1998; Chen et al., 2005; Kozlowski & Klein, 2000). Future directions should include exploring some of these concepts and theories for competency assessments in GIFT and STEEL-R.

## ACKNOWLEDGMENTS

---

The research reported in this document was performed in connection with contract number W911NF-18-C-0046 with the U.S. Army. The views presented in this paper are those of the authors and should not be interpreted as presenting official positions, either expressed or implied, of the U.S. Government.

## REFERENCES

---

- Brennan, R. L. (2000). Performance Assessments from the Perspective of Generalizability Theory. *Applied Psychological Measurement, 24*(4), 339–353. <https://doi.org/10.1177/01466210022031796>
- Brennan, R. L. (2010). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education, 24*(1), 1–21. <https://doi.org/10.1080/08957347.2011.532417>
- Chan, D. (1998). Functional Relations Among Constructs in the Same Content Domain at Different Levels of Analysis: A Typology of Composition Models. *Journal of Applied Psychology, 83*(2), 234–246.
- Chen, G., Mathieu, J. E., & Bliese, P. D. (2005). A framework for conducting multi-level construct validation. In *Multi-level issues in organizational behavior and processes* (pp. 273–303). Emerald Group Publishing Limited.
- Goldberg, B., Owens, K., Gupton, K., & Hellman, K. (2021). *Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy*. 15.
- Jeffcott, S. A., & Mackenzie, C. F. (2008). Measuring team performance in healthcare: Review of research and implications for patient safety. *Journal of Critical Care, 23*(2), 188–196. <https://doi.org/10.1016/j.jcrc.2007.12.005>
- Kozlowski, S. W., & Chao, G. T. (2012). The dynamics of emergence: Cognition and cohesion in work teams. *Managerial and Decision Economics, 33*(5–6), 335–354.



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Kozlowski, S. W., & Klein, K. J. (2000). *A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes*.
- Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, 2(1), 41–63. <https://doi.org/10.1007/BF03546488>
- Shavelson, R., Webb, N., & Rowley, G. (1989). Generalizability Theory. *American Psychologist*, 44, 922–932. <https://doi.org/10.1037/0003-066X.44.6.922>
- Sottolare, 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 (ARL-HRED)*.
- U.S. Army Training and Doctrine Command, Army Capabilities Integration Center. (2016). *The U.S. Army Human Dimension Concept*. [www.army.mil](http://www.army.mil). <https://www.army.mil/standto/archive/2016/11/29/>
- Webb, N., Rowley, G., & Shavelson, R. (1988). Using Generalizability Theory in Counseling and Development. *Measurement and Evaluation in Counseling and Development*, 21, 81–90. <https://doi.org/10.1080/07481756.1988.12022886>

## ABOUT THE AUTHORS

---

**Dr. Grace Teo** is a Senior Research Psychologist at *Quantum Improvements Consulting*. Grace's research involves understanding and improving human performance under various conditions and in different contexts, such as working with different technologies, and in teams. Other research interests include assessments, decision-making processes and measures, vigilance performance, human-robot teaming, automation, and individual differences. Grace earned her Ph.D. and M.A. in Applied Experimental and Human Factors Psychology from the University of Central Florida.

**Dr. Eric Sikorski** is Director of Programs and Research at *QIC* where he leads a team of Research Psychologists and Human Factors Engineers. His work at *QIC* includes staff development, strategic planning for the division, and managing research initiatives for multiple DoD customers. Eric holds a Ph.D. and M.S. in Instructional Systems from Florida State University and an M.A. in Human Factors and Applied Cognition from George Mason University.

**Dr. Michael King** is a Human Factors Engineer II at *Quantum Improvements Consulting (QIC)*. While at *QIC*, Michael has led various training and human performance improvement projects, including a major initiative studying team communication and performance for the U.S. Army and evaluating virtual reality technologies for Air Force pilot training. Michael earned his Ph.D. in Experimental Psychology at Case Western Reserve University.

**Dr. Jennifer Solberg** is the CEO of *Quantum Improvements Consulting, LLC*. She has 15 years of military selection and training research experience, with an emphasis on leveraging innovative technologies for improving training in a measurably effective way. Her current research focuses on developing assessments of Warfighter performance to enable adaptive training, predictive modeling, and improved training effectiveness. She has led a team of senior government scientists in developing measures for identifying the cognitive and perceptual skills critical to visual IED detection. Jennifer earned her Ph.D. from the University of Georgia.



# Supporting Data-Driven Team Feedback and Scenario Adaptations in GIFT

Andy Smith<sup>1</sup>, Randall Spain<sup>2</sup>, Nicholas Roberts<sup>3</sup>, Benjamin Goldberg<sup>2</sup>,  
Jonathan Rowe<sup>1</sup>, Bradford Mott<sup>1</sup>, and James Lester<sup>1</sup>

<sup>1</sup>North Carolina State University; <sup>2</sup>U.S. Army Combat Capabilities Development Command (DEVCOM) - Soldier Center; <sup>3</sup>Dignitas Technologies

## INTRODUCTION

---

The U.S. Army's training modernization initiative calls for the integration of advanced virtual and augmented reality, modeling and simulation, and intelligent tutoring capabilities into synthetic training environments to create highly immersive and effective training experiences. Decades of research show intelligent tutoring can improve learning outcomes, however, integrating automated coaching, feedback, and instructional support into synthetic training environments to support collective training is an open area of research. Of central importance is developing data-driven models that can determine when, what, and how to support and coach teams as they conduct training and mission rehearsals in highly dynamic and immersive training environments.

The Generalized Intelligent Framework for Tutoring (GIFT) has been used for nearly a decade to design and investigate intelligent tutoring capabilities in immersive training environments. As an open-source modular framework, researchers have used GIFT to design intelligent tutoring-based training courses to teach marksmanship fundamentals (Goldberg & Amburn, 2015), land navigation and terrain association skills (Goldberg & Boyce, 2018), dismounted infantry battle drill fundamentals (Folsom-Kovarik & Sinatra, 2020), and counterinsurgency doctrine (Spain et al., 2022). In recent years, GIFT has been enhanced to support intelligent tutoring for teams (Sottolare et al., 2018). These enhancements include the addition of a team modeling team structure in GIFT's Domain Knowledge File (DKF), new condition classes and scenario adaptations to support team assessment, the ability to deliver feedback at the individual and team levels, an instructor dashboard, referred to as the GIFT Game Master Interface that facilitates a "human in the loop" for assessing performance and injecting scenario adaptations during collective simulation-based training events, and the ability to facilitate improved after action reviews (AARs) of team performance using multimodal-based assessment capabilities.

However, creating pedagogical strategies and tactics in GIFT that mimic the flexibility and techniques used by expert human coaches and that support data-driven adaptability remains a research and development goal. Team-level coaching in GIFT is primarily driven by feedback provided immediately after an error has occurred. The feedback statements and strategies must be manually authored and programmed into GIFT. Authoring real-time assessment models that can assess taskwork and teamwork skills can be resource intensive, particularly for scenarios and simulation-based training exercises that include multiple events or trials. Devising empirical, data-driven models that drive key decisions about how to provide adaptive feedback and coaching is critical for delivering effective collective training experiences. Furthermore, creating DKF templates and models that can be easily reused between training exercises, structures and promotes reuse between scenarios, and is critical for supporting modular design and scalability.

In this paper, we describe how we are providing data-driven coaching and feedback to teams in simulation-based training exercises. Building on our previous work, we present enhancements to GIFT enabling the operationalization of the previously defined frameworks for data-driven models of adaptive team feedback (Rowe et al., 2020; Spain et al., 2021) capable of driving tutorial planning and feedback through reinforcement learning (RL)-driven models (Smith et al., 2022). RL is increasingly being leveraged to

optimize decisions about how best to provide feedback and instructional support to maximize rewards during a learning event or training scenario (Chi et al., 2011; Rowe & Lester, 2015; Shen et al., 2018). Using crew gunnery as a use-case, we describe how we are modifying GIFT to enable course adaptations to facilitate both the exploration of different coaching strategies in a way that can effectively train RL-based models while at the same time maintaining the ecological validity of existing training scenarios. Our modifications focus on enabling GIFT courses to deliver a wider variety of feedback. This includes modifications to allow for more flexibility in the content (i.e., feedback that covers multiple course concepts) and modality of the feedback (e.g. textual vs. graphical) as well as modifications to allow for varying the timing of the feedback to enable partially delayed feedback and summative feedback. These enhancements are accomplished by creating a custom strategy capable of passing assessment data from GIFT's DKF to an external server, and displaying feedback generated from that server within GIFT's Tutoring User Interface (TUI) or through a customizable web page. We also introduce the capability to support micro-adaptive sequencing within Virtual Battlespace 3 (VBS3) scenarios by modeling a training exercise as a series of course objects in GIFT's course creator interface that each have their own DKF. This facilitates the ability to automatically reset and sequence individual engagements with GIFT without trainees having to restart VBS3. That is, GIFT can now automatically restart a training exercise and seamlessly move between unique task-based DKFs without needing to restart VBS3 between engagements. We also discuss future directions for these enhancements, including incorporating feedback information into the user's xAPI (eXperience Application Programming Interface) record, as well as generalizing the modifications to better serve a wider range of training domains and scenarios.

## RESEARCH CONTEXT

---

Teamwork is critical for mission success in the military. Teamwork requires team members to coordinate actions, exchange information, communicate effectively, and be collectively aware and anticipate the needs of team members. Research shows that teams that engage in teamwork processes are more likely to perform well under stress and produce better team outcomes (McEwan et al., 2017). Therefore, considerable attention has been given to the best approaches and strategies for training and assessing teamwork skills (Salas et al., 2008).

Human instructors, teachers, and coaches may use a variety of instructional strategies and tactics to support teamwork training and team performance. The tactics could take the form of explanations, guided coaching, or feedback that highlight specific misconceptions or errors that a team or team member committed. Or they may be gentle nudges or reminders about a concept or topic that the team has previously learned or steps that a team member should take in order to solve a problem or correctly perform a task. In addition, coaches will also alter the pace, flow, and challenge level of training exercises to fit the needs of the team. Furthermore, coaches and instructors will dynamically change their tutoring strategy. They may offer immediate feedback or they may hold back and provide an opportunity for students to engage in self-corrective actions.

Determining how feedback tactics and strategies can be implemented in an adaptive instructional system (AIS) to support data-driven team coaching and feedback is a critical step towards the development of robust adaptive team training systems. In recent years, machine learning techniques, including RL, have shown significant potential in crafting adaptive learning experiences by systematically tracking student actions and outcomes in immersive learning environments and inducing data-driven policies that shape the learning experience. RL-based tutorial planning techniques have been effectively applied in narrative-centered learning environments (Rowe & Lester, 2015; Sawyer et al., 2017; Wang et al., 2018), and intelligent tutoring systems for logic and probability (Abdelshiheed et al., 2023), as well as for sequencing concepts for elementary mathematics education (Mandel et al., 2014), and for adaptive remediation to support cognitive engagement during online training (Fahid et al., 2021; Fahid et al., 2023; Spain et al.

2021). Our work aims to extend these techniques to support adaptive team training in synthetic training environments.

## **CREW GUNNERY COURSE DESIGN IN GIFT**

---

At last year's GIFT Users Symposium, our team discussed how RL can be applied to a team training framework to support automated coaching for teams during team-based synthetic training exercises (Smith et al., 2022). Specifically, we discussed how key components of RL-based tutorial planners, such as state definition, action definition, and reward function, can be constructed to operationalize key components of an adaptive team feedback model (Spain et al., 2021), including learner and team attributes, feedback content and timing, scenario adaptations, and within-task and longitudinal performance outcomes. We also discussed challenges and tradeoffs of different approaches for developing automated coaching policies, such as technical constraints relating to the amount of data required to train models depending on the granularity of representations.

Over the past year, our team has continued working toward the goal of expanding GIFT to support data-driven coaching and feedback. We have developed a new crew gunnery test course in GIFT that can serve as a testbed for investigating various team-level feedback and coaching tactics. The course includes a VBS3 crew gunnery training scenario that emulates a high fidelity training and qualification range that U.S. Army gunnery crews use in preparation for live-fire training. The virtual scenario was developed at the Warrior Skills Training Center (WSTC) in Fort Hood, Texas and includes a variety of target types, gunnery platforms, offensive and defense battle positions, and terrain features.

The VBS3 crew gunnery scenario we are using involves a series six engagements. Each engagement requires crew members to implement fundamentals of the direct fire engagement processes outlined in TC 3-20.31-4 to detect, identify, decide, and successfully engage a target. The engagements can be configured to present crews with varying target conditions to include different: vehicle postures (offensive, defensive), target counts (single or multiple), target ranges (short, medium, long), target movement (stationary, moving, mixed), and visibility (day, night). The scenario supports three crew positions: Driver, Gunner (GNR), and Vehicle Commander (VC). An external evaluator or session proctor can use GIFT's Game Master to interface with the training scenario and to facilitate observer-driven assessment of crew coordination behaviors.

The basic sequence of an engagement is as follows. The Driver drives the vehicle to the battle position or firing box in the virtual range. After the vehicle reaches this location, the crew receives instructions for the engagement, such as "ISR Report a Truck in sector: Observe, engage, and report. OVER". These instructions are delivered to crews in VBS and can be triggered either by the evaluator using the Game Master interface or automatically based on scripting logic and triggers. After a short delay, the target engagement begins. During this period, crew members actively scan and coordinate search behaviors to detect the target in the environment. When a target has been positively identified, crews engage in the fire command sequence, which is a well-defined protocol for communicating information and coordinating actions to facilitate a coordinated response to a threat. During this stage, the VC and the GNR follow a prescriptive set of directives, spoken commands, and acknowledgements for engaging identified targets. After the target has been eliminated, crews provide a report to the evaluator and wait for instructions from the Tower concerning the next engagement.

The first engagement in our updated course is a defense engagement in which crews engage a stationary target from a fire box. The second engagement requires the crew to engage a single stationary target in an open environment, without cover. The third engagement requires crews to identify two targets as they move from one battle position to the next. In the case of multiple target engagements, crew must prioritize which

target to engage first based on the target's threat level. The fourth, fifth, and final engagements require crews to engage multiple stationary and moving targets and targets of different threat levels.

We designed the crew gunnery scenario to support automated assessment of several target engagement variables to provide crews with feedback and coaching to improve crew coordination behaviors. The logic driving the assessment, scenario adaptation, and feedback functionalities have been initially specified in the course's DKF, which our team manually authored using Graphical User Interface (GUI)-based tools in GIFT's interface. Table 1 provides operational definitions of selected target engagement variables we are modeling in our DKF.

**Table 1. Sample of crew gunnery timing variables and performance measures assessed during target engagements.**

Variable	Description	Assessment Type
Detect Target	The amount of time it takes a GNR to orient towards the target	Automated
Target Exposure Start	Timestamp of when the target animates to UP position	Automated
Target Detection Time	The amount of time it takes crew to issue verbal alert for target	Manual
Target Orientation Time	The time it takes a crew to orient the gun towards target	Automated
Open Time (OP)	Timestamp of when a crew initiates firing at a threat	Automated
Target Hit Time	Timestamp of the strike of the round to the indicated target was a TARGET hit	Automated
Target Hit Efficiency	Amount of time it takes to eliminate a target after Open Time	Automated
Engagement time	The amount of time it takes to defeat all targets in an engagement	Automated

As previously noted, during engagements crew members must search and scan the environment for potential threats. To facilitate target detection assessment we are using the *detect object* condition class in GIFT. This trigger allows GIFT to assess when the gunner orients the reticle of the gun towards the target. Ideally, we would be able to use this condition class for each crew member to diagnose timely target detection, however, current programming constraints in VBS3 prevent this. Once a threat has been identified, crews engage in a fire command sequence. Because crews are trained to follow a specific verbal protocol it is critical to assess whether crews are using correct terminology and engaging in closed loop communication. The current iteration of our crew assessment model utilizes an observed assessment condition class to facilitate the measurement of crew communication and coordination during this portion of the task. The observed assessment condition class allows a human observer to assess concepts defined in a GIFT-based training course using the GIFT Game Master interface.

The instructional strategies we are implementing in the course are compatible with data-driven coaching requirements for our sample scenario (Smith et al., 2022). We are investigating four generalizable strategies: feedback timing; feedback directive, feedback specificity, and instructional sequencing. Each of these instructional strategies is mapped to a lower-level adaptive event sequence that encodes a set of corresponding instructional tactics. The instructional tactics include coaching and feedback statements and adaptation decisions that address the following questions: Upon observing an error, should the AIS provide feedback or wait and continue to observe performance? To whom should the feedback message be directed towards? The individual who committed the error, the entire crew, or only the crew leader? If the decision is to provide feedback, how much detail or process-related information should be included in the message? Should the system provide detailed corrective feedback or provide a general prompt or reminder that probes the trainee to engage in self-directed reflection to diagnose the error? How can the scenario be adapted to

promote mastery learning and facilitate guided practice opportunities? One of the strategies we observed instructors using during crew gunnery training trials was resetting engagements whenever a crew failed to terminate a target in a timely manner. One of our goals is to replicate this instructional strategy in GIFT.

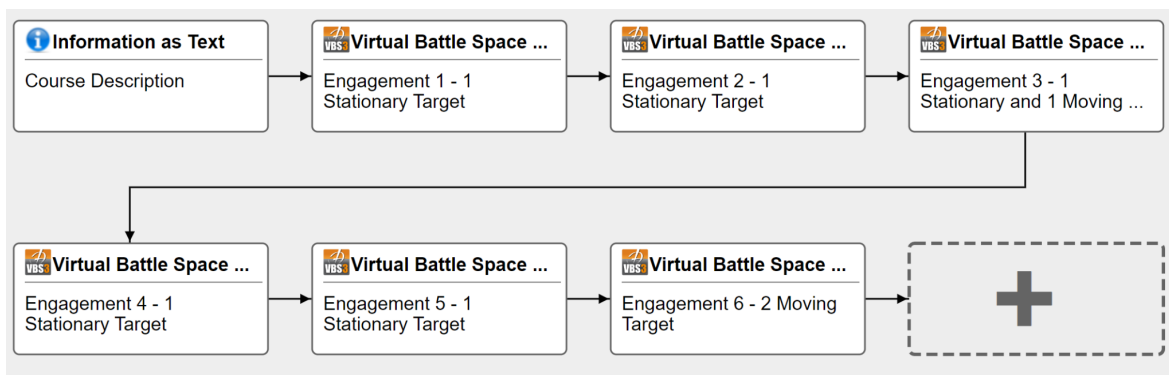
## GIFT MODIFICATIONS TO SUPPORT DATA-DRIVEN FEEDBACK AND SCENARIO ADAPTION

Although GIFT provides an expanding set of functionalities to develop adaptive training experiences there are several opportunities to enhance GIFT to support data-driven coaching and tutorial planning. In this section, we discuss two key extensions that we have implemented in GIFT to support improved team coaching feedback and assessment in GIFT: the Modular VBS3 Course Objects, and the External Feedback Server.

### Modular VBS3 Course Object

The first set of modifications we made to GIFT was the creation of a modular VBS3 course object to represent individual engagements in the training exercise. As described above, the Crew Gunnery Scenario we have targeted consists of 6 separate engagements, each of which is orchestrated by a similar set of phases, assessed by the same criteria, and meant to exercise similar learning objectives. Given this shared structure we sought to create a modular VBS course object, where each engagement is represented by a separate DKF in GIFT, while maintaining the same experience of one continuous VBS mission for the trainees. See Figure 1 for an example of the GIFT Course creator using the Modular VBS3 Course Object.

Each engagement consists of 4 phases. First, trainees are instructed to move to the starting point for the exercise. Next, a Detect phase is started where the crew must scan the environment and find the targets. After detection the crew must Engage the target and eliminate them within a defined time limit. Finally, a Feedback phase is triggered, after which the engagement is completed and the training course transitions to the next engagement. Additionally, a key requirement of this Modular VBS3 Course Object design is creating scenario injects allowing for the given engagement to be configured through the DKF, rather than relying on scripts and configurations in VBS. For example, the starting battle position for the engagement, the type of target(s), and position of the targets can be defined via the DKF.



**Figure 1. Example GIFT Course in GIFT Course Creator using Modular VBS3 Course Object**

This structure provides several advantages. First, it greatly simplifies authoring of a crew gunnery training scenario, as each engagement is self-contained in its own DKF reducing the total number of tasks, strategies, and state transitions that need to be managed in a given course object and DKF file. Second, the new structure allows for micro-sequencing and new scenario adaptations, including resetting and repeating a



given engagement. This was a common intervention used by trainers in the existing system we observed at WSTC. While not impossible in the existing GIFT structure, resetting as engagement midway through the scenario requires a large amount of logic/state transitions, which only increases in complexity as the scenario length increases. The new structure allows for an individual engagement to be easily reset/repeated an indefinite number of times, either by the instructor or a tutorial agent controlling the training activity.

Another benefit of the Modular VBS3 Course object is enabling easier orchestration of the training activity, as all scenario configuration and adaptation can be handled through GIFT and the Game Master interface. Without these new modifications, the instructor would have to navigate the VBS map editor interface to display targets, place trainees, etc., while at the same time entering assessments in the Game Master interface. In addition to focusing their attention on one interface, our changes also significantly declutter the Game Master interface, as only assessments and interventions for the given engagement are displayed at one time. Finally, the changes allow for xAPI statements to be generated after each engagement and each repetition of an engagement, rather than after the completion of the entire training activity. This is important as it can enable the generation of formative feedback during the training mission, or potentially more complex interventions such as dynamically re-sequencing engagements to maximize the alignment of scenario difficulty and training objectives.

### **External Feedback Server**

To best utilize the capabilities enabled by the Modular VBS3 Course Objects, our next step was to modify GIFT to allow for integration with an External Feedback Server. While GIFT currently supports much of the infrastructure required for collecting training data at scale, we looked to add modifications to better support data-driven feedback and scenario adaptations. As discussed at previous GIFT Symposiums, data-driven tutorial models have significant potential for improving learning outcomes. However, training these models requires the ability to collect data exploring the space of tutorial feedback, across a wide population of teams and abilities. By integrating the External Feedback Server with GIFT, we greatly increase the variety and quality of feedback we can provide to trainees, while also greatly enhancing the tools available for sampling the effectiveness of the feedback.

In the current configuration GIFT currently allows us to provide text messages to users via the TUI, or displayed in the VBS User Interface (UI). These messages are authored in the DKF, and assigned to a specific assessment concept and level. However, each engagement in the crew gunnery task consists of multiple assessment concepts, reflecting both task performance and team coordination. Additionally, to mirror what we observed at WSTC, feedback is often provided after an engagement has been completed, rather than immediately when the error occurred.

To address these issues, we created a feedback phase in the DKF of each engagement. At the beginning of the phase, GIFT pauses the VBS scenario, and sends a variety of information to an external server via HTTP request using a URL defined in the DKF. The JSON object generated by GIFT includes learner ability information on the team passed into GIFT at the start of the course, all assessment concepts for the engagement, and the assessment variables captured by GIFT during the engagement. The external feedback server receives this information, and returns a webpage which is displayed in the GIFT TUI interface. After interacting with the webpage, trainees can then continue the VBS session.

This functionality enables several benefits. First, it allows for rapid prototyping and testing of different feedback types and modalities. Web-based feedback provides an abundance of graphical and interactive options, allowing for both graphical and tabular representations of user performance and feedback, as well as more robust textual feedback such as templated messages incorporating multiple aspects of the users performance in the engagement. Constructive feedback techniques can also be easily leveraged via a web-based approach, though future modifications will be needed to ensure this information is passed back to



GIFT and incorporated into the xAPI record. Another benefit is the external server can be configured to enable weighted sampling of users at run-time. For example, a common policy when initially training RL-based systems is to randomly select tutorial actions for a period of time to collect initial data on how effective that feedback is for different levels of learners, situations, etc. This type of collection is currently not supported in GIFT, but is enabled by the external feedback server configuration. Additionally, more complex forms of weighting and sampling can be implemented, including “online” RL-based models that need the ability to vary feedback based on algorithms designed to manage exploitation vs. exploration tradeoffs.

## CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

---

Delivering tailored coaching and feedback to team members in synthetic training environments is a critical challenge. Using Crew Gunnery as an exemplar domain, we have identified and implemented several features that will reduce friction for developing data-driven feedback and training models for multi-task synthetic training scenarios using VBS3 and GIFT. These changes include a Module VBS3 Course Object allowing for complex DKFs to be decomposed into more manageable engagement level blocks, and adding the ability for GIFT to communicate assessments and display feedback from an External Feedback Server. Future research will integrate these advancements into the main GIFT distribution to be utilized by other projects. The External Feedback Server protocol will be documented and refined, as well as adding functionality to enable greater two way communication with GIFT, including passing information about what feedback was received, and potentially information explaining why the model made that decision. We will also add additional automated target engagement measures into the course and add functions to facilitate the point scoring system used by vehicle commander evaluators to score gunnery crews on training tables. Finally, we will use these new capabilities to collect data from gunnery crews, and train and evaluate data-driven team feedback models.

## ACKNOWLEDGEMENTS

---

The research described herein has been sponsored by the U.S. Army Combat Capabilities Development Command under cooperative agreement W912CG-19-2-0001. 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

---

- Abdelshiheed, M., Hostetter, J. W., Barnes, T., & Chi, M. (2023). *Bridging Declarative, Procedural, and Conditional Metacognitive Knowledge Gap Using Deep Reinforcement Learning* (arXiv:2304.11739). arXiv. <https://doi.org/10.48550/arXiv.2304.11739>
- Chi, M., VanLehn, K., Litman, D., & Jordan, P. (2011). Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. *User Modeling and User-Adapted Interaction*, 21(1–2), 137–180.
- Fahid, F. M., Rowe, J. P., Spain, R. D., Goldberg, B. S., Pokorny, R., & Lester, J. (2021, June). Adaptively scaffolding cognitive engagement with batch constrained deep Q-networks. In *Artificial Intelligence in Education: 22nd International Conference, AIED 2021, Utrecht, The Netherlands, June 14–18, 2021, Proceedings, Part I* (pp. 113-124). Cham: Springer International Publishing.
- Fahid, F. M., Rowe, J. P., Spain, R. D., Goldberg, B. S., Pokorny, R., & Lester, J. (2022). Robust Adaptive Scaffolding with Inverse Reinforcement Learning-Based Reward Design. *Artificial Intelligence in*

## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, 204–207.
- Folsom-Kovarik, J. T., & Sinatra, A. M. (2020, May). Automating assessment and feedback for teamwork to operationalize team functional resilience. In Proceedings of the 8th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym8) (p. 126). US Army Combat Capabilities Development Command–Soldier Center.
- Goldberg, B., & Amburn, C. (2015, August). The application of GIFT in a psychomotor domain of instruction: a marksmanship use case. In Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3) (pp. 115-124).
- Goldberg, B., & Boyce, M. (2018). Experiential Intelligent Tutoring: Using the Environment to Contextualize the Didactic. In Augmented Cognition: Users and Contexts: 12th International Conference, AC 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part II (pp. 192-204). Springer International Publishing.
- Mandel, T., Liu, Y. E., Levine, S., Brunskill, E., & Popovic, Z. (2014, May). Offline policy evaluation across representations with applications to educational games. In *AAMAS* (Vol. 1077).
- McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: a systematic review and meta-analysis of controlled interventions. *PloS one*, *12*(1), e0169604.
- Rowe, J., & Lester, J. (2015). Improving student problem solving in narrative-centered learning environments: A modular reinforcement learning framework. In *Proceedings of the 17<sup>th</sup> international conference on artificial intelligence in education*, (pp. 419-428).
- Rowe, J., Spain, R., Goldberg, B., Pokorny, R., Mott, B., & Lester, J. (2020, May). Toward data-driven models of team feedback in synthetic training environments with GIFT. In Proceedings of the 8th annual generalized intelligent framework for tutoring (GIFT) users symposium (p. 136). US Army Combat Capabilities Development Command–Soldier Center.
- Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human factors*, *50*(6), 903-933.
- Sawyer, R., Rowe, J., & Lester, J. (2017). Balancing learning and engagement in game-based learning environments with multi-objective reinforcement learning. In *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings 18* (pp. 323-334). Springer International Publishing.
- Shen, S., Mostafavi, B., Barnes, T., & Chi, M. (2018). Exploring induced pedagogical strategies through a Markov decision process framework: Lessons learned. *Journal of Educational Data Mining*, *10*(3), 27-68.
- Smith, A., Spain, R. D., Rowe, J., Goldberg, B., & Lester, J. (2022, May). Formalizing Adaptive Team Feedback in Synthetic Training Environments with Reinforcement Learning. In Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym10) (p. 117).
- Sottolare, R. A., Graesser, A. C., Hu, X., & Sinatra, A. M. (Eds.). (2018). *Design recommendations for intelligent tutoring systems: Volume 6-team tutoring* (Vol. 6). US Army Research Laboratory.
- Spain, R., Rowe, J., Goldberg, B., Pokorny, R., Mott, B., & Lester, J. (2021, May). Automated coaching in synthetic training environments: Developing an adaptive team feedback framework. In Proceedings of the ninth annual GIFT users symposium (GIFTSym9) (pp. 187-199).
- Spain, R., Rowe, J., Smith, A., Goldberg, B., Pokorny, R., Mott, B., & Lester, J. (2022). A reinforcement learning approach to adaptive remediation in online training. *The Journal of Defense Modeling and Simulation*, *19*(2), 173-193.
- Wang, P., Rowe, J. P., Min, W., Mott, B. W., & Lester, J. C. (2018, July). High-Fidelity Simulated Players for Interactive Narrative Planning. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 3884-3890).

## ABOUT THE AUTHORS

---

**Dr. Andy Smith** is a Research Scientist in the Center for Educational Informatics at North Carolina State University. He received his M.S and PhD. in Computer Science from North Carolina State University, and his B.S. degrees in Computer Science and Electrical and Computer engineering from Duke University. 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. Randall Spain** is a Research Scientist in the Soldier Effectiveness Directorate at U.S. Army DEVCOM, Soldier Center, Simulation and Training Technology Center (STTC). He holds a PhD in Human Factors Psychology from Old Dominion University and an M.S. in Experimental Psychology. His research focuses on the design and evaluation of advanced training technologies on learning and performance.

**Nicholas Roberts** is a Senior Software Engineer at Dignitas Technologies and the engineering lead for the GIFT project. Nick has been involved in the engineering of GIFT and supported collaboration and research with the intelligent tutoring system (ITS) community for nearly 10 years. Nicholas contributes to the GIFT community by maintaining the GIFT portal ([www.GIFTTutoring.org](http://www.GIFTTutoring.org)) and GIFT Cloud ([cloud.gifttutoring.org](http://cloud.gifttutoring.org)), supporting conferences such as the GIFT Symposium, and technical exchanges with Soldier Center and their contractors. He has also been heavily involved in integrating GIFT into TSS/TMT.

**Dr. Benjamin Goldberg** is a senior research scientist at the U.S. Army Combat Capability Development Command – Soldier Center, and is co-creator of the Generalized Intelligent Framework for Tutoring (GIFT). Dr. Goldberg is the technical lead for a research program focused on the development and evaluation of Training Management Tools for future Army training systems. His research is focused on the application of intelligent tutoring and artificial intelligence techniques to build adaptive training programs that improve performance and accelerate mastery and readiness. Dr. Goldberg has researched adaptive instructional systems for the last 15 years and has been published across several high-impact proceedings. He holds a Ph.D. in Modeling & Simulation from the University of Central Florida.

**Dr. Jonathan Rowe** is a Senior Research Scientist in the Center for Educational Informatics at North Carolina State University, Adjunct Assistant Professor in the Department of Computer Science, and Managing Director of the National Science Foundation AI Institute for Engaged Learning. He earned his Ph.D. in Computer Science from North Carolina State University in 2013. His research focuses on designing, developing, and evaluating AI-augmented learning and training technologies, with an emphasis on game-based learning environments, multimodal learning analytics, interactive narrative generation, affective computing, user modeling, and intelligent tutoring systems.

**Dr. Bradford Mott** is a Senior Research Scientist in the Center for Educational Informatics at North Carolina State University. He received his B.S., M.C.S., and Ph.D. in Computer Science from North Carolina State University. Prior to joining North Carolina State University, he worked in the video game industry developing cross-platform middleware solutions used extensively in commercial games and training applications. His research interests include AI and human-computer interaction, with applications in educational technology.

**Dr. James Lester** is a Distinguished University Professor of Computer Science at North Carolina State University, where he is Director of the Center for Educational Informatics. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI). His research on personalized learning technologies ranges from intelligent tutoring systems and game-based learning environments to affective computing, computational models of narrative, and natural language tutorial dialogue.



# Refining a Team Training Approach to Prepare Command Staff for Wargaming

Randy Jensen<sup>1</sup>, Evan Finnigan<sup>1</sup>, Grace Teo<sup>2</sup>, and Lisa N. Townsend<sup>3</sup>

<sup>1</sup>Stottler Henke Associates, Inc.; <sup>2</sup>Quantum Improvements Consulting, LLC;

<sup>3</sup>U.S. Army Combat Capabilities Development Command (DEVCOM)– Soldier Center

## INTRODUCTION

---

This paper gives an update on the design and implementation of a team trainer for preparing command staff for the collaborative processes of course of action (COA) development and wargaming. The target training audience are command staff at battalion and higher echelons, who must bring together the expertise of different warfighting functions as they synchronize plans in a key stage of the Army's military decision making process (MDMP). This process requires staff members to understand the interrelationships between roles and functions, which relates to individual and team-related training objectives for participants to effectively gain and use knowledge at three levels: within their own roles, provided to other roles, and needed from other roles.

For this team training application, the technical approach has evolved over several iterations in terms of the training use case and target audience, the role of human instructors driven by the use case, the assessment model and methods, and the plans for standards-oriented implementation methods. The earliest stage involved analysis of the relevant team dimensions, resulting in a model for teamwork constructs applicable to wargaming performance assessment (Teo et al., 2021). Also initial prototyping included preliminary design of a standards-oriented implementation (Jensen et al., 2021) with the structure for a distributed team trainer and mechanisms making use of the GIFT architecture (Generalized Intelligent Framework for Tutoring, Sottolare et al., 2012). An initial prototype version of the training application was presented in a previous GIFTSym paper (Jensen et al., 2022), with examples of team assessment methods in the context of wargaming exercises. Initial feedback from instructors led to a modified prototype design to support a use case adapted for a specific need in current wargaming training. Where the initial prototype attempted to create an approximation of a wargaming exercise in a distributed trainer, the modified design aims more toward *preparing* staff for wargaming, rather than *conducting* wargaming. This is because an important prerequisite for being able to conduct wargaming effectively is the understanding of the team roles. When the staff team has a shared understanding of the various team roles, they would be able to provide better information in support of each other and preemptively avoid errors when presented with an operating situation (Salas et al., 2014). Hence, the priority became targeting a need for training staff to first become familiar with the process of cross-functional coordination with different roles, and how they work together to synchronize plans for a COA. This prepares the staff for the next step where the focus moves more heavily to scenario-specific wargaming decisions. In order to support this modified exercise design and training objectives, a second prototype has been developed.

This paper gives an overview of the initial implementation of the second prototype, as context for a discussion of planned directions for conforming and integrating with GIFT. Although the earliest design outlined planned methods for building the system based on the GIFT architecture, the use case has changed and shifted away from earlier plans for using GIFT artifacts like the Domain Knowledge File (DKF). For example, as instructors elaborated on the objectives and practices for wargaming preparation exercises, they described team feedback methods that do not readily fit the mold for the kinds of performance scoring mechanisms that may be used for other domains. The team competencies that instructors reference in this domain tend to be limited in number and expressed at a high level. In practice, the more nuanced elements of team competencies are covered informally in after action review through instructor-led discussion and

team self-reflection. However, there is structure provided by the model for team dimensions in wargaming, and this is the starting point for revisiting how this application can be integrated with GIFT and benefit from its conventions. This paper outlines considerations for how exercise management, data flow, and assessment can be constructed with GIFT for this application. These considerations also may ultimately inform more general practices for other similar collaborative team decision-making domains. For instance, the GIFT DKF construct may be used to collect instances of teamwork markers generated by either the instructor or the automated rules in the environment. An example of where a marker is created is when a participant identifies information relevant to the COA that can be supplied by another role, which is an indicator of team cognition.

The following topics are discussed in this paper:

- Training application overview with example team learning objectives
- Architecture and planned interoperability with GIFT
- Assessment methods

## **TRAINING APPLICATION OVERVIEW**

---

As discussed above, the application described in this paper aims to provide team familiarization exercises to train U.S. Army command staff at battalion and higher echelons in the team processes associated with COA development, analysis, and wargaming. The prototype is under development as part of an effort called Reusable Automated Assessment and Feedback for Teams (RAAFT), which is being conducted for the U.S. Army Combat Capabilities Development Command – Soldier Center.

### **Training Environment**

The RAAFT prototype is constructed as a distributed, browser-based synchronous team trainer, which allows participants to be either remote or co-located. Exercises are led by a human instructor who also plays the part of the Lead planner, and participants are each assigned to one of six command staff roles representing different warfighting functions – Intel, Movement & Maneuver (MM), Indirect Fire (Fires), Aviation (AV), Protection (PRO), and Sustainment (SUS). Participants initially review pre-briefing materials about the operational scenario, such as a tactical map, mission statement, and enemy order of battle. As a familiarization exercise, when it comes to the consideration of a specific scenario and COA, one of the overarching goals is to orient the training audience more toward simply asking the right questions (process), as opposed to trying to arrive at the best decisions (performance outcomes). The purpose is to encourage thinking about the contributions of different warfighting functions, and also keep the exercise overhead to a minimum in terms of the amount of scenario-specific knowledge that participants must consume.

The exercise itself is conducted mostly by collaborating in a shared message panel, where the Lead / Instructor posts Prompts to initiate staff discussion on different Topics. The Lead / Instructor has a prescribed list of Topics with associated Prompts that can be sent to the message panel, but these Topics are only visible to the Lead / Instructor, and not to the training audience. Figure 1 below shows the Lead / Instructor screen, which includes the panel with instructor tools that are unique to this role (lower left, with a list of Topics and a Prompt field). The map panel and message panel are common to all participants, who engage in discussions triggered by Prompts, by either replying inline on existing threaded discussions, or adding new messages (using the input box) which are treated as new threads. Since the training environment is intended to support a variety of settings, participants may also communicate by other means (even verbally if in close proximity), but the training application has no access to such interactions. As



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

participants type messages, a selection of auto-complete options is available based on their partial inputs, or they can proceed to just type free text. Since the auto-completes come from a predefined library with markup for their meaning and relevance to Topics, this is one avenue for the system to understand messages coming from participants, and apply automated rules. For free-text inputs, most of the burden of understanding is on the Lead / Instructor. In the exercise snapshot below, the Lead / Instructor has sent two Prompts (seen in the message panel), most recently leading to discussion on the current active Topic “Indicators for EN repositioning.” In the Instructor tools panel (bottom left) the current Topic is marked with a check, to reflect the judgment that the staff team has adequately covered the Topic.

The screenshot displays the Lead / Instructor interface during an exercise. At the top, the title is "Critical Event - 1 BCT preventing the threat on OBJ TOM from reinforcing to OBJ HARRY". Below this are tabs for "Map", "Statement", "Red OB", "Resources", and "Exercise Control". The main map area shows a desert terrain with various objectives (OBJ TOM, OBJ HARRY, EA GUN) and units (PL DOG, PL CAT, PL RAT, PL MOUSE). A central panel displays a list of Topics, with "Indicators for EN repositioning" selected. A message panel on the right shows a conversation between participants, with a prompt from the Lead / Instructor. The bottom left panel shows a "Topics" list with checkboxes for each topic, and a "Send" button.

Figure 12. Training Application View of Lead / Instructor Screen During Exercise

Instructor Prompts are typically directed to a specific role, but anyone can take part in the discussion at any time. This is often essential, as the assigned roles in the exercise may not necessarily correspond to real-world expertise, sometimes by intention, so cross-functional participation is encouraged and in fact fundamental to effective wargaming processes.

As the exercise unfolds, noteworthy examples of teamwork (either positive or negative) are identified with markers. Markers can be created either by the Lead / Instructor or by automated rules in the system, and can be associated with a specific message and/or a Topic and/or a more general observation. In the example above, there is a marker (small bookmark icon) that has been created and associated both with the current Topic (“Indicators for EN repositioning”) and with the most recent message from Sustainment (“Consider movements of support elements...”). The markers capture significant examples of teamwork from the exercise for further discussion in after action review (AAR). In this case, the marker was created by an automated system rule. The analysis is made possible by the fact that the participant in the Sustainment

role used an auto-complete message, which could be recognized by the system for its relevance to the current Topic. Two operations happened as a result. First, from internal associations authored in the auto-complete library, the system determines that the message is considered an optimal response for the current Topic. Specifically, the enemy support elements are an effective indicator that maneuver units are repositioning, and this is a better choice than the initial response from Intel (which refers to enemy command elements). So the Topic is automatically marked as covered in this case, although more generally it can be either the system or the Lead / Instructor making judgments about when a Topic is adequately covered. The second operation is the creation of a marker based on observed teamwork processes. Although the original Topic involved a Prompt that was directed to Intel, it was the Sustainment role that contributed the optimal response. This is one situation that the system identifies as positive teamwork, so it creates a marker tagged with the Supporting Behavior team dimension. Using the popup window below in Figure 2, the Lead / Instructor reviews the marker and adds further information.

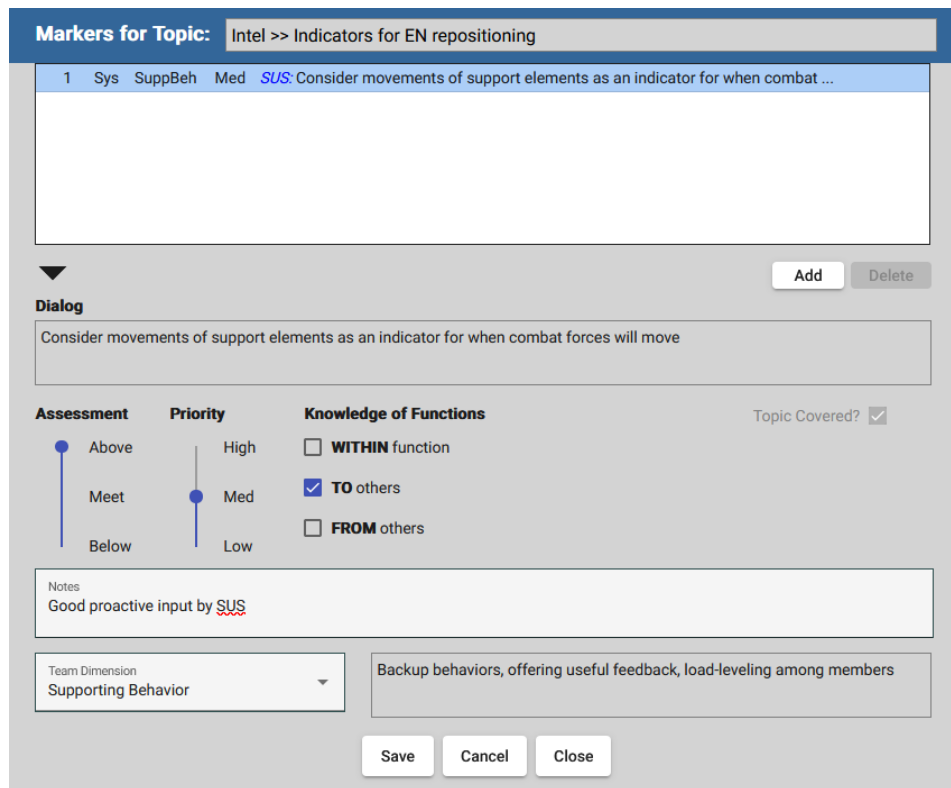


Figure 13. Marker Review Popup for Lead / Instructor

The representation for markers supports several kinds of supplemental information, including:

- Contextual information
  - Relevant message and Topic, if any
- Qualitative assessment
  - Above, Meet, or Below expectations
- Priority
  - High, Med, or Low, for reference when preparing AAR
- Functional knowledge
  - Concept relating to whether the marker relates to a particular participant’s knowledge WITHIN their own functional area, to be provided TO another function, or needed FROM another function



## Proceedings of the 11th Annual GIFT Users Symposium (GIFTSym11)

- Relevant team dimension
  - Referencing a model of team dimensions applicable to this domain, such as Team Cognition, Information Exchange, and Supporting Behavior (Teo et al., 2021)
- Additional notes from the Lead / Instructor

The Lead / Instructor input in this example includes an Assessment notation of Above expectation, and a Knowledge of Functions notation that this exchange involved information flow To others (in this case, Sustainment knowledge about logistics elements, provided to Intel and the rest of the staff). The Lead / Instructor also adds a comment that this is good proactive input.

### Learning Objectives

The learning objectives targeted by the RAAFT training application are mostly associated with team processes. In contrast to team performance which places more emphasis on outcomes, team processes relate to the dynamics of interactions or cognitive states within team members (Grand et al, 2013), and are foremost in this application since a major exercise goal is to familiarize participants with the different staff roles and their cross-functional dependencies. There are several team processes activated in the training application, and so one of the design questions for structuring a competency model for this domain involves defining how competencies will be measured, retained, and tracked over multiple exercise iterations, and potentially within and across teams. Specifically, we consider four different kinds of measures:

- **Topic coverage.** For the Topics associated with the operational scenario in the exercise, did the team adequately cover the questions that needed to be considered or addressed? This corresponds directly to existing practices followed by instructors when they monitor wargaming exercises and maintain a checklist to mark notes organized by topics in a mental model tailored for the scenario. In this sense, each Topic in the exercise has an associated learning objective that can go into a competency model (Trigger for EN repositioning, Indicators for EN repositioning, NAI for repositioning indicators, EN reinforcement position...).
  - Example from above: did the team adequately identify considerations about the indicators for enemy repositioning?
- **Team dimensional constructs.** For teamwork processes exhibited in the exercise, what are the relevant team dimensions? This makes reference to a model for teamwork constructs applicable to wargaming performance assessment developed earlier in this effort (Teo et al., 2021). However, it is challenging to treat the team dimensions as competencies in the conventional way that other competencies are handled, because of their abstract nature. It remains a research question whether and how it is suitable to construct an exercise that would aim to mark an individual or team as having reached a certain scored proficiency on Team Cognition or other similar measures. Instead, the team dimensions are treated as supplemental information to accompany markers for more concrete measures such as the instances of Topics covered well or poorly. Thus the initial implementation treats the team dimensional constructs as having a secondary role rather than being independent competencies tied to learning objectives.
  - Example from above: what team dimensions were exhibited in the process of the staff's discussion of indicators for enemy repositioning?
- **Post-exercise self-reporting.** As participants reflect on the exercise during AAR, what did they learn about their own roles and others? This kind of explicit reflection (often called “sustains and improves”) is a common practice with existing training, so it is mirrored in exercises to be conducted in this environment. Since an AAR is conducted within the training environment, where

participants can refer back to the discussion of different Topics, self-reported takeaways can be compared with markers from the Lead / Instructor and also add enhanced information.

- Example from above: does the participant playing the Intel role echo anything they learned from the Sustainment suggestion to consider enemy support elements rather than command elements?
- **Specific performance measures.** Aside from other measures above, what are concrete indications of team processes that can be inferred from analysis of exercise data? Examples include communication dynamics (e.g., statistics about relative levels of participation from different roles), time to complete Topics, repetition or disjointed communication across threads.
  - Example from above: How long did the staff spend on the discussion of indicators for enemy repositioning?

## **ARCHITECTURE AND PLANNED INTEROPERABILITY WITH GIFT**

---

At the highest level, the RAAFT prototype is a server-based team training application that users must access synchronously from their browsers. Exercises require users to take unique roles corresponding to the different staff warfighting functions, so the process for joining a training exercise must include the assignment of roles. For the purposes of GIFT integration, the RAAFT prototype is treated as an external training application, which is readily supported in an architecture where a GIFT Cloud instance communicates with the RAAFT application and its server. This section describes the architecture in more detail, starting with the structure of the training application itself, and then the integration and data flow with GIFT. The architecture for interoperability with GIFT is currently under development, so the discussion of integration methods is aimed at the current design concept.

### **RAAFT Training Application**

The RAAFT application architecture takes the form of a MEAN architecture which stands for MongoDB, Express, AngularJS, and Node.js. This is a standard architecture used to build web applications.

#### ***RAAFT Client***

The RAAFT client is an Angular application. Angular is a front-end web framework designed for creating single-page web applications. Angular is created on top of Node.js which is a JavaScript runtime for building server-side or desktop applications. The RAAFT client is a thin interface that delivers information and takes user interaction but leaves all significant processing to the server. Figure 1 in the previous section shows the client interface as seen by the Lead / Instructor.

#### ***RAAFT Server***

The server is a Node application that provides a REST API (Representational State Transfer Application Programming Interface) for the web client application. This REST API manages exercise data flow to and from users, including the delivery of scenario information and the recording of all user inputs including exercise-related communications as well as actions by the Lead / Instructor such as the use of teamwork markers. One of the key requirements of a REST API is that it is stateless, which means that the server does not store any information related to previous requests. All data that needs to be persisted is stored in the database. The REST API receives all data in JSON (JavaScript Object Notation) format. This is a natural choice for a MEAN stack because JavaScript and TypeScript objects can be easily serialized into JSON, and the MongoDB database can easily process data in JSON format for storage.

In addition to receiving user-created data and persisting it to the database, the RAAFT application server also performs a range of processing tasks. The RAAFT application server can use rules to create assessments, and also do some basic processing to manage the session. For example, when the Lead role ends a session, the server clears all of the messages, the currently selected topic, markers, and topic statuses.

### Database

A MongoDB database stores and persists all application-related data. A MongoDB database uses a non-relational approach to storing data with flexible data models. The MongoDB database stores all of the dynamic exercise data: messages, status of Topics, and markers as well as the static data like the Prompts and the full library of auto-complete message options. All of this static data is pulled into the database from JSON files at the beginning of the exercise, so content is added or modified via those JSON files. An export capability is supported in order to save a log of the dynamic exercise data at the end of an exercise.

### Deployment

The RAAFT client, server, and database software components all need to run simultaneously, so they are containerized using Docker, which packages software and execution requirements into a lightweight environment to alleviate the need for separate manual installation. The three software components are deployed on a Linux server running Docker. The Linux server is also configured to accept requests on specified ports to allow users to request the web application and then to allow the web application to connect to the REST API to receive data.

### Interoperability with GIFT

Figure 3 below shows a simplified architecture for the RAAFT training application and GIFT.

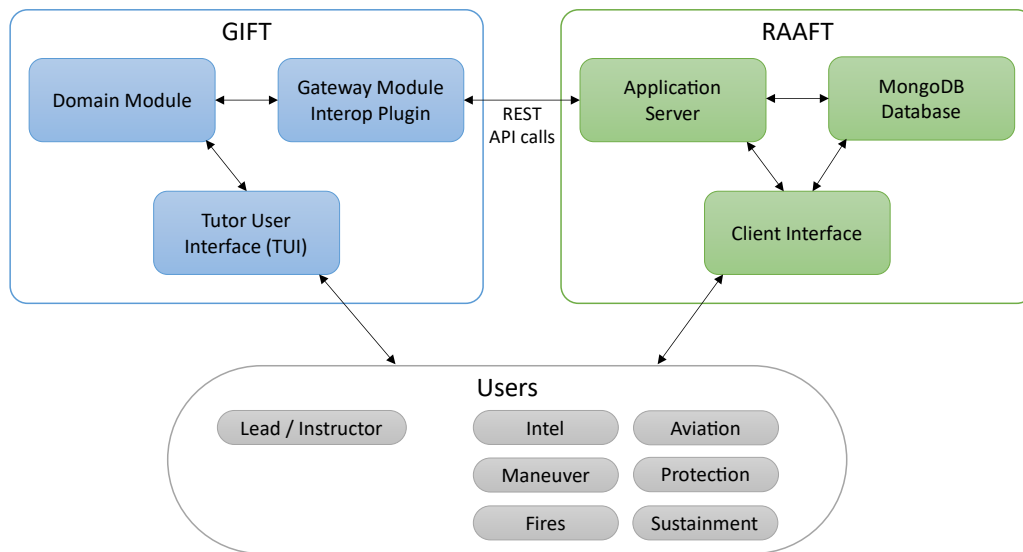


Figure 14. Interoperable Architecture with the RAAFT Training Application and GIFT

All users interact with both applications through a browser, with simultaneous active sessions to access (i) a GIFT Cloud instance through the Tutor User Interface, and (ii) the training application through the RAAFT Client. RAAFT and GIFT communications are managed with REST API calls between the RAAFT Server and the GIFT Gateway Module Interop Plugin. For the current implementation efforts, the

focus is on interoperability within the scope of a single exercise. Although the mechanism is designed to be in place to maintain records of exercise results and team competencies in GIFT, the initial implementation does not make use of GIFT capabilities to manage exercise sequencing or instructional models.

The initial integration between the training application and GIFT consists of two parts: coordinating the launch of an exercise with users assigned to specific roles, and delivering training application state messages to the GIFT Domain Module.

***Exercise Launch***

The goal for an ideal configuration is to make the exercise launch process as seamless as possible, where an exercise participant can go from a GIFT lobby straight to participation in the training application, with robust support for loss of connectivity to either application (the GIFT Cloud instance or the RAAFT training application) without loss of session role assignments and data. The initial architecture reflects a step toward this goal, but with the requirement that users open independent browser instances for both applications. In this configuration, the sequence is as follows in Table 1:

**Table 3. Exercise Launch Sequence**

<b>Event</b>	<b>Role</b>	<b>GIFT</b>	<b>RAAFT</b>
<b>Pre-launch</b>	All users (including the Lead / Instructor as well as all participants taking command staff roles)	Lobby interface open in browser, showing a page to select a role for the exercise. The Lead / Instructor has unique credentials for this step, as the only person authorized to select that role. All others can only select from the command staff roles.	Client open in browser, showing a blank landing page confirming successful connection, and awaiting role selection.
<b>Role assignment</b>	Any user	Upon selecting a role, the Gateway Module Interop Plugin sends a configuration message to the RAAFT server, associating the user with a role.	After receiving the configuration message from GIFT, the landing page shows the role assignment. Until the exercise starts, all participants see their roles, but do not enter the exercise environment yet.
<b>Exercise start</b>	Lead / Instructor		The client interface for the Lead / Instructor includes exercise control tools. When the Lead / Instructor starts the exercise, the exercise environment becomes active in the client interfaces for all participants. Any participant joining after this step goes directly from role assignment to the active exercise.

***Data Flow***

GIFT treats RAAFT as an external training application in this architecture, in the sense that assessment logic is internal to the RAAFT application, and the assessment outputs are primarily the teamwork markers created either/both by the system and the human Lead / Instructor role. Fundamentally all data flow between the training application and GIFT before, during, and after an exercise takes place via REST API calls. This approach matches well with the existing implementation of the RAAFT prototype which already

makes use of a REST API. For the purposes of GIFT integration, the REST API is implemented inside of a GIFT Gateway Module Interop Plugin, which then allows Java code written on the RAAFT side to interact with Condition Classes in the GIFT Domain Module. During an exercise, the updates to Topic status and the Teamwork markers exercise are conveyed to GIFT by this mechanism. Both of these types of data are serialized as JSONs using a schema with the necessary data.

For markers originating from humans, the data flow will likely remain primarily a relay function as these markers are sent as state messages to a GIFT Condition Class. There is also the possibility of using or adapting the Observed Assessment mechanism in the GIFT Game Master tool as a means for the Lead / Instructor to input markers, as long as this does not create confusion switching applications during the exercise. However, for system generated markers, there is logic for rules that process exercise data to identify situations where certain markers can be created and tagged. Although the initial plan in terms of interoperability is for this logic to remain internal on the RAAFT Server, it is a future goal to experiment with abstraction by implementing at least this portion of assessment logic to an implementation within GIFT Condition Classes associated with nodes in the DKF concept hierarchy. This next phase of integration requires dynamic exercise data such as Prompts and participant messages to be sent during the exercise. In addition, the Condition Classes need to be initialized with scenario knowledge consisting of static data for Prompts, Topics, and the auto-complete library. The following section goes into further detail about assessment methods in the training application, and the relationship to GIFT mechanisms for assessment via the Domain Module.

## ASSESSMENT METHODS

---

As discussed above, the concept for the implementation of assessment methods is phased. The plan for a future phase is to explore abstracting assessment logic to run using GIFT constructs external to the training application. However, initially, from a GIFT integration standpoint, RAAFT is considered an existing training application with its own internal assessment engine. In this configuration, the markers are treated as the primary training application state messages passed from RAAFT to GIFT, and the markers are collected by a Condition Class in the GIFT Domain Module. DKFs on the GIFT side associate incoming marker data with concept hierarchy nodes, which predominantly correspond to the Topics in the exercise.

Referring back to the view of the training environment in Figure 1, the hierarchy of Topics is organized by warfighting functions corresponding to the staff roles occupied by participants. Under the Intel role, four topics are visible: Trigger for EN repositioning, Indicators for EN repositioning, NAI for repositioning indicators, and EN reinforcement position. Each of these Topics (and all Topics associated with other roles) has a counterpart in the concept hierarchy in GIFT. When the team of participants covers a Topic, with any associated markers created to convey supplemental information such as relevant team dimensions or the relationship to knowledge of functions, the delivery of a message to GIFT becomes the basis for a record of the results for the corresponding concept.

As a collaborative team problem-solving domain, it is difficult to define a set of competencies that matches the scope of instructional training objectives. Both for the broad areas of wargaming and MDMP, and the more specific wargaming preparation and team familiarization exercise delivered by the RAAFT training application, instructional objectives are mainly oriented at giving cadres of exercise participants opportunities to practice team collaborative processes. Conventionally implementable artifacts like enumerated competencies with scored levels of mastery tend to be only indirectly used in existing training for these domains, and only at a high level (example: ELO-AOC-12.1 “Use the Military Decision-making Process (MDMP) to plan a tactical operation”). Since team compositions routinely shift from one exercise to another, there is little meaning to an approach that would involve persistent team-level models of mastery. If a team of six individuals demonstrates effective teamwork processes in the handling of Intel related topics

in an exercise, that may or may not be an indicator of likely teamwork in a future exercise involving four of those six, with two roles swapped out. Another approach as an alternative to a persistent team model for a specific team composition might be to simply distribute findings about teamwork competencies to each participant, stored in their individual student models. But this approach also has downsides, because there are many situations where it is not appropriate to attribute positive and negative team processes to each individual on the team.

A further complication comes from the fact that teamwork inherently relates to processes or situations that may arise from differences across participants or roles. Referring back to the exercise example shown in Figure 1, the situation in this case started with the Intel role initially suggesting a sub-optimal response to a Prompt (to consider enemy command elements as an indicator of repositioning). From a straightforward assessment perspective, this response from Intel could be considered a negative training point, perhaps an individual error. However, the emphasis of the exercise is on teamwork processes, as opposed to performance outcomes. And the process in the same situation involved a follow-up contribution from Sustainment, suggesting a better response (to consider enemy support elements as an indicator). The goal of teamwork assessment in this situation is to recognize the effective collaborative process that led to the collective team's response. In this case the Prompt was directed at Intel, but Sustainment gave the good response, and in fact any role could have. As a result, when the Topic is marked as covered, and supplemented with information from the Lead / Instructor, this information is conveyed to a GIFT DKF at the team level.

This example illustrates the thinking behind the current design for how the GIFT DKFs are structured for this domain. The DKFs apply to the entire team as a whole. Although a Topic may have a notional association to the Intel function, any staff member may contribute to the discussion, and any positive or negative teamwork processes involved in the discussion are attributed to the entire team. If specific roles made key contributions, this can be preserved in the supplemental information conveyed with markers for the Topic, but at least from a representational perspective, the Topic is not strictly confined to the one role. The DKFs also essentially make use of Tasks that apply throughout the exercise, since there are no specific bounds on when a staff team may discuss a Topic. They may discuss a Topic before receiving a related Prompt from the Lead / Instructor, and they may also revisit a Topic after having moved on to other Topics or even after the original Topic has been marked as covered. One advantage to this approach is that it reflects a relatively simple structure for DKFs and team organization, rather than having large numbers of DKFs for individual, pair-wise, or n-wise subsets of roles.

## **CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH**

---

The design approach for interoperability with GIFT discussed in this paper is under development, so the intention is to document lessons learned from the development process in the future. Although there are complexities in fitting a training application for collaborative problem-solving to a structure designed for more enumerable, discrete, and scorable competencies, the tools in the GIFT framework provide an effective basis for building an approach that can be reused for similar domains.

A relevant research question in developing the integrated training approach is where the opportunities are for reusability across scenarios and potentially even training environments. This is an ongoing area of investigation from a couple of different perspectives. First, in terms of using the information embedded in markers about abstract team dimensions (team cognition, information exchange, etc.), a potential question is: if a team using this training applications demonstrates effective teamwork indicated by markers for certain dimensions, how likely is it that this represents findings that would hold true for the same team and same dimensions, but in other scenarios and even other collective tasks? There may be some dimensions

that translate better than others; for example supporting behavior might be a recurring feature of a team, whereas information exchange or team cognition might be more closely related to individual knowledge.

Second, in terms of the teamwork markers tied to specific exercise Topics, there's a similar question: if the team performed well on that Topic for this scenario, would it be reasonable to expect them to perform well on the same Topic in a different scenario? This seems more likely to recur across scenarios, but it remains an area of investigation

This research focus was highlighted in a previous GIFTSym paper (Jensen et al., 2022), which conveyed that the purpose of assessment and the review of teamwork markers generated in an exercise is initially limited to the scope of the exercise itself. However, using GIFT to enable the tracking of data over time, future development should expand the use case to situations where the same team composition engages in an exercise multiple times (or a series of exercises over time), where the goal is to see improvement among the team as a result of repeating the cycle and team review. As discussed earlier, this can become a complex team modeling problem when dealing with changes in team composition, but a baseline case can assume iterations with the same team composition.

One of the observations from instructors has been that front-end elements may also be an effective area for reuse. As the wargaming preparation training application may be more broadly considered a team familiarization exercise, the collection of user interface panels may also have value if implemented as reusable components to be readily adapted for other domains. This observation from instructors arose especially as a result of their experiences with challenges conducting team training events during the pandemic, where existing chat and teleconference tools became the norm but also fell short of instructional needs in certain ways. A common set of panels that can be composed into a browser-based training environment (map or image panel, chat / message panel, instructor tools, etc.) could be developed as a set of reusable front-end elements in the GIFT toolset, to support a potential cluster of related team training applications. Ultimately the aim is not solely to develop a single trainer for a specific application, but to derive lessons from this process for other similar future training needs.

Finally, another area for future work relates to the authoring of scenario content. Since a library of auto-complete messages plays a key role in the automated rules executed in the RAAFT training application, one of the future development goals is to grow the library over time by mining exercise data. When participants type free-text messages, and these are tagged in markers created by Lead / Instructor roles, the accumulated data set amounts to a source for semi-automated expansion of the library. Over time, this may reduce burdens on both instructors and authors.

## References

---

- Grand, J.A., Pearce, M., Rench, T.A., Chao, G.T., Fernandez, R., & Kozlowski, S.W. (2013). Going DEEP: guidelines for building simulation-based team assessments. *BMJ Qual Saf.* 2013 May;22(5):436-48. doi: 10.1136/bmjqs-2012-000957. Epub 2013 Jan 25. PMID: 23355693.
- Jensen, R., Presnell, B., DeFalco, J., and Goodwin, G. (2021). Designing a Distributed Trainer Using GIFT for Team Tutoring in Command Level Decision Making and Coordination. In *Proceedings of the 9th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym9)*. 176-185.
- Jensen, R., Finnigan, E., Teo, G., & Goodwin, G. (2022) Team Assessment Methods for Training Command Staff Decision Making. In *Proceedings of the 10th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym10)*. 157-167.
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2014). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, 54(4), 599–622.



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

Teo, G., Jensen, R., Johnston, J., DeFalco, J., & Goodwin, G. (2021). Measures for Assessing Command Staff Team Performance in Wargaming Training. Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC 2021).

## ABOUT THE AUTHORS

---

***Randy Jensen** has been a project manager for over 20 years at Stottler Henke Associates, Inc., in San Mateo, California, where he has led projects to develop intelligent tutoring systems for domains involving complex decision-making. Examples include trainers for air attack planning, unmanned vehicle command and control, small unit tactics, combined arms team exercises, information systems troubleshooting, satellite scheduling, and a current effort for division level command group wargaming. Randy has a B.S. with honors from Stanford University.*

***Evan Finnigan** is a software engineer at Stottler Henke Associates Inc., where his work includes developing intelligent tutoring software and building interactive web-based user interfaces. Previously, Evan was a student researcher in a human robot interaction lab where he built web-based interfaces to control an assistive robot. Evan has a B.S and M.S. from the University of California at Berkeley.*

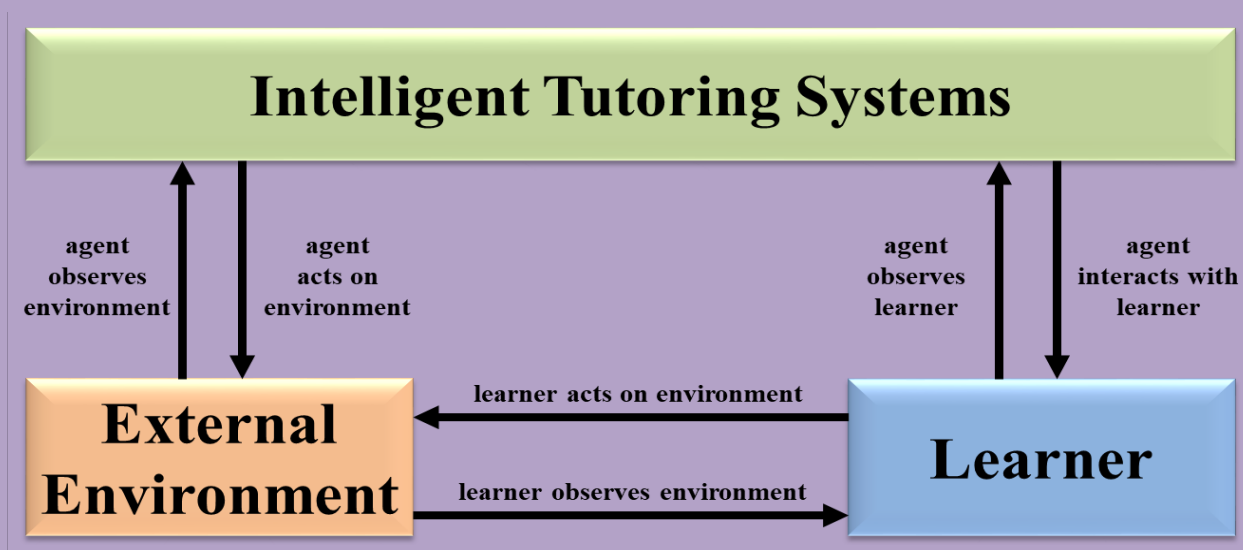
***Dr. Grace Teo** is a Senior Research Psychologist at Quantum Improvements Consulting. Grace's research involves understanding and improving human performance under various conditions and in different contexts such as working with different technologies, and in teams. Other research interests include assessments, decision making processes and measures, vigilance performance, human-robot teaming, automation, and individual differences.*

***Lisa N. Townsend** is a Psychologist who has been working at STTC (SED/DEVCOM SC) in Orlando, FL since August 2022. Prior to STTC, she spent 27 years as a Research Psychologist at the Naval Air Warfare Center Training Systems Division (NAWCTSD) in Orlando, FL. She has a Master of Science in Industrial/Organizational Psychology and a Bachelor of Arts in Psychology, both from the University of Central Florida (UCF). She has worked on many diverse teams including those within Research and Development, Technology Transfer, Instructional Systems Design, and Human Systems Integration, in efforts that have spanned across Services and platforms.*



# Proceedings of the Eleventh 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 technologies for evaluating the effectiveness of tutoring applications. 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.



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:*

- *Dr. Anne M. Sinatra is a Research Psychologist at the U.S. Army Combat Capability Development Command – Solider Center and has been a part of the Generalized Intelligent Framework for Tutoring (GIFT) team since 2012. Dr. Sinatra was the GIFTSym11 Program Chair.*

