Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy

Benjamin Goldberg
US Army DEVCOM Soldier Center
Orlando, FL
benjamin.s.goldberg.civ@mail.mil

Kevin Owens, Kevin Gupton
Applied Research Laboratories University of Texas at Austin
Austin, TX
kowens@arlut.utexas.edu; kgupton@arlut.utexas.edu

Kevin Hellman
US Army TRADOC Proponent Office Synthetic Training Environment
Fort Leavenworth, KS
kevin.m.hellman2.civ@mail.mil

Robby Robson
Eduworks, Inc.
Corvallis, OR
robby.robson@eduworks.com

Shelly Blake-Plock
Yet Analytics
Baltimore, MD
shelly@yetanalytics.com

Michael Hoffman
Dignitas Technologies
Orlando, FL
mhoffman@ dignitastechnologies.com

ABSTRACT

The Army's Synthetic Training Environment (STE) modernization program's Training Management Tools require capabilities that objectively measures and evaluates performance over time. Persistent tracking of individual and team performance data enables these tools to better infer proficiency levels, identify strengths and weaknesses, and adaptively tailor coaching and remediation. The STE Experiential Learning for Readiness (STEEL-R) project addresses this requirement by establishing an interconnected system of systems built on open source software and commonly applied data standards.

The STEEL-R team is developing an extensible data strategy that interoperates across Live, Virtual and Constructive environments. It uses real-time processing to translate data sources into meaningful assessments that align to warfighter competency requirements. To demonstrate this concept, STEEL-R leverages tools and methods from the Army's Generalized Intelligent Framework for Tutoring (GIFT) and the Advanced Distributed Learning (ADL) Initiative's Total Learning Architecture (TLA) projects. GIFT is used to capture and interpret raw learner data, then TLA standards and business practices are applied to communicate outcomes to a competency management system for readiness and talent tracking and to a persistent data lake to support decision analytics pipelines.

In this paper, we describe the functional components of the STEEL-R architecture and illustrate it in the context of a Rifle Squad use case, focusing on data flows and processing from the training point-of-need to the Army enterprise cloud. The STEEL-R architecture serves as a reproducible data strategy for STE that can extend cross-service. It aligns evidence-based metrics derived from operational training exercises with established competency frameworks for every echelon and individual role. These frameworks inform the performance metrics and type of data that must be reported to make meaningful inferences on competency proficiency. We will conclude with a discussion on the future capabilities a training management architecture and set of data strategies of this nature can potentially support.

ABOUT THE AUTHORS

Dr. Benjamin Goldberg is a Senior Scientist at the U.S. Army Combat Capabilities Development Command (DEVCOM) Soldier Center, Simulation and Training Technology Center (STTC) in Orlando, FL. He has been conducting research in the Modeling & Simulation community for the past twelve-years with a focus on adaptive learning in simulation-based environments and how to leverage Artificial Intelligence tools and methods to create personalized learning experiences. Currently, he is the lead scientist on a research program developing adaptive training solutions in support of the Synthetic Training Environment. Dr. Goldberg is co-creator of the award winning Generalized Intelligent Framework for Tutoring (GIFT) and holds a PHD from the University of Central Florida in the program of Modeling & Simulation.

Kevin Owens is an Engineering Scientist at the Applied Research Laboratories, The University of Texas at Austin – a US Navy University Affiliated Research Center (UARC). After retiring from the US Navy, Mr. Owens earned a BS and MS in Instructional Design and has 19-years' experience in DoD learning engineering science and technology (S&T) and DoD operational research. Mr. Owens supports US Army PEO-STRI and the Synthetic Training Environment (STE) requirements process, the Squad Immersive Virtual Trainer (SIVT) and Integrated Visual Augmentation System (IVAS) engineering team; STE Experiential Learning for Readiness (STEELR) S&T applied research; and US Navy learning engineering applied research.

Kevin Hellman is the Lead Capability Developer for the STE Information System (IS) and Training Management Tool (TMT) at the Combined Arms Center (CAC) Proponent Office in Fort Leavenworth, KS. He was a core and original member of the STE CFT, co-creator of the STEELR research project, a core member of the SiVT-IVAS engineering team, and is a former US Army Cavalry Scout. He has 19 + years' experience in Data Management, Process Management, Data Conversion Services and Data Fusion in both industry and government. He has lead data efforts at the Tactical, Operational and Joint levels for the US Army. Mr. Hellman holds a Bachelor's degree, Master's degree and an MBA-Finance.

Robby Robson is a researcher, entrepreneur, and standards professional who is co-founder and CEO of Eduworks Corporation, a member of the IEEE Standards Association Board of Governors, and Principal Investigator on the STEEL-R project. He has made contributions to areas ranging from semi-algebraic geometry and computational number theory to web-based learning, digital libraries, and applications of AI to learning, education, and training. His recent efforts focus on competency-based approaches to talent management and experiential learning and providing effective and equitable reskilling opportunities for current and future workers. Robby holds a doctorate in Mathematics from Stanford University.

Shelly Blake-Plock is president and CEO of Yet Analytics, Inc. As principal investigator for multiple xAPI-related R&D projects sponsored by the Advanced Distributed Learning Initiative, he has contributed to the development of new capabilities in xAPI Profile design and validation, domain-based query, and data simulation. He holds a Master's of Science degree in Education from The Johns Hopkins University.

Michael Hoffman is a senior software engineer at Dignitas Technologies and the technical lead for the Generalized Intelligent Framework for Tutoring (GIFT) project. For almost a decade he has been responsible for contributing too and managing the development of GIFT, meeting community requirements, and supporting ITS research. Michael contributes support for the GIFT community through various mediums including the GIFT portal (www.GIFTTutoring.org), annual GIFT Symposium conferences, Expert workshops on ITS related topics *and technical exchanges with* CCDC *and their contractors.* He holds a Master's degree in Computer Science from the University of Central Florida.

Kevin Gupton is a systems architect in the M&S Engineering Group at the Applied Research Laboratories, The University of Texas at Austin (ARL:UT). He has over 18 years' experience in enterprise system engineering, data modeling, and knowledge management in the simulation domain. Projects to which he has provided system architecture and data management expertise include STE, SiVT, IVAS, STEELR, Joint Federated Common Data Services, Integrated LVC Test Environment (ILTE), and NATO MSG-164 M&S as a Service. Mr. Gupton holds a BS in Mathematics and a MS in Computer Science from Texas A&M University.

Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy

Benjamin Goldberg
US Army DEVCOM Soldier Center
Orlando, FL
benjamin.s.goldberg.civ@mail.mil

Kevin Owens, Kevin Gupton
Applied Research Laboratories University of Texas at Austin
Austin, TX
kowens@arlut.utexas.edu; kgupton@arlut.utexas.edu

Kevin Hellman
US Army TRADOC Proponent Office Synthetic Training Environment
Fort Leavenworth, KS
kevin.m.hellman2.civ@mail.mil

Robby Robson
Eduworks, Inc.
Corvallis, OR
robby.robson@eduworks.com

Shelly Blake-Plock
Yet Analytics
Baltimore, MD
shelly@yetanalytics.com

Michael Hoffman
Dignitas Technologies
Orlando, FL
mhoffman@dignitastechnologies.com

INTRODUCTION

The U.S. Army is making significant investments in the modernization of its core Synthetic Training Environment (STE) capability. It will leverage the state of the art in gaming and simulation technology to provide a point of need collective training solution that immerses teams of soldiers in realistic scenarios that target core skills and challenge team dynamics. To ensure this modernization effort facilitates its intended function in supporting training and readiness needs, a robust training management capability is required. In its defined end-state, the STE will leverage intelligent tutoring and adaptive training services to assist in the preparation and delivery of scenarios that addresses a given teams or units training needs. This is facilitated through robust adaptive instructional functionality that leverages adaptive instructional services that use data to drive real-time assessment and training management with a goal of maximizing training benefit (i.e., acquiring skill that transfers to an operational setting).

In this paper we present the STE Experiential Learning for Readiness (STEEL-R) data strategy, and its underlying architecture used to create a state-of-the-art intelligent tutoring capability designed around a training readiness through STE use case. Ultimately, STEEL-R is being implemented to facilitate and better inform requirements within STE's Training Management Tools (TMT) that focus specifically on intelligent tutoring for teams, and how data will drive and optimize workflows that associate with the preparation and delivery of a training event. This involves establishing a distributed architecture that enables multiple layers of data inferencing for the purpose of informing both real-time and persistent long-term assessment representations, and using these assessment models to drive training strategies that target individual-, team-, small unit-level competency acquisition. With this objective in mind, STEEL-R is being implemented using existing standards and best-practices linked to experiential learning, game-based intelligent tutoring and competency-informed performance tracking. The result is a referential architecture that supports data translation and contextualization at the human performance level, providing a re-usable strategy that converts raw data into metrics, applies metrics with defined criteria to measure/assess performance, and persistently and strategically stores evidence-based assessment data over time to track skill development and infer task proficiency.

In the following sections, we will review the theoretical underpinnings of experiential learning and adaptive training informing this work, followed by a detailed breakdown of the STEEL-R architecture implemented and its underlying functional components. We will then provide a use case of the architecture and data strategy being applied in the infantry squad domain, with a focus on its application across STE's Plan, Prepare, Execute, and Assess training management workflows, while highlighting its intended extensibility across an unlimited set of team and task structures. Before we focus on STEEL-R, it will be important to review the core STE-Information System (STE-IS) tools and methods under development to drive this future synthetic training infrastructure.

STE-Information System (STE-IS) Components

STE-IS is comprised of three foundational components: (1) Training Management Tools (TMT), (2) Training Simulation Software, and (3) One-World Terrain. The core technologies of STEEL-R will be housed within the TMT infrastructure. While TMT comprises multiple logistical functions supporting the delivery of training, STEEL-R pays specific attention on how STE will utilize data to optimize training outcomes, while remaining agnostic to specific training modalities, environments, and data sources. Ultimately, the goal is to create a service-oriented TMT infrastructure that applies tools to automate exercise design, facilitate scenario execution with adaptive functions, and enable objective evidence-centered assessment across multi-modal data sources. These data are then gathered in "data lakes" for further analytics, with services being designed to infer competency over time.

These training management objectives create explicit dependencies with the other STE-IS components. For instance, the STE Training Simulation Software (TSS) represents the underlying simulation/game engine and set of interface modalities a trainee or team will engage with during training execution. This accounts for the inputs and controls a trainee has during run-time, the behaviors and models of computer generated forces, and all of the data/telemetry produced as a scenario is completed in support of a specified training objective. The TMT requires an ability to consume these raw data sources and apply algorithms trained to generate automated assessments; however, most TSS solutions were designed without this requirement in mind. To achieve its desired end-state, the TMT must have a direct socket connection with TSS to consume relevant data, along with adaptive training services that translate these sources into meaningful measures of performance. These metrics can be used to inform real-time coaching and adaptation strategies, along with providing granular evidential statements on ability that can be used to monitor performance and progress over time. STEEL-R is investigating the intersection between TSS and TMT, with a focus on building definitive best practices on how to establish data-driven evidence during a STE exercise. Another element for consideration is that the TSS will operate across varying modes of immersion and interactivity, leveraging virtual, augmented, and mixed reality mediums to create a more dynamic and realistic experience. The TMT must be able to monitor engagement across all supported modes through data interoperability, enabling cross-resource inference.

The third STE-IS component is One World Terrain (OWT). OWT is being implemented to provide a realistic and extensible representation of the real-world, providing terrain models and formats that are applied across all STE modes of interaction. It provides a set of tools, methods, and workflows to generate synthetic terrain databases through on-location data collections, enabling up-to-the-minute representation of the environment and mission or task that will be executed within. From a STEEL-R perspective, OWT is critical for establishing context and logic around a set of otherwise generalizable assessments that are defined at the competency level, but require OWT to enable calibration of specific assessments based on placement and configuration of scenario elements. In theory, with all STE-IS components, we have a training capability that enables the delivery of immersive scenarios on relevant terrains with tools to guide the training experience. In practice, a training strategy based on learning theory is required to guide the TMT. With this in mind, we are implementing a competency-development informed approach based on tenets of experiential learning and building expertise over time and through repetition (Kolb, 2014; Ericsson, 2006).

Competency-Development Strategies and Synthetic Training Environments

In its most basic form, a competency represents a human performed capability comprised of Knowledge, Skills, Abilities, Attitudes (KSAAs; Fletcher, 2005; Walcutt & Schatz, 2019). A competency-development strategy is based on the idea that sound instructional design can be applied to define what capabilities a team and individual requires to be successful at a task and/or role. Through this definition process, a competency-development strategy is applied to create and align experiential resources to the representative KSAAs across all defined and interconnected competencies (i.e., competency framework). Most important to this development strategy, these definitive performance capabilities can be measured, and monitored over time, using objective assessment techniques and mathematical modeling to inform a competency state. Competencies provide a structure to "bin" data into specific categories of performance and to pre-define a stable standard for how data should be collected and classified. For further breakdown on competency development strategies and underlying theory, see Owens (2021) and its application with small-unit teams, see Owens, Gupton, Hellman & Goldberg (2020).

Competence is measured as a vectored state of human capability and probability; in other words, a state of what one can do now and a probability of how one will perform in the immediate future. A competence state is based on three data-informed scalars representing: how-well, how-hard and how-often an individual or team has performed a task or set of competencies (Owens 2021). These scalars are calculated from outcomes across multiple training exercises and are dependent on granular assessments at the task, step, and behavior level, with associated metadata to track characteristics that inform the three scalars above (Gilbert, 2007). Measures of competency occur at two stages: once at the point of performance which result in one of three levels: *below-expectation*, *at- expectation* or *above-*

expectation (i.e., expert) as shown on the left side of Figure 1. The three-levels can be defined for each KSAA element represented in a competency, and are applied to show the trend of an actor's capability. It also informs staff and leadership if a team or individual are ready for higher or lower levels of difficulty in future experiences.

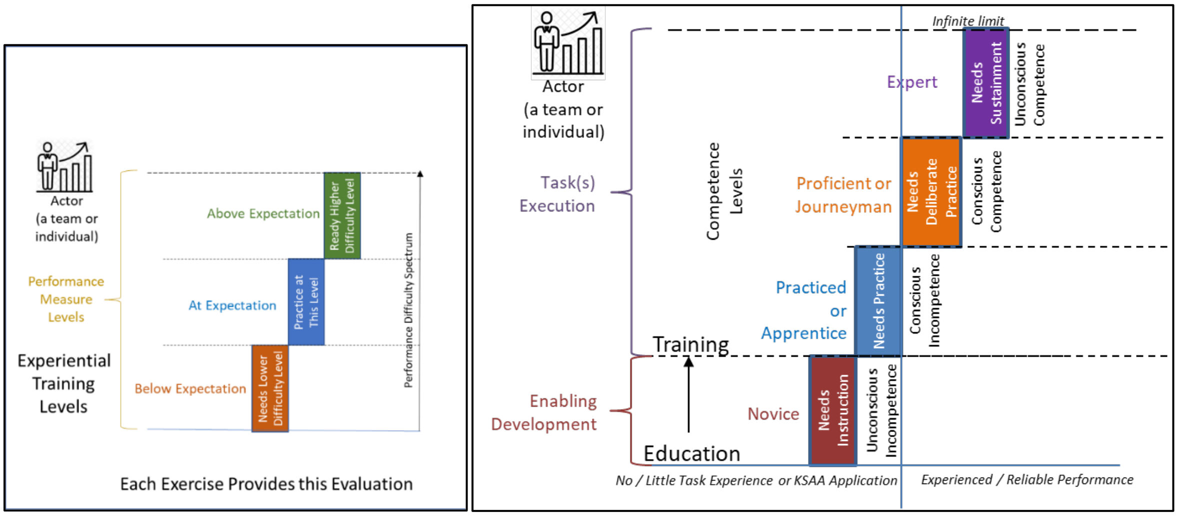

Figure 1. Evaluations at the individual exercise and longitudinal trends.

The second stage of competence measurement is persistent in nature and based on a longitudinal math model (discussed in detail below) using data collected across multiple exercises, modalities, and sessions. A key purpose of this stage is to objectively assert a level of current and future capability (i.e., competence) to perform a given task, based on empirical – and inspectable – past performance data. The levels of competence shown on the right side of figure 1 are an amalgamation of existing competency-based learning models like the Dreyfus' model of skill acquisition (Dreyfus, 2004), Miller's model of clinical competence (Miller, 1990), as well as the more fringe but sound Broadwell's model (Broadwell, 1969). These models are applied in a logical manner that fits with the existing Army operational training evaluation structure (FM 7-0), which currently uses three-levels of *proficiency-based* competence ratings: *untrained*, *practiced*, and *trained*. This competency-based development strategy will continue to use these levels, but they will be expanded to include more appropriate and internationally accepted terms in the spectrum of competence (see right side of Figure 1). In addition, this strategy will recommend expanding this overall spectrum of proficiency with the addition of a fourth level of competence termed *expert.* With an intent to establish a data-driven TMT capability informed by competency-development strategy, the STEEL-R data strategy was created.

STE EXPERIENTIAL LEARNING FOR READINESS (STEEL-R)

The STEEL-R effort is focused on explicit development and integration activities to create a functional proof-of-concept that supports the TMT and competency-development requirements introduced above. It is based on a strategy of collecting context-rich human performance data at the training point-of-need, and storing these data over time to objectively measure and infer proficiency across a set of defined competency structures. In the following sub-sections, we provide a detailed breakdown of the STEEL-R architecture and its core components. We then introduce experiential-derived competency frameworks that will guide the integration activities across supporting data services.

Functional Architecture and Components

The STEEL-R architecture (see Figure 2) is realized through the integration of several open-source technologies. When combined together these tools create a persistent data ecosystem capable of capturing and logging granular metrics of performance at the training edge, establishing rich cloud-enabled data lakes housing aggregated evidence of performance, and using Competency Management System (CMS) techniques, mathematical modeling and machine learning to ultimately measure team and individual competencies across all representative tasks and recorded training events.

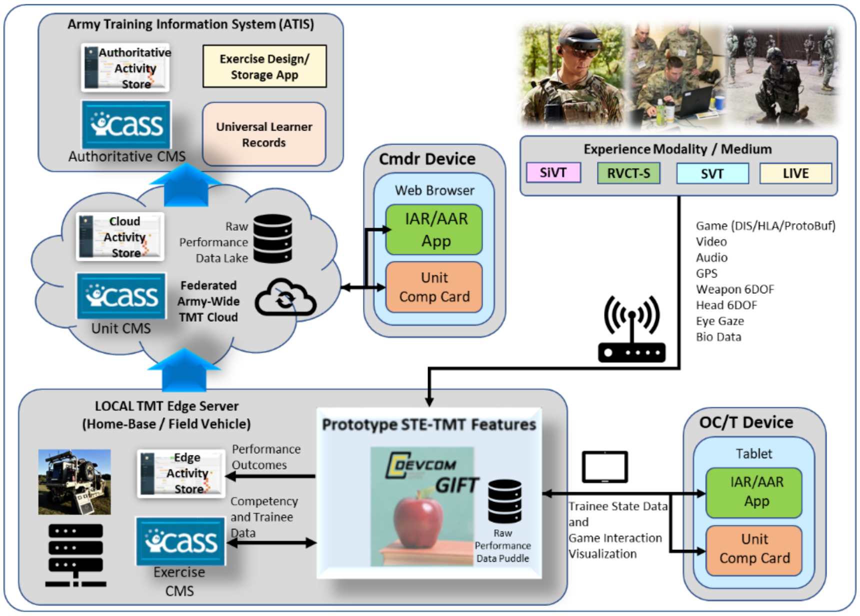

Figure 2. STEEL-R Functional Architecture

Specifically, we are leveraging the U.S. Army's Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Goldberg, Brawner & Holden, 2012) to sync TMT with a STE training modality to capture data in real-time to create evidential statements of performance. With an ability to convert raw data into meaningful assessments, we are leveraging the Advanced Distributed Learning (ADL) Initiative's eXperience Application Interface (xAPI) and Learner Record Stores (LRS; ADL, 2018, 2021b) to provide a mechanism to report out performance in a standard format for long-term storage, and extending ADL's Competency and Skill System (CaSS; Robson, 2019) to apply an xAPI informed math model used to infer individual and team competencies. This integration of components creates a cloud-enabled, hybrid TMT network that enables streamlined transfer of performance-derived metrics to a series of LRSs for direct use by CaSS' competency modeling services. This will lead provide a connection with STE TMT and Army authoritative data sources that will use these training data to better track readiness across the force. Each component will be discussed in further detail below.

Generalized Intelligent Framework for Tutoring (GIFT)

The Generalized Intelligent Framework for Tutoring (GIFT; Sottilare et al., 2012) is an open-source, domain-independent, modular, service-oriented architecture used to author, deliver and evaluate Adaptive Instructional Systems. GIFT was designed to provide reusability across domains, training applications and technologies. It also works as a backbone for the integration of disparate data streams across different sources, including biometric sensors, training system state information, observer input and historical records. The architecture consists of several core modules that compartmentalize the operation of synchronizing, classifying, interpreting and applying learner state during training, creating Adaptive Learning Effect Chain (ALEC, Sottilare, Ragusa, Hoffman & Goldberg, 2013). Figure 3 depicts these core modules and the data that is transferred between them. The figure also shows how long-term learner records can be exchanged with an LRS to initialize learner state upon starting a session and provide evidence of training results in GIFT to other external systems not shown (e.g. CaSS).

In support of STEEL-R, GIFT is used to establish the task and assessment structures that will be applied during a training event, and used to capture evidential statements of performance. GIFT is being applied as a direct translation layer, using its real-time assessment functions in a hybrid tutoring approach (Goldberg, Hoffman & Graesser, 2020), to create a set of formative (i.e., real-time observation) and summative (e.g., post-task summation) performance metrics across all of the objectives and KSAAs measured during task execution. GIFT provides a re-usable schema and set of authoring tools/workflows to quickly populate an assessment model that will build this performance context. With the ability to translate multi-modal data and interaction into meaningful evidence, STEEL-R applies xAPI as a mechanism to report out this information.

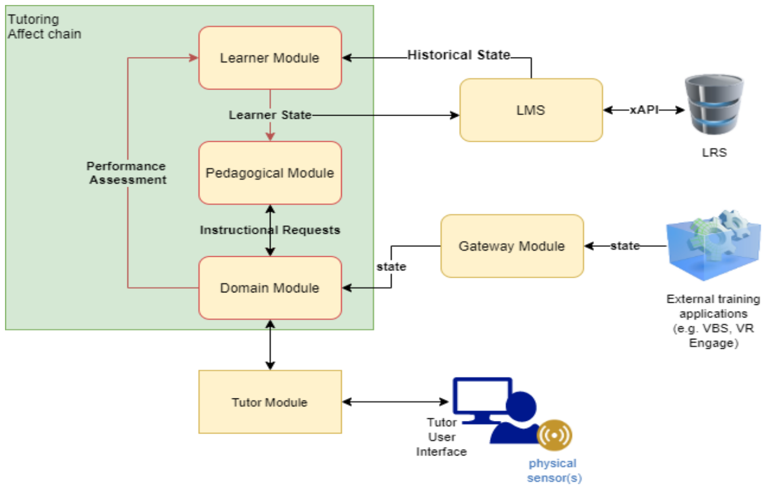

Figure 3. GIFT real-time assessment components that contextualize data in evidence-based performance statements.

xAPI and Learning Record Stores

The Experience API (xAPI) enables the interoperable exchange of activity data from a training resource and is used to describe a learner's behavior and performance (ADL, 2021b). The data follows strict adherence to a standard format and is validated by a Learning Record Store (LRS). The LRS is essentially the server-side abstraction of the xAPI data specification.

STEEL-R supports the model of the ADL Initiative's Total Learning Architecture (TLA; ADL, 2021a; Walcutt & Schatz, 2019) with regard to the design flow of xAPI data. First, GIFT emits xAPI statements that describe granular aspects of a learners engagement, including event-based metrics describing an individual or teams affective state, cognitive state, and activity within the learning experience. All of this data follows the model of a xAPI Profile (see GitHub) which has been designed to represent the range of learner engagements that can be measured by GIFT within the execution of a scenario. The xAPI Profile itself includes machine-readable documentation of the vocabulary, patterns, sequences, and concepts included in the range of learning experiences present in the scenario. In STEEL-R, data flows from a Learning Record Provider (such as GIFT) into a Noisy LRS where the data is validated and stored. The benefit of using Noisy LRSs at the edge of a federated system is in the management of data collection closest to the source—therefore a federated and distributed learning ecosystem could contain several LRSs, each responsible for the validation of data either from a specific source such as an intelligent tutor or a learning experience platform or from a specific region or authority such as from several locations or from different agencies.

Data from the Noisy LRS is forwarded to a Transactional LRS governed by the Master Object Model xAPI Profile. This profile acts as a statement filter and provides the ability to streamline the data from any of several edge data sources into a single source of truth with a common vocabulary. Data in the Transactional LRS may be consumed by a Learning Record Consumer such as the Competency and Skill System (CaSS; explained in detail in next sub-section). In STEEL-R, CaSS consumes xAPI statements which have recorded activities that are mathematically applied for the purpose of asserting competencies.

The data emitted as xAPI from GIFT follows the GIFT xAPI Profile and (as described in Blake-Plock, et al., 2021) includes the following:

- As a domain session is requested, data is cached for use in the creation of the xAPI statements
- As a domain session starts, an xAPI statement is generated which identifies the user and the course selected
- As a knowledge session is created, one or more xAPI statements indicate that either a session host created and started a session or that upon creating the session lobby, other users joined the lobby and then the host started the session
- As an updated request passed through a knowledge session, the team position of the session member provides information for use in the creation of xAPI statements
- The knowledge session begins for the team and statements are emitted

- The learner state is derived from the relevant GIFT components regarding: cognitive state, affective state, and performance state; user interaction within the STE causes an update to these attributes
- As formative assessment is completed, a request is made to publish the lesson score and summative results are recorded as xAPI statements
- The session is closed, an xAPI statement is emitted indicating that the user has exited the course

The xAPI statements contain information about who a team member is, the structure of the team that they are operating within, and who their counterparts are on that team. They provide mission results for individuals and capture the human performance evidence necessary to assert competencies, including affective, cognitive, and performance states. The statements also describe the training environment/experience modality used to establish confidence around the assessment for building assertions. The activity they have tracked can be represented as trends as well as be used to instrument the analysis of short-term, long-term, and predicted assessment.

Competency and Skill System (CaSS)

The primary function of CaSS in STEEL-R is to gather evidence of competency from multiple events/sources and to store this evidence in a standardized internal format known as an assertion (Robson & Poltrack, 2017). An assertion has an agent, which is the person, organization or system making the assertion; a source of evidence or data used by the agent (which could be the agent itself but is more often an assessment, scenario, practice session, credential, or some other source); an entity and a competency about which the assertion is made; a timestamp that indicates when the evidence was produced or the assertion was made; and a verb which is either "has", "does not have", or "attempted." Optional parameters include a confidence between 0 and 1 that indicates the confidence the agent has in the assertion; a link to evidence that is the basis of the assertion; an expiration time; and a decay function that indicates how confidence diminishes as a function of time. A typical STEEL-R assertion collected by CaSS is (in human-readable form):

> "CaSS asserts at 11:34:23 on 17-March-2021 with confidence .75 and source GIFT that Fireteam A has the competency 'enter and clear a room' based on (link to replay file from Battle Drill 6, Scenario 5, run on 17-March-2021)."

In STEEL-R, a critical part of this CaSS function is generating assertions from GIFT-generated xAPI statements and xAPI statements from other Training Aids, Devices, Simulations and Simulators (TADDS). When CaSS detects a new xAPI statement in the Transactional LRS that is part of the STEEL-R architecture, it decodes it to create an assertion. As a competency management system, CaSS stores and manages competency objects that represent competencies, competency frameworks, KSAAs, and related types of objects that define what a person, team, or organization – called entities in CaSS – knows or can do. Competency objects in CaSS can have a name, unique ID, description, relations to other competency objects, and one or more types. In STEEL-R, type is used to label competencies as KSAAs, as discrete "tasks" and "roles" (which means a team/individual has the ability to successfully perform the task/role), and evidence-centered "performance metrics" that are used to calculate a competency state. The object type is also used to identify whether a competency is an individual or team competency. As an example, a performance metric represented in GIFT such as "minimize collateral damage" for a battle drill becomes a competency object in CaSS with type "performance metric", type "team competency," and description "minimizes collateral damage when engaging in a battle drill."

In the case of GIFT, xAPI statements report performance on tasks and concepts. They include an assessment as to whether the performance was below, at, or above expectations and a link back to the scenario where that assessment was made. CaSS maintains a mapping between GIFT tasks and concepts and CaSS competency objects and uses this mapping, together with the performance level reported by GIFT, to create an assertion. For example, suppose that "communicates clearly with teammates" is an individual competency object in CaSS (of type KSA), that "team leader gives order to withdraw" is concept in GIFT, and that a soldier has the position of team leader in a GIFT battle drill scenario. If GIFT reports that this concept was performed below expectation via a xAPI statement, CaSS will see this xAPI statement and translate it into an assertion that the soldier does not have the "communicates clearly with teammates" competency object. The confidence associated with this assertion may be arbitrarily set to 1 or may be set to a smaller number to reflect the scenario being monitored by GIFT. As another example, if the xAPI statement indicates performance at expectation, the assertion may be positive, but with lower confidence than had the performance been above expectation.

Competencies can be asserted against at different levels. In STEEL-R, for example, a competency can currently be untrained (U), practiced (P), or trained (T). In CaSS, different levels of a competency are represented as distinct competency objects. This is counter to the way that people think about levels but facilitates having different performance criteria and, more importantly, different enabling or required competencies for different levels. Thus, the task "Enter and Clear a Room," which is a team competency of type task in CaSS, is represented as three competencies, one for each level. In CaSS, competency objects can also have relations to other competency objects and to external resources such as documents that define or clarify the definitions and training resources that can be used to acquire or assess the competency. The most important relations in STEEL-R are relations that indicate whether one competency requires another, enables another (in the sense of instructional design theory, see e.g. (TRADOC, 2021), or broadens or narrows another, which is discussed below.

The third function of CaSS in STEEL-R is to compute various experiential learner models. STEEL-R will eventually have three such models – a state model that reflects which competencies an entity possesses and includes an indication of how much each competency has been practiced; a predictive model that estimates the probability that an entity will successfully perform or demonstrate a competency given the opportunity to do so; and a training model that is used to estimate the probability that an entity will acquire a competency (or a new level of a competency) by engaging in a particular training scenario in a particular TADDS. It is a longer-term goal of STEEL-R to generate the predictive model that helps determine when to advance from TADDS to live training and to generate a training model that can be used to select or configure scenarios to increase the efficiency and efficacy of training. Currently STEEL-R is focused on computing a state model that O/CTs and commanders can use to make more informed evaluations and training decisions. This model and the associated computations are discussed below in the Training Strategy section.

Building Experiential Competency Frameworks

The STEEL-R functional components integrated within its architecture were originally designed to be generalizable in nature. When applied in this new context, establishing competency frameworks adhering to experiential learning principles will be critical to guiding its implementation. As opposed to traditional static task or learning-objective structures, a competency structure is used to produce a "living", cloud-based, and machine-readable set of standards that enables an organization to define reusable performance criteria (Owens, Gupton, Hellman & Goldberg, 2020). STEEL-R will define competency structures that account for granular steps, processes and procedures that require consistent application for an actor to be gauged as proficient. Measures aligned to these KSAA elements will be produced by GIFT using an evidence centered design approach, with a focus on using real-time interaction data to objectively monitor these critical behaviors across all task interactions. In this instance, specific CaSS objects will be defined that establish a hierarchical task tree structure enabling granular representation and measurement of task characteristics, along with rules to aggregate and roll-up those metrics for inferring higher order competencies. These associations are being defined in mathematical algorithms trained to objectively infer competence-states across teams. Figure 4 provides an illustration of each of the key competency structure elements being represented in CaSS.

STEEL-R is also working to implement the first competency structures for modeling and managing the expertise of occupational teams. When considering the variables that makes one team better than another, this expertise rating is not only based on performance linked to task and role execution, but also accounts for competencies linked to teamwork and cohesion (Patton et al., 2018; Sottilare et al., 2018). This will provide a means to track if task objectives are being met, as well as how efficiently the team works together while achieving an objective. To support this extension in the competency modeling landscape, building evidence (via xAPI) of teamwork is critical, with heavy reliance on establishing task events and triggers that explicitly measure teamwork dimensions. With the infrastructure in place, future work will focus on defining these team level frameworks that will used to infer teamwork across all events.

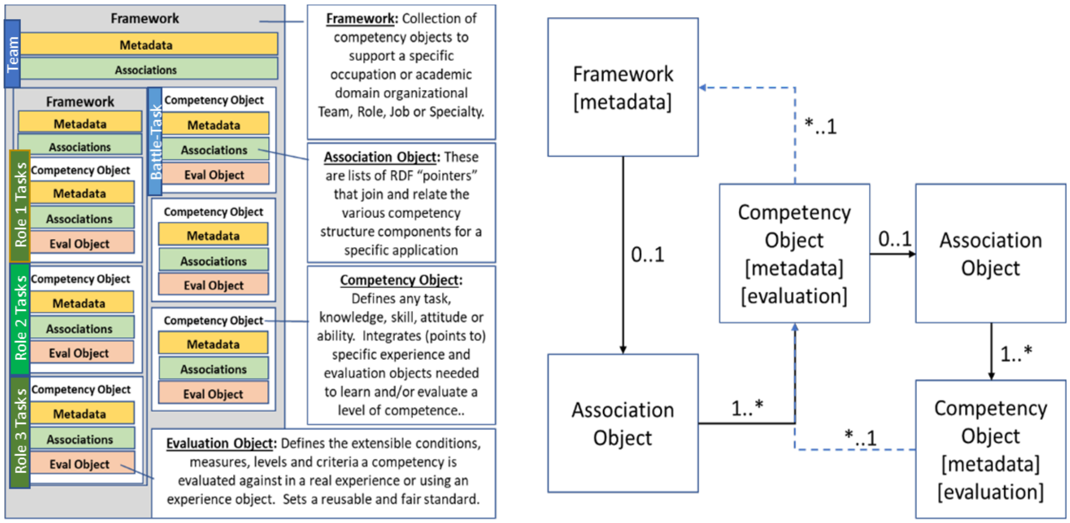

Figure 4. IEEE Competency Based Structure

At the use case level, it's important to establish these framework across all critical KSAA associations. As an example, a defined competency object in CaSS at the infantry squad level is 'Engage Targets with an M4' (see Figure 5). As represented above, this competency will associate with specific role frameworks that align to this weapon system. In this instance, we are defining this CaSS object as a framework that is comprised of underlying KSAAs. This highlights that 'Engaging Targets with an M4' is not directly observable, and that it requires multiple congruent skills applied in unison for skillful application. In addition, each KSAA can be directly measured to assert if an individual is able to consistently perform this competency when operationally required, while providing granular diagnostics on all associated skills. In this instance, each constituent skill is represented as an independent CaSS object with direct steps/processes/procedures/behaviors that can be directly measured/assessed against. The goal is to establish evidence during task execution that is used to infer competency.

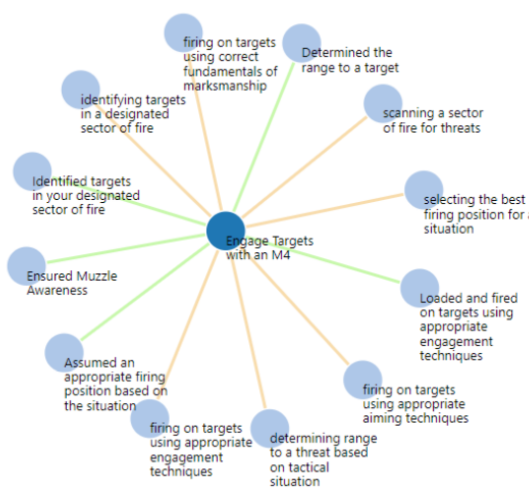

Figure 5. Visual representation of KSAAs aligned to the Engage Targets with an M4 competency

STEEL-R TRAINING STRATEGY

With an ability to establish context rich performance measures from a training event, and being able to persistently track those metrics against competency structures, it is important to establish a strategy that uses these data functions and models to assist STE users in defining training objectives with a focus on accelerating competency development. This requires careful interpretation of tracked xAPI statements to answer meaningful questions; questions directly related to what competencies an individual or team possesses, what competencies require attention, and what types of scenarios and conditions will support further progression in development. To inform this capability, an experiential math model is being established to create a proficiency index based on the experiential evidence captured using GIFT.

Inferring Proficiency through an Experiential Math Model

Most instructional systems, including intelligent tutors, view the state of having acquired a competency or skill as a binary state (i.e., the learner either has it or does not) that can be assessed within the context of the instructional resource. These are generally point-in-time assessments that come from a single source that fails to consider performance over time and under varying real-world conditions. As pointed out in Robson (2019), this is not ideal even for cognitive skills (e.g., math skills), whose value greatly diminish if they cannot be retained and applied in context. To account for this more longitudinal, nuanced, and sophisticated approaches are needed to represent and evaluate the state of a learner's skills requiring repetition and experience to master and achieve consistent performance.

This persistent modeling is accounted for in systems such as ACT-R (Anderson et al. 1996) and its successors. A recent contribution of interest associates with performance prediction models (Jastrzembski and Gluck, 2009), which introduce variables that impact skill acquisition and retention, including forgetting curves (Averell & Heathcote, 2011), activation thresholds, and spacing effects (Pavlik & Anderson, 2003). The models used in STEEL-R, as implemented in CaSS, consider similar parameters, but with key enhancements:

(A) The STEEL-R models used to evaluate the state of a learner are designed to consider data from multiple training sources that support the same competency structures (e.g., scenarios within Virtual Battlespace 3, the Integrated Visual Augmentation System, and the Squad Advanced Marksmanship Trainer). These models can weigh the corresponding assertions differently and use them all to infer the state of a learner.

(B) The state of a competency includes a quantity (called a *practice score* in STEEL-R) that reflects the amount of practice an entity has had in performing or demonstrating a competency. The practice score counts the number of attempts at performing or demonstrating the competency and weights them by a forgetting function, so recent attempts count more, and by a spacing function, so repeated attempts within a short period contribute less.

(C) CaSS can use relations among competencies to make inferences and apply rollup rules that influence states. These include required and enabling relations discussed earlier plus relations among team and individual competencies. STEEL-R tracks which soldiers are in which positions and which units play which roles in a scenario. Given a team competency C, relations and rollup rules can specify that a specific position or unit must possess one or more other competencies for the team to possess *C*.

When computing the state of a soldier or unit with respect to a competency, CaSS examines all assertions about that soldier or unit. The state computation uses this data to compute an assertion score that combines all assertions into a single number between -1 and 1. The state is then computed based on practice scores, assertion scores, and rollup rules. For details on the type of computations and models used, see Robson et al. (2021).

STEEL-R USE CASE: INFANTRY SQUAD

In this section we will provide a concise use-case highlighting the role STEEL-R could play within the context of an infantry squad's training progression. The use-case follows a modified form of the Army's Plan, Prepare, Execute and Assess (PPEA) model, with dedicated activities and workflows leading up to an experiential training event. To enable this use case, a few assumptions are in play: (1) an infantry squads' competencies and frameworks exists in CaSS that are aligned to the tasks a squad will exercise in the training event, (2) there is a library of existing experiential training support packages (TSP) that are designed with multiple experience-events that prompt the squad and individual tasks and KSAAs, and (3) there are task and competency evaluation assessment rubrics that auto-program the GIFT real-time assessment based on the tasks and competencies a TSP targets. Future work will focus on how STEEL-R supports the TSP design, which will be discussed in the conclusion.

Training Plan Development

In contrast to traditional long-range or short-range training planning, especially as it relates to reserving and preparing for live training, Squad and Individual experiential learning with the STEEL-R capabilities will allow for more broad and responsive training opportunities at the point-of-need. This process begins with the concept that home-base or deployed units will now have continuous online feedback and status in the form of a competence "leaderboard" that will provide small-unit leaders and Soldiers their progress and probability of proficient performance for a given team, role or task. The idea being this continuous feedback will motivate Platoon leadership, squad NCOs and individual Soldiers to initiate more short-concise experiential training events as deemed necessary.

When a training event is desired, STEEL-R will enable unit staff to digitally select the competence item they wish to train on. In future versions this will automatically assist with front-end logistics, providing a list of the best TADSS (i.e., STE-IS training modalities) to meet objectives, the closest location to the capability, along with its availability. It will also provide a list of available TSPs designed to support the selected competency, as well as provide an option to schedule a training event. From this point, the trainer or unit leadership can begin preparing for the training event based on availability and/or, design a new TSP tailored for their specific training needs.

Scenario(s) Preparation

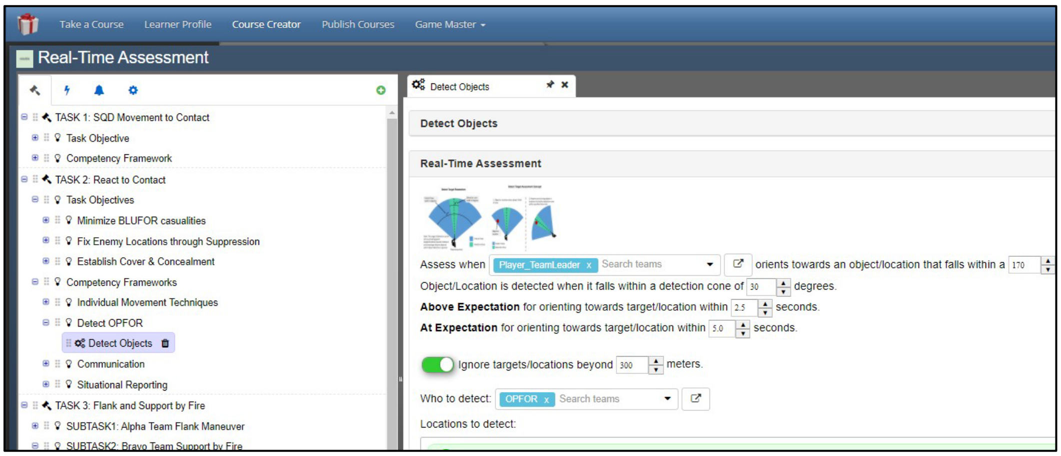

Figure 6. GIFT Real-Time Assessment (i.e., Domain Knowledge File) Authoring Tool

With a defined training plan, unit staff have specified a set of task objectives and competency structures that will guide preparation activities. Existing scenarios will be imported into STE TMT that match task and competency structure specifications. In other words, scenarios with explicit task and event pairings aligned to KSAAs of focus are prepared for delivery within the STE-IS. In the context of STEEL-R, preparation activities are focused on configuration and calibration of assessment models and underlying metrics linked to the scenario events and triggers. GIFT's real-time assessment authoring tool (see Figure 6) creates the task/metric schema, and users' edit/modify/create measurement parameters that dictate how tasks and KSAAs will be assessed. This involves designing evaluation triggers to fixed geographic locations (points, lines or areas), objects or entities that a trainee will encounter. These triggered evaluation events will activate GIFT's real-time assessment logic. Initially, these schemas will be directly informed by doctrine, with an ability to modify and extend based on evolving standards and operational characteristics. Each node represented in the task tree is used by the GIFT's xAPI profile to create evidential statements based on the below-, at-, and above-expectation inputs. This provides a flexible and extensible tool to establish any performance metric.

Scenario(s) Execution

Following preparation workflows, it is time to execute the training. In the execution phase, the trainees are interacting with the training environment which can include any combination of live, virtual, or constructive applications. Each system provides some level of state information during a session that can be combined with one or more physiological sensors to determine and track performance in real time. Performance is measured based on the structure and rules defined in the aforementioned assessment models established during the prepare phase. Ideally this assessment would be calculated automatically with a high level of objectivity, data precision, and environment fidelity with instructor trust across the tasks being performed, but this can't always be guaranteed. Often times the training system might not provide STEEL-R with adequate details of the unfolding events and GIFT may lack the necessary interpretation of human behavior to completely replace an instructor's insight. To alleviate this issue, OC/Ts (Observer Controller/Trainers) can use the GIFT Game Master application to actively monitor the tasks modeled in GIFT for adaptive exercise control (see Figure 7).

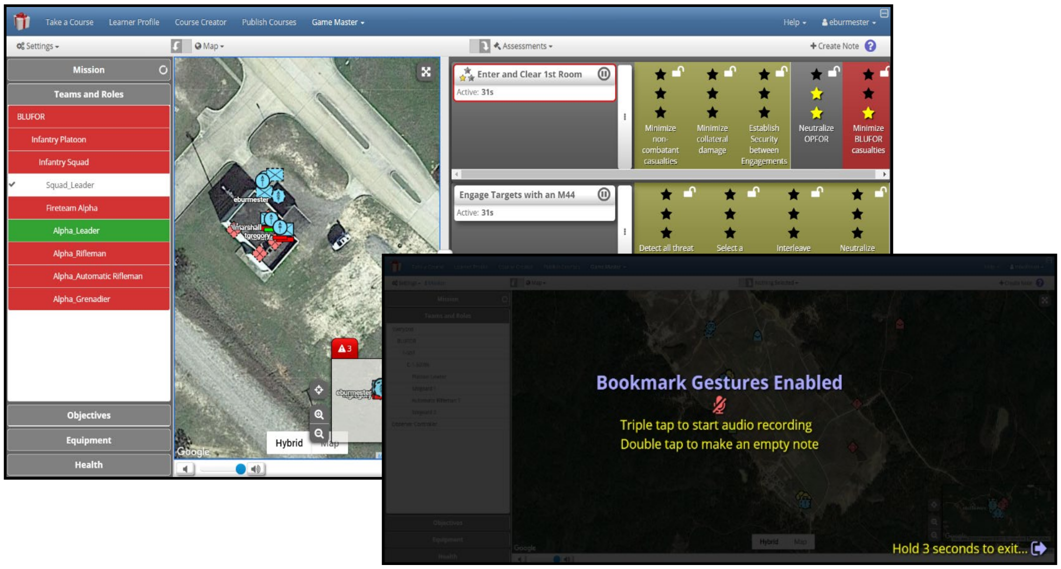

Figure 7. Upper Image screenshot of the GIFT Game Master Active Session showing the mission, map, and assessment panels with hybrid assessments. Bottom image Game Master Bookmark Gesture mode used to create real-time annotations and audio capture.

The Game Master delivers a customizable user interface for the OC/T to organize panels in a dashboard style fashion. There is an assessment panel that exposes the configured assessment model for each active and completed task. A map can be shown that contains 2525 military symbols overlaid with state information for all entities in the scenario. Another panel provides access to a list of scenario injects and coaching strategies the OC/T can manually execute. In addition, one of the most important features that the Game Master provides is the ability to establish in-situ bookmarks or notes during training (see Figure 7). These bookmarks are time synched with the other data streams being collected. Bookmarks can either be in the form of text or an audio recording. To facilitate ease of use the user can enter bookmark gesture mode to create bookmarks of both types by simply tapping on a touch screen device such as a tablet. These UIs provide a hybrid intelligent tutoring capability (Goldberg et al., 2020) that interfaces relevant staff with STEEL-R's adaptive instructional logic, putting the human on the adaptive loop.

Training Assessment

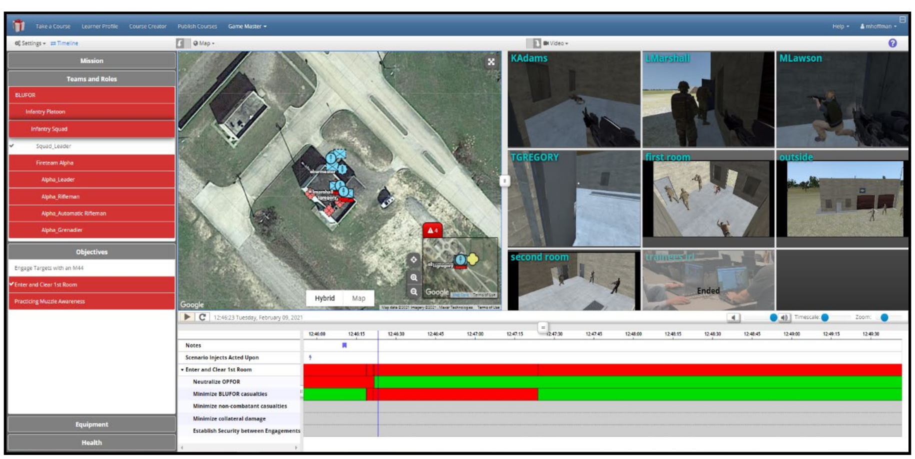

Figure 8. Screenshot of the GIFT Game Master Past Session user interface showing an explorable time-line experience with the time-synchronized mission, map, and video panels visible.

Once scenario execution has concluded, the training experience continues in the assessment phase. First, unit staff would use the STEEL-R tools to review the scenario that was just completed, with a focus on inputting and updating all associated assessments at the individual and team level. In the Game Master Past Session experience, unit staff will explore a time line of events across multiple panels and visualizations with the ability to play, pause, rewind and skip ahead to any event of interest (Figure 8). During this playback exploration the user might decide to provide an assessment that was not automated at run-time; or they can also review and disagree with an automated assessment, providing an ability to modify the evidence captured from that event that will be used to assert competency. Both of these actions not only update the learner state for that concept, triggering an update to the rendered timeline, but it also invalidates any previously recorded xAPI statement in the connected LRS. One or more new xAPI statements will be created to capture each change made. What results is a record of all statements that were created during and after a session. Each statement contains who was responsible for the assessment, what was being assessed and who was being assessed at a minimum. Following the initial playback phase to complete the assessment inputs, the Game Master playback is then used to facilitate an interactive and data-supported After Action Review. This enables structured reflection across all tasks and events, with direct and annotated observations used to assist in conversation.

Competency Assessment

CaSS computes the state of each soldier and unit with respect to each competency in a framework. This is based on applying newly captured xAPI evidence to the proficiency math model for direct updates across all tracked competency objects. In STEEL-R, these and related data are displayed to a unit staff in a *proficiency dashboard*. The purpose of this proficiency dashboard is to visualize competency levels and team progression across all critical task and KSAA structures based on aggregated xAPI statements. These data will provide better insight into Soldier and unit readiness, and is intended to be directly applied to support training plan activities. Eventually, this dashboard may display predictive analytics, but in current STEEL-R implementations the only data available is learner state. This data is made available through CaSS Application Programming Interfaces (APIs).

Design of the proficiency dashboard is challenging and is expected to go through multiple iterations. One challenge is that trainers think in terms of performance on tasks whereas data in CaSS is about competencies. These competencies may represent underlying skills/abilities that explain performance, which is the long-term intent of STEEL-R, but in current state the competencies closely parallel battle drill tasks and the underlying KSAAs. The data gathered and reported by GIFT and other TADDS represents performance at a given time (i.e., what a soldier or unit *has* done in the past), whereas assertions and state in CaSS represent competency (i.e., what a soldier or unit *can* do, in present or future). These are easily confused, and while competency is predictive of performance, it is not the same. For example, a soldier who is an expert marksman, meaning they possess a marksmanship competency at the expert level, does not hit the target every time and may perform below expectation in any given scenario. Competency states provide a probabilistic determinant of success, based on experiential evidence, and does not guarantee success. Current thinking is that *both* performance and competency should be displayed, with the challenge of distinguishing them.

Other challenges include providing the ability to drill down through echelons, displaying how competencies have been acquired (or lost) over time, and displaying data such as practice scores and assertion scores without running the risk that they are over-interpreted or misinterpreted. Everyone loves an easy way to keep score, but these quantities have complex definitions and depend on complex forgetting and spacing functions. At this stage what information to display and how to display it in support evaluation and decision making is a research question being explored.

FUTRE WORK AND CONCLUSIONS

As of the writing of this paper, STEEL-R has achieved its early goals of creating a viable architecture, gathering data from GIFT via xAPI statements, converting xAPI statements to CaSS assertions, computing an experiential learner state from those assertions for both individuals and teams, and displaying data in a competency proficiency dashboard. Much of this has involved breaking new ground. Existing training doctrine and methods focus on binary "go/no-go" decisions and on task performance rather than competency, and little work has been done on gathering live performance data from multiple training systems and mapping it to assertions about competencies. The experiential and inferential models used in STEEL-R to estimate whether a team or individual possesses a given competency go beyond those used by current intelligent tutoring and similar system and have not yet been adjusted or validated in live trials. Our future planned work includes field experiments, and we anticipate implementing models that, unlike the ones used to date, learn parameters from data. We also want to create predictive models that can inform transitions from TADDS to live exercises and models that help determine what underlying KSAAs need to be trained and what scenarios are likely to help a soldier or unit make the most progress.

ACKNOWLEDGEMENTS

The research described herein has been sponsored by the U.S Army DEVCOM – Soldier Center – Simulation and Training Technology Center. 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

ADL (2021a). Total Learning Architecture. Retrieved June 2, 2021 from: https://adlnet.gov/guides/tla/

ADL, (2021b). Experience API (xAPI) Standard. Retrieved June 2, 2021 from https://adlnet.gov/projects/xapi/

ADL, (2018). xAPI Profiles. Retrieved June 2, 2021 from https://github.com/adlnet/xapi-profiles

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher-level cognition and its relation to visual attention. Human–Computer Interaction, 12(4), 439-462.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. Journal of Mathematical Psychology, 55(1), 25-35.

Blake-Plock, S., Hoyt, W., & Casey, C. Instrumenting GIFT with xAPI: a use case for IEEE P9274.3.x standards activity and implications for the broader field of ITS and AIS. (2021). Proceedings of the Ninth Annual GIFT Users Symposium. Orlando, FL: US Army Combat Capabilities Development Command - Soldier Center.

Broadwell, M. M (1969). Teaching for learning (XVI). *The Gospel Guardian*, Vol 20, No 41, 1969.

Dreyfus, S. (2004). 5-Stage Model of Adult Skill Acquisition. *Bulletin of Science, Tech and Society*,24(3), 177–181.

Ericsson, K. A. (2006). *The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance*. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), The Cambridge handbook of expertise and expert performance (pp. 683–703). Cambridge University Press.

Fletcher, J.D. (2005). *The Advanced Distributed Learning Vision and Getting from Here to There*. Institute for Defense Analyses (IDA) Report D-3212. Retrieved from DTIC at https://apps.dtic.mil/sti/citations/ADA452053.

Goldberg, B., Hoffman, M. & Graesser, A. (2020). Adding a Human to the Adaptive Instructional System Loop: Integrating GIFT and Battle Space Visualization. In A. Graesser, X. Hu, Rus, V. and B. Goldberg (Eds.) Design Recommendations for Intelligent Tutoring Systems, Vol. 7: Data Visualization, U.S. Army DEVCOM.

Jastrzembski, T. S., & Gluck, K. A. (2009). A formal comparison of model variants for performance prediction. In Proceedings of the 9th International Conference of Cognitive Modeling, Manchester, United Kingdom, 2009.

Kolb, D. (2014). *Experiential Learning: Experience as the Source of Learning and Development*. NJ: Pearson.

Miller GE (1990). The Assessment of Clinical Skills/Competence/Performance. *Acad Med.* 1990 Sep;65(9 Suppl):S63-7.

Owens, Gupton, Hellman & Goldberg (2020). Re-Thinking the Tactical Small Unit Synthetic Training Model. Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Owens, K.P. (2021). *Competency-Based Experiential-Expertise and Future Adaptive Learning Systems*. In Proceedings of Conference on Human-Computer Interaction International, Adaptive Instructional Systems.

Patton, D., Townsend, L., Milham, L., Johnston, J., Riddle, D., Start, A. & Adler, A. (2018). *Optimizing Team Performance When Resilience Falters: An Integrated Approach*. In Inter Conf on Augmented Cognition.

Pavlik, P & Anderson, J. (2003). An ACT-R model of the spacing effect. In 5th Intern Conf of Cognitive Modeling.

Robson, R., & Poltrack, J. (2017). Using competencies to map performance across multiple activities. In Proceedings of I/ITSEC 2017.

Robson, R. (2019). Assessing competency instead of learning outcomes. In International Conference on Artificial Intelligence in Adaptive Education (AIED), Beijing, China.

Robson, E., Robson, R., Buskirk, T., Ray, F. & Owens, K. (2021). An Experiential Competency Application Framework. To appear in proceedings of the 23rd International Conference on Human-Computer Interaction.

Sottilare, R., Goldberg, B., Brawner, K. & Holden, H. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). Paper presented at the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Sottilare, R. A., Ragusa, C., Hoffman, M., & Goldberg, B. (2013). Characterizing an Adaptive Tutoring Learning Effect Chain for Individual and Team Tutoring. In Interservice/Industry Training, Sim & Education Conference.

Sottilare, R., Shawn Burke, C., Salas, E., Sinatra, A., Johnston, J & Gilbert, S. (2018). Designing Adaptive Instruction for Teams: a Meta-Analysis. *Int J Artif Intell Educ*, 28, 225–264

TRADOC. (2021). Training and Educational Development in Support of the Institutional Domain. TRADOC Pamphlet 350-70-14. 15 April 202.

Walcutt, J.J. & Schatz, S. (Eds.) (2019). Modernizing Learning: Building the Future Learning Ecosystem. Washington, DC: Government Publishing Office.