

# A Machine Learning-Based External Assessment Engine for GIFT to Support Team Training in Dismounted Battle Drill Operations

Caleb Vatal<sup>1</sup>, Naveeduddin Mohammed<sup>1</sup> and Gautam Biswas<sup>1</sup>

<sup>1</sup> *Vanderbilt University, Institute for Software Integrated Systems, Nashville TN, USA*

## Abstract

In urban combat scenarios performed by the armed forces, effective teamwork is essential to mission success. Soldiers must be able to rapidly make decisions regarding their own and their team behavior and effectively communicate and synchronize those decisions with their squad members. Developing proficiency in such team-oriented tasks requires repeated practice and training exercises, preferably accompanied by feedback and scaffolding of the team. In this work, we discuss our research in the development of intelligent training systems to support evaluation of individual and team performance during training exercises in synthetic training environments (STEs). To demonstrate the effectiveness of our research approach, we develop an external assessment engine (EAE) for the Generalized Intelligent Framework for Tutoring (GIFT). Our engine utilizes the multi-modal data generated by Soldiers' interactions with the STE environment. Through a combination of machine learning methods (e.g., motion tracking, and posture and gesture detection) and log data generated of Soldiers' actions and their outcomes from the STE, we compute automated performance metrics for squads of Soldiers that operate as teams in the scenario. The generated metrics span both individual and team performance across a variety of psychomotor, cognitive, and coordination skills and strategies. In this work, we discuss the high-level framework and design our EAE with application to dismounted battle drills for the armed forces, though the analysis techniques we develop are general they can be easily modified to apply to a variety of team training scenarios.

## Keywords

Teamwork, Tutoring, Training, Cognitive Task Analysis

## 1. Introduction

In many complex systems and environments, effective performance relies on a high degree of teamwork and shared decision making among multiple people. Breakdowns in teamwork can lead to loss of revenue, expenditure of extra resources, and in many domains, even loss of life. In order to mitigate such potential breakdowns, teamwork tasks must be effectively trained. In recent years, computer-based environments for simulating real-world tasks in low-stakes environments have become a popular method for executing such training. Such computer-based environments provide a rich source of multi-modal data for evaluating the performance of learners and trainees. By designing intelligent systems to analyze this multi-modal data, such performance evaluations can be automated so that trainees can be given feedback and scaffolding more frequently than if such feedback came only from domain experts.

While much of the past applications of learner evaluation in these computer-based learning environments has focused evaluations of the individual, the same environments and data can also be analyzed to study team-performance. As a first step toward such automated evaluation of teamwork, we develop a cognitive task model for the Enter and Clear a Room dismounted battle drill of the armed forces. Our model leverages a combination of the data generated by a synthetic training environment

Proceedings of the Challenges and Advances in Team Tutoring Workshop during the Artificial Intelligence in Education Conference, June 15<sup>th</sup>, 2021, Virtual from Utrecht, The Netherlands

EMAIL: caleb.m.vatal@vanderbilt.edu (A. 1); naveeduddin.mohammed@vanderbilt.edu (A. 2); gautam.biswas@vanderbilt.edu (A. 3)



© 2022 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
CEUR Workshop Proceedings (CEUR-WS.org)

(STE), machine learning techniques, and cognitive modeling to generated performance metrics for both individuals and teams within this domain. This paper discusses the conceptual design of this framework, as well as its implementation in the Generalized Intelligent Framework for Tutoring (GIFT). In addition, we also provide a conceptual overview of how our framework could be applied to a variety of other domains, both within the military and in other applications.

## **2. Case Study: Enter and Clear a Room**

Enter and Clear a Room (ECR) is a dismounted battle drill designed to operationalize procedures for urban warfare. The goal is to neutralize a set of enemy personnel who are located within a building which houses unique obstacles and may also house non-combatants. Squads of three to four Soldiers are assigned to neutralize combatants within a given room as they move through clearing an entire building. The operation begins with the squad taking formation outside a doorway to the assigned room. After a signal from the squad leader to commence, the squad members rapidly enter the room in succession following tactical procedures for movement and sectors of fire. While entering and moving through the room, the squad neutralizes enemy combatants while minimizing collateral damage to the squad, any present noncombatants, and the property. Once all enemies have been neutralized, the squad leader gives the clear signal, and the operation concludes. For the remainder of this paper, we will discuss the ECR domain as a case study for implementation of our team training tools.

## **3. Data from Synthetic Training Environments**

Within the U.S. Army, STEs are often used for streamlining and enhancing the training process of Soldiers on a variety of combat drills. The goal is to increase soldier proficiency in psychomotor skills, coordinated movement and teamwork, and cognitive skills and strategies. In addition to providing repeated practice in an inexpensive but realistic environment, STEs also have the added benefit of generating multi-modal user data for the Soldiers who utilize them. By analyzing this data, in-depth data-driven performance analysis can be produced. This section will discuss the specific STE used in our case study for the ECR domain and the data which it produces.

### **3.1. Squad Advanced Marksmanship Trainer**

The squad advanced marksmanship trainer (SAM-T) is a STE designed for simulating live-fire weapon training. The SAM-T serves to accelerate training and development of individual Soldier and squad close combat skills to increase readiness and performance by allowing for repeated practice and drills in a realistic operational environment. SAM-T operates as a mixed reality training environment, where squads see urban combat scenarios played out on a U-shaped arena (see Figure 1). The projected virtual scenarios are designed using Virtual Battle Simulator 3 (VBS3). The Soldiers interact with the virtual scenario by moving around within the physical arena space and firing their weapons, which contain a digital interface that maps the weapon aim onto the relative position in the virtual space. The SAM-T system logs a variety of events which occur throughout the course of the scenario including Soldier weapon fire data, Soldier weapon aim data, and Soldier biometric data. In addition, other events from the VBS scenario are also logged including virtual agent position, virtual agent weapon aim, and virtual agent weapon fire [10].



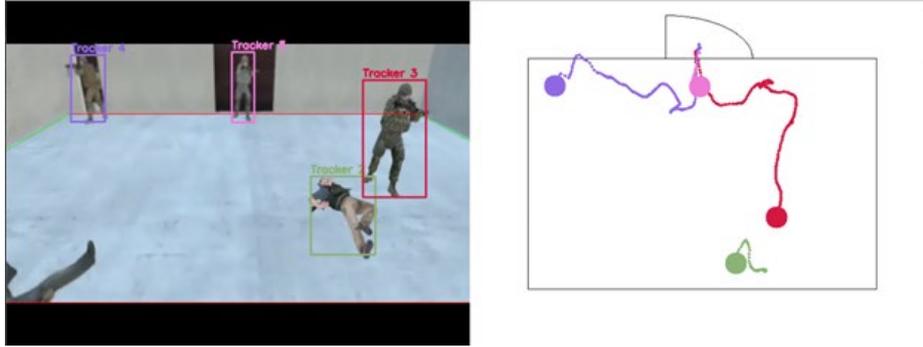
**Figure 1:** The SAM-T synthetic training environment. A virtual training scenario is projected onto screens in a U-shaped arena. Soldiers move around in the physical area space and interact with the digital scenario using weapons designed to record fire events in digital space based on their relative aim.

### 3.2. Computer Vision Techniques

The SAM-T environment discussed above provides a variety of data related to both the Soldier actions, as well as the virtual scenario. However, the SAM-T does not capture information about the Soldiers' states in the physical environment. This physical data includes Soldier position, pose, and gesture, as well as information about other objects in the physical space (e.g., obstacles and cover). To supplement the data supplied by the training environment, we employ computer vision techniques applied to video captured from an overhead camera mounted above the SAM-T arena space.

To capture Soldier position in the physical space, we utilize visual motion tracking applied to the camera video. Since we are primarily focused on team-based exercises, multiple object tracking (MOT) is required and will produce a track of each individual Soldier's positions over the course of the scenario. In recent years, the most successful paradigm for MOT is the *tracking-by-detection* paradigm, which combined machine learning methods for object (person) detection with signal processing methods and other algorithms for matching these detections between frames of video [5]. By running the object detector in each frame and matching detections between two frames to the same objects, a cohesive track of the object's motion can be created. In our work, we utilize a modification of the well-known SORT algorithm for MOT [1]. Our modification, called *Fusion-SORT*, combines multiple sources of detection, specifically detection of heads and detection of bodies, to improve tracking performance during times of subject occlusion. For a more detailed discussion of the Fusion-SORT algorithm, see Vatrak et al. [11].

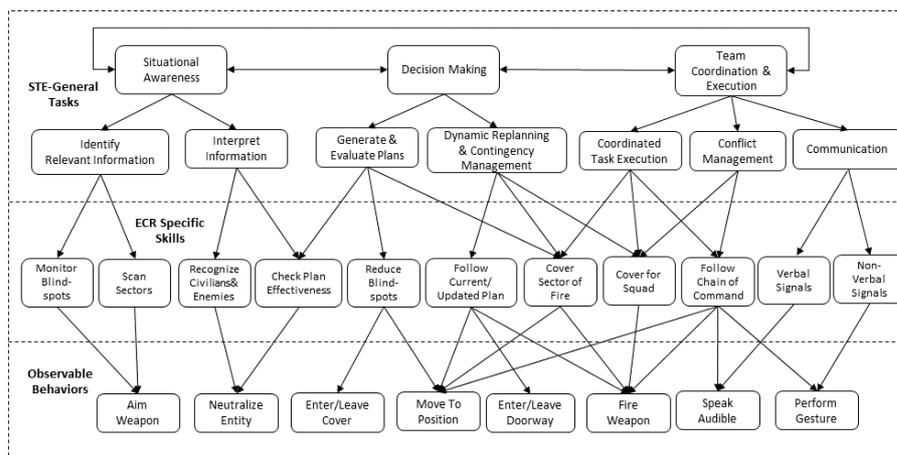
Fusion-SORT produces tracks of Soldier positions in reference to the overhead camera's view, which we call the *camera frame*. In order to analyze Soldier behavior in the 3D area, we project these camera frame tracks to the *map frame*, which is a 2D bird's eye view of the arena space. The projection to the map frame utilizes 3-point planar homography based on the bounds (walls) of the scenario space. During the setup of the SAM-T system, users create a configuration file which defines line segments representing the walls of the SAM-T arena. During execution, the tracking framework computes the intersection of the defined wall line segments to determine the corners of the SAM-T scenario space. These four corners are used as the reference points for the homography to convert the camera frame to the map frame. This projection allows conversion of the position tracks generated by Fusion-SORT to an interpretable position in the 3D space of the SAM-T arena. An example of this projection on simulated data is shown in Figure 2.



**Figure 2:** Example of the planar projection from the overhead camera in the SAM-T arena to the generated top-down map used for analysis.

#### 4. Cognitive Task Model for Team Evaluation

The learner modeling procedures that we use for analyzing both individual and team performance are derived from cognitive task analysis methods e.g., [4]. We create a hierarchical model of task and sub-task competencies with low levels of the hierarchy mapping onto directly observable Soldier behaviors in the STE and higher levels mapping onto more general cognitive skills utilized by the team and its members. An example of this hierarchical task model applied to the ECR domain is shown in Figure 3.



**Figure 3:** Hierarchical task model for the dismounted battle drill called *Enter and Clear a Room*

Cognitive task analysis methods have been shown to be an effective method for training on complex decision-making tasks. Previous work has mapped observable actions of learners in intelligent tutoring environments onto task models to provide data-driven performance and learner evaluations. However, most work to date using task models have focused on evaluation of individuals only. In our work, we extend the task model to capture team behavior. This addition of team-based tasks and behaviors to the model are motivated by understanding of team performance and team decision making in terms of shared mental models of the task and the team [2], [9].

Work in understanding of dynamic team decision making has suggested that effective team performance requires team members to hold highly overlapping cognitive representations of executed tasks and team member roles and responsibilities [3]. Such cognitive representations are often referred to as mental models. In the context of the ECR domain, Soldiers must make decisions under severe time pressure, relying on predetermined procedures and prototype matching based on their previous training experiences [8]. This procedural execution and prototype matching represent activation of an individual's mental model of the scenario. If the mental models of members of a squad were significantly mismatched, it could lead to collateral damage and fratricide as different members of the

squad execute conflicting plans [12]. The constructed task model shown in Figure 3 is method of capturing the Soldiers' mental models. Construction of our task model follows from the findings of Glickman et al. [6] that two tracks of behavior are involved with team training: taskwork and teamwork.

The taskwork behavior track contains the tasks, skills, and procedures related to direct execution of the mission. In order to avoid confusion of terminology between the hierarchical task model and the taskwork behavior track, for the remainder of the paper, we will refer to Glickman's concept of the taskwork behavior track as the problem-solving behavior track. In the case of the ECR domain, the problem-solving behavior track represents the overall goal of neutralizing the room, as well as the operational objectives that have been standardized to do so, including procedures for entering the room, moving through the space, covering sectors of fire, neutralizing enemy combatants, etc. Through review of ECR-related literature and discussion with domain experts, we determined that these problem-solving behaviors are instances of two high-level STE-General tasks: Situational Awareness and Decision Making, as shown at the highest level of the task model in Figure 3. These STE-General problem-solving skills then breakdown into skills more specific to the ECR domain such as specific movement patterns, monitoring blind spots, covering sector of fire, etc. These problem-solving skills inherently do not require specific team skills; rather, they are procedurally executed knowledge and plans (sequence of tasks) based on pre-specified Army procedures and the mental models that Soldiers acquire during their training.

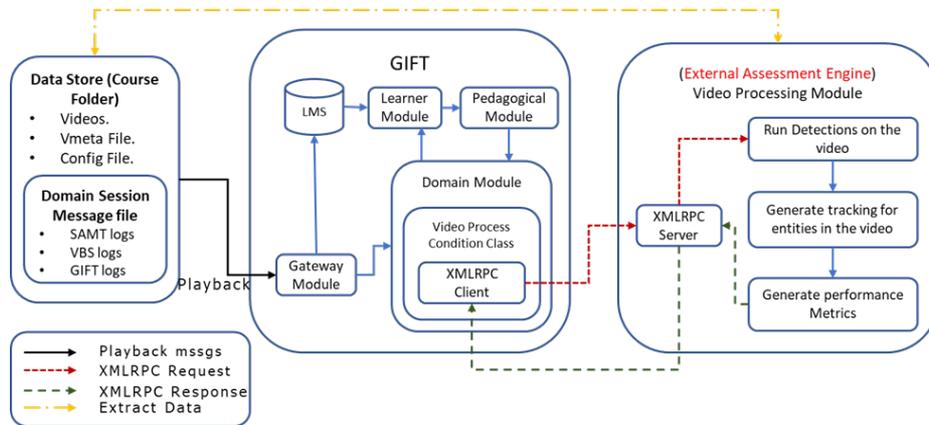
The teamwork behavior track contains the skills needed to function within the context of the larger team. At a high level, many of these teamwork skills are useful and consistent across a variety of domains [6]. In our task model, we breakdown the highest-level teamwork skill into three subtasks: coordinated task execution, conflict management, and communication. These skills are largely based on the taxonomy of shared cognition for teamwork presented in Wilson et al. [12]. Their taxonomy breaks down teamwork into 3 skills: *communication*, which maps onto communication skills in our model, *coordination*, which maps onto coordinated task execution skill in our model, and *cooperation*, which maps onto conflict management skills in our model. These high-level teamwork skills are not specific to the ECR domain or even military domains in general. Rather, these high-level teamwork behaviors are necessary for nearly all team-based problem-solving environments.

However, when these skills are operationalized, they become somewhat more domain specific and more interconnected with the problem-solving skills. For example, in the ECR domain, coordinated task execution requires a common understanding of roles within the group and what those roles mean for movement patterns throughout the room and covering sectors of fire. If team members did not have a common understanding of these roles, it is likely that multiple Soldiers would attempt to move into or cover the same sector leading to blind spots in coverage and potential fratricide. The interconnected nature of the problem-solving and teamwork tracks is the primary reason that we combine them into a single task model instead of doing analysis separately, as is suggested by the discussion of multiple mental models by Cannon-Bower et al. [3]. When the procedural nature of the problem-solving track begins to fail, the teamwork track becomes increasingly important. This is why most of the overlap between the taskwork and teamwork tracks occurs in the decision-making section of our task model. Appropriate decision making within a team context requires individual members to predict the needs of their teammates and adjust their procedures accordingly. This is particularly true when decisions are under severe time pressure, such as in ECR, as needs cannot be directly communicated effectively [3].

## 5. Implementation in GIFT

Generalized Intelligent Framework for Tutoring (GIFT) is a modular, interoperable, and reusable computer-based tutoring framework for military training. GIFT has two different assessment engines that provide assessments for specific concepts that an individual or a team is being trained on: 1) default assessment engine and, 2) external assessment engine. Our team tutoring model as described in the previous section is implemented in GIFT as an external assessment engine. The module interacts with GIFT through a server communication model. A GIFT condition class establishes connection between GIFT and the external assessment engine. Figure 4 shows the computational architecture of the video processing module and its integration in GIFT. The current integration architecture focuses on post data analysis, but in the future, we will modify it to support online analysis and assessment. In this section,

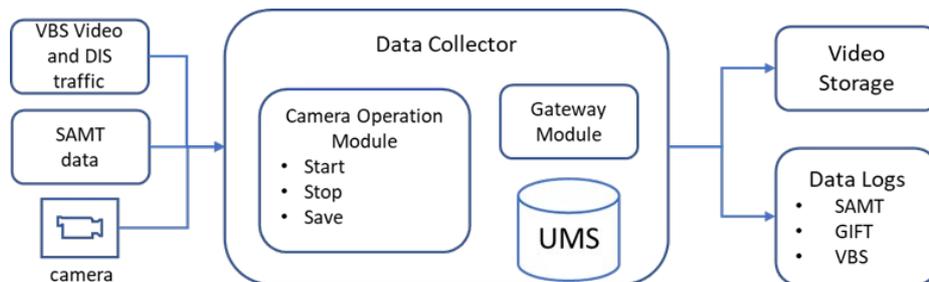
we briefly discuss the data collector, and communication modules of the GIFT integration. For a complete discussion of the implementation in GIFT, see Vatrul et al. [11].



**Figure 4:** General architecture for integration of the external assessment engine with GIFT

## 5.1. Data Collector

Data is collected at the SAM-T testing site prior to analysis. The data collector is a subset of GIFT, and it incorporates the Gateway module and the User Management store (UMS). The architecture is shown in Figure 5. This setup does not follow the traditional GIFT framework of real time analysis by running a domain Module course to generate assessments and a Pedagogical module to determine instructional strategies based on the assessments. The data collector is designed to collect data from multiple sources: (1) a camera placed on top of the SAM-T screen facing the soldiers; (2) VBS3 video and events derived from the ECR scenario as it plays out in the environment in the form of DIS; and (3) SAM-T behavior and shot event data captured across wearable and weapon-embedded sensors. The data collector captures videos from VBS and camera and stores them in a video storage location. The DIS messages from VBS and SAM-T logs are transformed into predefined logging format and are stored at a location.



**Figure 5:** General architecture of the GIFT data collector

## 5.2. Communication Module

The communication between GIFT and external assessment engine (EAE) is established through an XML-RPC interface. The communication mostly takes place through XML-RPC setup except the initialization which happens through command line interface. As GIFT initializes, the condition class associated with EAE initializes the XML-RPC client connection to EAE. The EAE server is initialized through a command line interface. The server implementation for EAE is done in Python 3.6. Once the connection is established between EAE XML-RPC server and the condition class XML-RPC Client, GIFT sends information to EAE consisting of location of data store on the local file system and a dump of all the data collected by the SAM-T and VBS systems. Once the data packet is received by the EAE, it applies the machine learning and cognitive task modeling approaches described in the previous

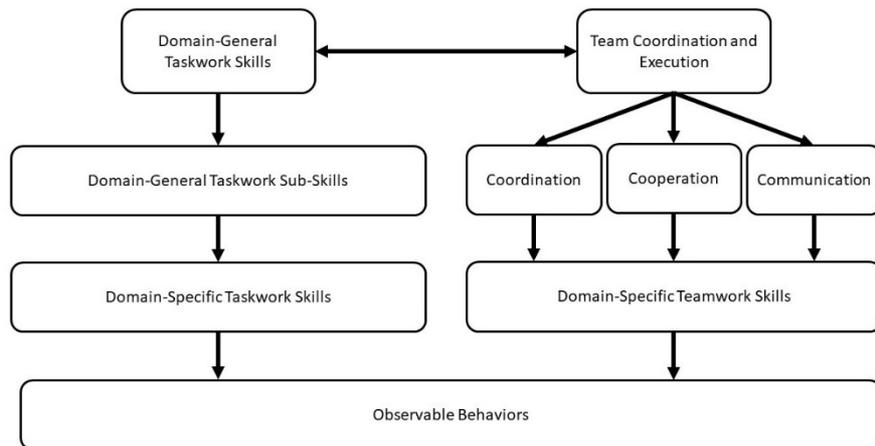
sections to derive metrics. These metrics are then entered into a queue data structure with a timestamp that represents the time in the scenario at which the performance metric was generated. This is repeated for the entire data packet and once the processing is done, EAE generates a start signal for GIFT to continue initialization and move on to playback mode.

When GIFT initializes playback mode, the request is sent to the EAE, and a timer is initialized. The timer is used to match the timestamp of the event in playback to the timestamp of the next evaluated performance metric in the queue. If the timer value matches, then the condition class is updated with that associated value. This process is repeated until the entire metric queue is empty and the EAE shuts down soon after.

## 6. Toward Domain Generality

In this paper, we have discussed the development of a model for team tutoring in the context of the Enter and Clear a Room domain and implemented in the GIFT framework as an external assessment engine. However, the framework of combining the two skill tracks of Glickman et al. [6] with the teamwork taxonomy of Wilson et al. [12] in a cognitive task model can be applied to a variety of domains including other dismounted battle drills, mounted battle drills, and even domains outside of the armed forces, such as K12 education. In this section, we discuss a starting point for the generalization of the framework presented in this paper to other domains.

Figure 6 shows a high-level generalization of our hierarchical cognitive task model. Overall, this generalization follows the same structure as the model presented in this paper. At the top levels of the model, we have domain-general skills, and as we move down the hierarchy, the skills become more domain-specific and more observable. In addition, the model is largely divided into two sections, following the two skill tracks of Glickman et al. [6].



**Figure 6:** Generalized cognitive task model for team tutoring based on the two skills tracks of Glickman et al. [6] and the team taxonomy of Wilson et al. [12].

The left side of the model contains the problem-solving (taskwork) track, representing the skills and procedures directly related to the execution of the given domain. Just as before, taskwork skills are primarily individual and related to the given domain. At the top level, we have domain-general problem-solving skills. In the model described in this paper, these are the situational awareness and decision-making skills; however, depending on the domain for which the model is built, these skills may have to be replaced. For example, in a K12 education setting, these skills may be concepts such as information gathering and solution construction, as in Kinnebrew et al. ([7]). In general, these top-level skills must be domain-general but also relevant to the given domain. Below this level is the domain-general problem-solving sub-skills. This level represents skills which are still domain-general, but more specific and observable than at the highest level. At the third level, we begin domain-specific tasks. These are specific skills which must be specified for each domain on which the model is applied. For example, in

the ECR domain this level contains concepts such as monitoring blind spots, cover sectors of fire, etc., but in a mounted battle drill, this level might contain skills related to vehicle operation instead. Finally, the lowest level contains observable behaviors. This level represents concepts which are specific to the environment in which the agents are acting. For example, in the ECR domain under the SAM-T, this layer contains behaviors such as aiming a weapon, moving throughout the room, etc., but in a K12 domain in a computer-based learning environment, this layer might contain behaviors such as mouse clicks, reading an assigned article, etc. These observable behaviors are a result of the combination of the domain and the environment.

The right side of the model contains the teamwork track, representing the skills needed to function in the context of a larger team. Because teamwork skills are highly transferrable between domains, the teamwork track of the model remains largely similar between domains, especially at the high levels. At the top two levels of the teamwork track, the model utilizes the teamwork taxonomy of Wilson et al. [12]. In this taxonomy, teamwork skills breakdown into coordination, cooperation, and communication. Coordination represents the ability of team to integrate, synchronize, and sequence activities without wasting resources. Cooperation represents the team's desire to coordinate. In this sense, cooperation is highly affective and encompasses the team members' attitudes and motivations. Finally, communication represents the ability of a team to transfer knowledge and information among its members. All these concepts are highly general and applicable to any domain requiring teamwork. Below this level is the domain specific teamwork skills, which represents the realization of the taxonomy's high-level skills within a specific domain. For example, with the ECR domain, communication is realized primarily with gestures, but with a K12 domain, communication is likely primarily with speech. Finally, at the lowest level is again the observable behaviors layer, which is not specific to the teamwork or problem-solving tracks but rather spans both. As described earlier, observable behaviors result from a combination of the domain and the environment.

## 7. Conclusions and Future Work

In this paper, we have described our new framework for evaluating team tutoring based on a combination of the dual problem-solving (taskwork) and teamwork skill tracks proposed by Glickman et al. [6] and the teamwork taxonomy proposed by Wilson et al. [12], and structured around a hierarchical task model based on cognitive task analysis. We discussed our implementation of the framework in GIFT using the ECR dismounted battle drill of the armed forces as a case study. While this work has primarily been focused on this domain during development, we believe that the framework presented represents a highly generalizable model for evaluation of team tutoring for both military domains such as mounted and dismounted battle drills, as well as other domains such as K12 education or workplace training. Future work will focus on evaluating this framework in-depth on a variety of case studies, both within military domains and others, while comparing to expert analysis of team performance for validation.

## 8. References

- [1] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Uproft, Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP) (2016): 3464-3468.
- [2] J.A. Cannon-Bower, E. Salas, S.A. Converse, Cognitive psychology and team training: Shared mental models in complex Systems, *Human Factors Society Bulletin* 22 (1990): 1-4.
- [3] J.A. Cannon-Bower, E. Salas, S.A. Converse, Shared mental models in expert team decision making. In Castellan Jr., N.J. (ed.) *Individual and group decision making: Current issues* (1993): 221-246. Lawrence Erlbaum Associates Inc.
- [4] R. Clark, F. Estes, Cognitive Task Analysis for Training. *International Journal of Educational Research* 25(5) (1996): 403-417.
- [5] M. Fiaz, A. Mahmood, S. Javed, S.K. Jung, Handcrafted and deep trackers: Recent visual object tracking approaches and trends. *ACM Computing Surveys (CSUR)*, 52(2) (2019): 1-44.

- [6] A.S. Glickman, S. Zimmer, R.C. Montero, P.J. Guerette, W.J. Campbell, B.B. Morgan, E. Salas, The evolution of teamwork skills: An empirical assessment with implications for training (Tech. Rep. No. 87-016). Naval Training Systems Center, Orlando FL (1987).
- [7] J. Kinnebrew, J. Segedy, G. Biswas, Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning behaviors. *IEEE Transactions on Learning Technologies* 10(2) (2017): 140-153.
- [8] G. Klein, R. Calderwood, A. Clinton-Cirocco, Rapid decision making on the fire ground. In: *Proceedings of the Human Factors Society 30th annual meeting*, Human Factors Society (1986): 576-580.
- [9] D.L. Kleinman, D. Serfaty, D., Team performance assessment in distributed decision making. In: Gilson, R., Kincaid, J.P., Goldiez, B. (eds.) *Proceedings for Interactive Networked Simulation for Training Conference*, Institute for Simulation and Training, Orlando FL (1989): 22-27.
- [10] M. Pargett, Squad tactics tested on new virtual marksmanship trainer, <https://www.army.mil/article/219231/squadtacticstestedonnewvirtualmarksmanshiptrainer>. Accessed April 2, 2019.
- [11] C. Vatrail, N. Mohammed, G. Biswas, B.S. Goldberg: GIFT External Assessment Engine for Analyzing Individual and Team Performance for Dismounted Battle Drills. In *9th Annual Generalized Intelligent Framework for Tutoring User Symposium*, US Army Combat Capabilities Command Center, Orlando FL (2021): 109-127.
- [12] K. Wilson, E. Salas, H. Priest, D. Andrews, Errors in the heat of battle: taking a closer look at shared cognition breakdowns through teamwork. *Human Factors* 49(2) (2007): 243-256.