Scenarios for Training Teamwork Skills in Virtual Environments with GIFT

Robert K. McCormack[1], Tara Kilcullen[1], Anne M. Sinatra[2], Tara Brown[1], Jeffrey M. Beaubien[1]
Aptima, Inc.[1], US Army Research Laboratory, Human Research & Engineering Directorate[2]

INTRODUCTION

Breakdowns in teamwork are often cited as a cause for poor, and at times, devastating outcomes that lead to loss of life, limb, and material resources (Wilson, Salas, Priest, & Andrews, 2007). Such failures are often attributed to breakdowns in essential teamwork skills, such as coordination and communication, and emergent team states, such as cohesion and shared situational awareness. The relevance of these team constructs is evidenced in the academic literature across a variety of domains, including both medicine (Hughes et al., 2016) and the military (Sottilare, Burke, Salas, Sinatra, Johnston & Gilbert, 2017; Wilson et al., 2007). Additionally, the Army has recognized the importance of Soldiers demonstrating these teamwork concepts. Specifically, several of the principles of Mission Command outlined in ADRP 6-0 align with these concepts, including "Build cohesive teams through mutual trust" and "Create shared understanding" (U.S. Department of the Army, 2012).

While the importance of these teamwork concepts is recognized, there remain challenges to training them efficiently. To maximize effects while simultaneously minimizing costs, there has been a push toward the use of intelligent tutoring systems (ITSs) and ITS frameworks in training. However, these systems have been almost universally designed to train individuals, not teams. In particular, the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012; Sottilare, Brawner, Sinatra & Johnston, 2017) enables the ITS community to efficiently achieve learning effects for individuals. To date, all of the GIFT-based team tutoring applications have been limited to dyads or triads (Bonner, Walton, Dorneich, Gilbert, Winer, & Sottilare, 2015; Gilbert, et al., 2017). However, in order for GIFT to fully support Army training needs, it must scale to larger team structures, such as squads, platoons, and above. In this paper, the authors outline development of such a system within the GIFT framework and development of supporting training scenarios within the Virtual Battlespace (VBS3) simulation environment. There are two overarching objectives to this effort:

1. The first objective is to demonstrate the utility of GIFT for adaptive team training in rich Virtual Environments (VEs), and specifically VBS3. Previous team tutoring implementations using GIFT focused on dyads and/or triads. The current effort aims to assess the utility of using GIFT for larger organizational structures (e.g., squads).

2. Previous efforts have relied heavily upon expert observer rating scales and self-report surveys of team processes and performance. The current effort seeks to assess the utility of using unobtrusive measurement methods (Orvis, Duchon, & DeCostanza, 2013) instead.

To accomplish these two objectives, the authors are developing a prototype training system in GIFT that can capture meaningful team processes and emergent states in a virtual training environment. In addition, the authors are developing realistic training scenarios that provide sufficient complexity and team interaction opportunities to enable effective team training. Specifically, the authors are developing scenario frameworks that enable GIFT to read data collected unobtrusively from teams training using the VBS3 platform, and computing measures of key teamwork constructs that will be used to assess and debrief team performance. This paper will summarize developments to date towards achieving the above-mentioned goals.

TRAINING ENVIRONMENT

Intelligent Tutoring System

GIFT is a domain-independent intelligent tutoring system framework (Sottilare, Brawner, Sinatra & Johnston, 2017). Much of the research and the efforts to date in GIFT have focused on individual tutoring. However, the ultimate goal of GIFT is for tutoring to be conducted with teams. Both theoretical and practical work has been done with GIFT that will prepare it for scaling up for use with teams. A large-scale literature search and meta-analysis has served as the theoretical foundation for team tutoring in GIFT (Sottilare, Burke, Salas, Sinatra, Johnston, & GilbertOr 2017). As part of this effort, relevant behavioral markers were identified for several team constructs. Additionally, initial work has also been done to adapt GIFT for use by multiple users engaging in the same scenario simultaneously.

The first work to implement team tutoring in GIFT created a two-person surveillance task using the Virtual Battlespace 2 (VBS2) software. The task consisted of two individuals (spotters) each monitoring their own sector and reporting to their teammate if a threat was passing to the other's sector (Gilbert et al., 2017). This task demonstrated that GIFT could have two individuals simultaneously engage in a simulation-based environment, and was able to provide feedback based on the actions of both individuals separately, as well as the team as a whole. The next step was adjusting the surveillance task such that it had three individuals working together as a team to achieve their goals. Two spotters continued to monitor their respective areas or responsibility,and a third role – a "sniper" who received information from both spotters – was added. The role of the sniper included receiving information from the two spotters, acknowledging receipt of that information, and making decisions based on it. Through the development of this scenario, it was shown that GIFT was capable of providing tutoring and real-time feedback to a three-person team in a simulation-based environment.

looking toward the future, it is important to demonstrate that GIFT is capable of tutoring large numbers of individuals simultaneously, such as a squad of, which is typically composed to two four-person fire teams plus a squad leader. Scaling GIFT's team tutoring capabilities will require consideration of not only how to deal with the data of nine separate team members, but also how to measure teamwork within a VE and how to handle different team member roles. Therefore, new approaches should focus on defining roles within a GIFT tutoring scenario, simultaneously assessing the behavior of multiple Soldiers, and efficiently determining the team's overall performance in real-time. The teams' performance will then need to result in the proper feedback being given to the team either during or after engagement with a game scenario. Additionally, future work should find efficient ways to implement team behavioral markers in the GIFT software so that the team's performance can be assessed in real-time.

Virtual Training Environment

To train and assess teamwork skills, the authors utilize the VBS3 software. The decision to use VBS3 was two-fold. First, GIFT has integrated with VBS3 (and previous versions of the Virtual Battlespace software) throughout its development. Therefore, it already interoperates with the VBS3 architecture and data structures. Secondly, VBS3 is widely used throughout the Army. While we are not assuming that all research participants that will come through the training will have had exposure to VBS3, it is a readily available training asset at many Army installations. This will ensure adequate locations and candidates for validation of the training and teamwork measures.

When it comes to infantry, virtual training proves to be an overall challenge. Many virtual training platforms have proven to be ineffective for numerous reasons. These include overly cumbersome or counter-intuitive software interfaces, the system being too time-consuming to set up and tear down, and the lack of validated human performance measures. With the instantiation of VBS3 into their virtual training toolbox, infantry Soldiers and Marines are able to gain valuable, training experiences prior to completing live training. The flexibility of VBS3 – in terms of actions, assets, and customization – means that it can support the development of scenarios that are rich enough to enable measurement of teamwork skills.

Inherent in training and assessing teamwork skills is the ability for individual Soldiers to interact and communicate with one another. VBS3 includes a built-in text chat feature will serve as a primary means of team communication and information coordination, as well as providing a rich set of data for teamwork measurement. Team members will need to communicate about a number of issues throughout the scenario, such as detecting a threat, reporting a threat, and handing off a threat. Interaction and communication complexity can be manipulated by putting constraints on the communication structure. For example, the communication structure can be set up such that certain members of a team cannot communicate directly with members of another team, which mimics communication breakdowns during a mission.

The specifications for an initial VBS3 scenario, as just described, should provide enough complexity to require sufficient teamwork interaction. However, the goal is to make the scenario readily scalable to accommodate different team sizes, as well as to support the training of Soldiers at different expertise levels. The VBS3 simulation engine itself has been shown to support over 100 simultaneous learners, and the structure and number of the teams, assets, threats and communication constraints can be scaled to support more or less complex conditions, as desired.

SCENARIO DEVELOPMENT FOR VBS3

To support teamwork training within VBS3, realistic scenarios are needed that provide ample opportunities for assessment and feedback. The authors have identified a number of constraints for training scenarios:

1. Must be implemented within the constraints of the simulation environment (VBS3);
2. Must represent realistic tasks, interactions, and outcomes to ensure Soldier engagement and buy-in;
3. Must support the training and assessment of teamwork-related constructs (e.g., coordination, communication, cohesion) that emerge as a function of the team members' interactions;
4. Must allow team members to communicate both naturally and in a manner that enables assessment of communications for measurement purposes;
5. Must initially focus on the squad level, but also enable larger team structures to train within the simulation environment;
6. Be scalable to support higher echelon training objectives with more complex scenarios.

Scenario Overview

Working with an active duty Army infantry Subject Matter Expert (SME), the research team modified an existing Combat Search and Rescue (CSAR) training scenario that is currently being used at the Army's Basic Leader Course (BLC) to train and assess small unit leadership skills (See Figure 1 for an overview of the scenario). The scenario focuses on search and rescue of a downed pilot within the Area of Operations (AO). The team is a squad-sized element that is comprised of two four-person fire teams. The squad is led by a squad leader (a Sergeant); each fire team includes a fire team leader (a Corporal), as well as a Rifleman, a Squad Automatic Weapon (SAW) operator, and a Grenadier.

The scenario unfolds over a roughly 1-mile linear pathway through a forest which includes a mixture of tall trees and scrub brush that are common to northern Florida (Camp Blanding Joint Training Center). While Soldiers were able to venture from the path into the forest, it both slowed their movement and impaired their visual scan. Because of this, the forest also served as an ideal place for small groups of enemy fighters to launch ambushes against the squad.

Prior to starting the scenario, the Squad Leader is provided with a tactical map of the AO, along with a Fragmentary Order (FRAGO) that describes their mission. The squad leader is also provided with available intelligence (INTEL) about the location of the downed pilot as well as the number and disposition of enemy forces in the AO.

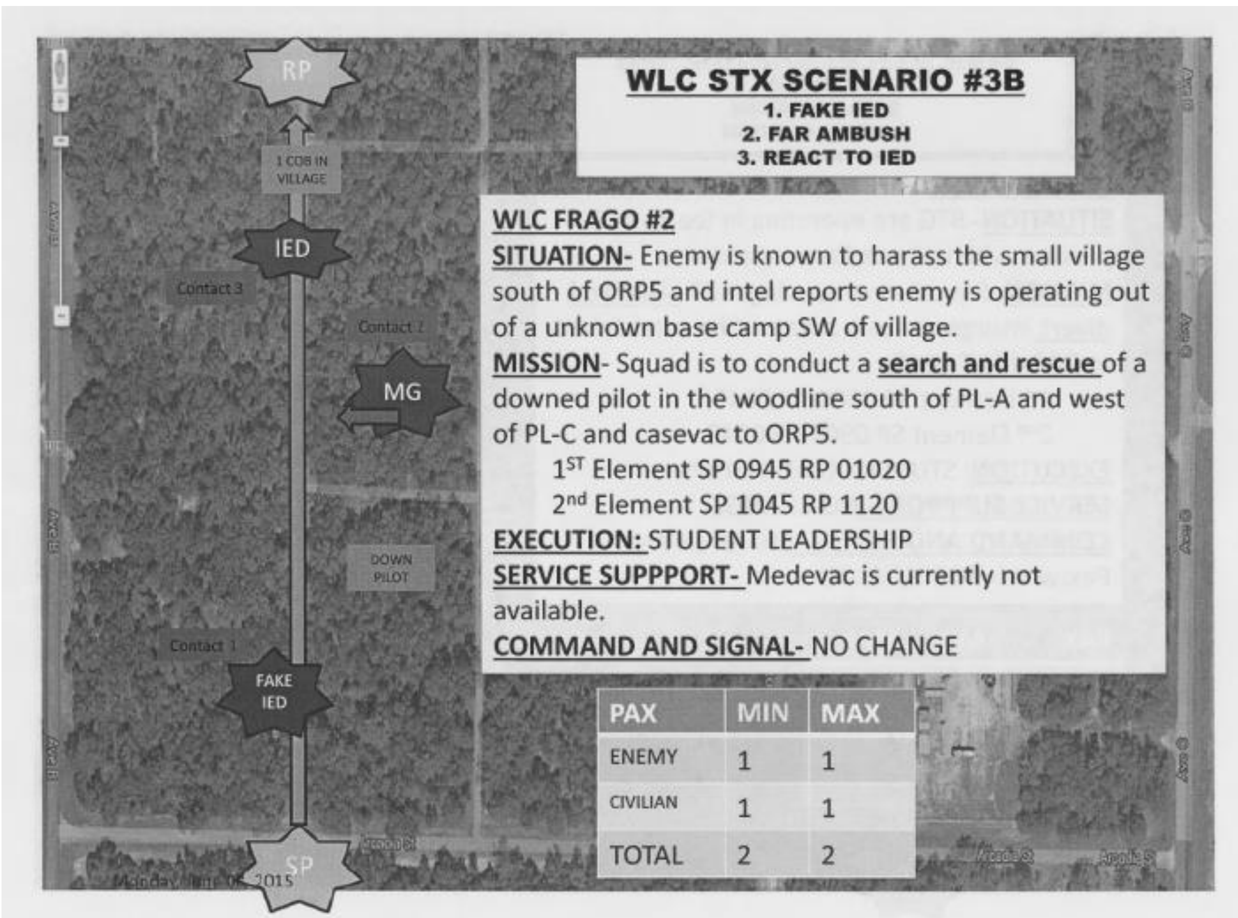

Figure 1. Notional CSAR Scenario

The squad's primary goal is to rescue a downed pilot. Their secondary goal is to complete a presence patrol in a local village in order to sustain their support against local enemy fighters. Along the way, the squad has to overcome several challenges.

After leaving the starting point, the squad first encounters a suspected Improved Explosive Device (IED). Despite being an enemy hoax, the squad is still required to perform a series of threat-relevant tasks, such as: Confirming (and communicating) the exact location and description of the device; Clearing all personnel to a safe distance; Cordoning (marking) the area to prevent anyone else from entering; Controlling access to the perimeter; and Checking for secondary devices.

The squad then continues down the path toward the estimated location of the downed pilot. Upon reaching the pilot's location, the squad must physically secure the pilot, cordon off the area, apply first aid, and radio headquarters to request medical evacuation. During this time, a local farmer arrives upon the scene towing a wagon full of goods. Before the helicopter can arrive, the squad then needs to apply escalating force to prevent the farmer – who could be an enemy fighter in disguise – from getting within "danger close" proximity to the pilot.

After the pilot is evacuated, the squad continues down the path toward the village. Along the way, they are ambushed by 3-4 enemy fighters who are equipped with small arms, such as AK-47 rifles. The fighters are largely unskilled and have poor aim. As a result, they cause little (or no) injuries to the squad, but this element provides an opportunity to measure how well the squad maintains their formation and responds to the threat, while maintaining their primary and secondary objectives. After dispatching the enemy fighters, the squad leader issues a Situation Report (SITREP) to headquarters, and redistributes ammo among the team. The squad then continues down the path. Along the way, they encounter a second IED, which requires the same set of behaviors that were described previously. Finally, the squad enters the village. At this time, they interact with village leaders – including the mayor, religious leader, and elders.

It is anticipated that during the scenario there will be several points where the scenario is paused and immediate feedback is given. This will provide opportunity for adjustment and recalibration among the team, but requires that opportunities for teamwork measurement occur throughout the entire scenario.

TEAMWORK MEASUREMENT

Based on a review of existing theory and measures, Sottilare et al. (2017) developed a list of behavioral indicators for several teamwork constructs. This set of behavioral markers provides the foundation upon which the research team is developing unobtrusively metrics of teamwork skills. The authors initially decided to target two areas for measure development – task cohesion and physical coordination – which will highlight how different types of data (e.g., communications, scenario interaction data) can be used to measure teamwork skills.

To develop our teamwork measures, the team uses a process based on the Rational Approach to Developing Systems-based Measures (RADSM; Orvis et al., 2013; see Figure 2), which has been successfully used to develop indicators and measures of team states (McCormack, Brown, Orvis, Perry, Myers, 2017).

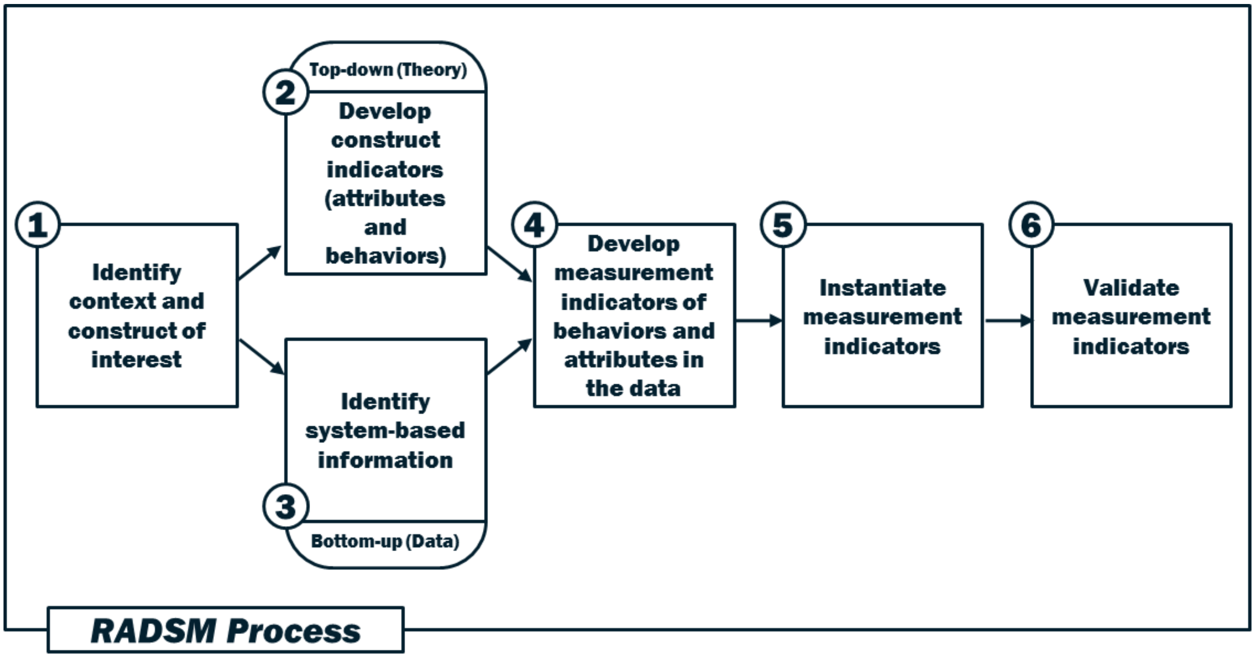

Figure 2. The RADSM Process for Measurement Development

The RADSM process consists of several steps, as highlighted in Figure 1, to ensure that developed measures are conceptually sound and contextually relevant. The end result of this process is a set of measures that can be assessed automatically and unobtrusively (that is, not requiring human coding or input) given the data available in the system.

Step 1 is focused on identifying the context and construct of interest for measurement. For the current effort, the context is a teamwork task, described above, performed within VBS3, while the constructs of interest are cohesion and cooperation. Steps 2 and 3 apply top-down and bottom-up approaches, respectively, to measure development. Specifically, in Step 2, the goal is to leverage existing theory to identify behavioral indicators of the constructs that are conceptually grounded. Sottilare and colleagues (2017) have provided a basis for this step. In Step 3, the focus shifts to identifying the available data sources, and specific system-based information, that is available from the environment of interest. The RADSM process is data source-agnostic, supporting data available from a variety of sources. In this case, the goal is to document the various data elements that can be captured from the training scenario. Within VBS3, this data might include text chat, positional data of all entities, sensor actions and results, and weapon fires and remaining ammunitions.

Once the behavioral indicators and list of available data or information is completed, Step 4 consists of bringing these two pieces together to operationalize the indicators using the types of data available in the environment. This transitions the conceptual nature of the behavioral indicators to specific, data-defined performance measures that can be implemented within GIFT. The intent is to develop several operationalized indicators of each teamwork skill, which could each provide unique insight into how the team is doing on that particular skill. It is important to note that any one indicator could be conceptually relevant to a number of teamwork skills, given the conceptual overlap of the teamwork constructs. The goal is to identify a set of indicators, that when used together, do the best job at assessing a unique teamwork skill, such as cohesion. The indicators tied to any one teamwork skill can be implemented and assessed individually, or aggregated to form a single, more comprehensive assessment of a teamwork skill. Table 1 provides examples of what this process looks like when developing measures of task cohesion.

Table 1. Example of RADSM Step 4 for Development of Task Cohesion Measures

Behavioral Marker	How would this be demonstrated?	Data Source(s)	Data Features	Analysis Method(s)
Members are actively working together and pitching in to reach team goals	All team members are communicating with each other	Chat logs	Sender/receiver of chats; number of chat messages sent by person	Compute # of messages sent by each team member; Assess the distribution of communication actions across team members
	Each team member is taking the actions that they are responsible for (e.g., detecting threats in their area)	Movement and action logs; List of team member responsibilities	Who did what action and when	Comparison of user movements/actions against their responsibilities
Occurrences of phrases like "great job everyone", "go team", "you're the best", "good work"; positive affirmations toward the team's work	Team members using these phrases in their chat communications with one another	Chat logs	Sender/receiver of chats; content of chat communications	Dialogue act analysis – sum instances of the use of words and phrases matching those associated with "positive affirmation"

Once the team has compiled a set of operationalized indicators, Step 5 of the RADSM process will focus on implementing these measures in the GIFT environment. During this step, the team specifies the criterion for each measure (e.g., thresholds for effective and ineffective assessments). For example, if the distribution of chat messages across the team is concentrated on one or two individuals, this may indicate low task cohesion and would signal the need for feedback.

Finally, in Step 6, the goal is to validate the measures of the teamwork skills developed in the previous steps. During the development phases of this effort, the primary focus of validation is establishing the face validity of measures. That is, individuals with expertise in teamwork measurement as well as Army SMEs will provide assessment of the utility and accuracy of each conceptual measure. In subsequent efforts, the team will develop and execute controlled experiments of the system using teams of active duty infantry Soldiers to establish and verify the validity of each measure.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This ongoing effort is aimed at training and assessing team performance within the GIFT environment. This serves two purposes: demonstrating that GIFT can be effectively extended from individual training to team training, and demonstrating that reliable and valid measurements of teamwork can be assessed in a virtual team training environment such as VBS3. Our progress to date has shown that there is ample opportunity to deliver rich training experiences through a VE and that there are a plethora of behavioral indicators and measurement opportunities within the scenario. Next steps for this effort include continued development, refinement, and implementation of the scenario inVBS3; development and implementation of the unobtrusive teamwork measures; development of feedback strategies; and validation of the measures through both face validation and rigorous human-in-the-loop experimentation. Future efforts will build upon this work, both in terms of the revised GIFT architecture for supporting team training, but also the scenario and team measures created.

REFERENCES

Army Doctrine Reference Publication (ADRP) 6-0, Mission Command. Washington, DC: Headquarters, Department of the Army, 2012.

Bonner, D., Walton, J., Dorneich, M. C., Gilbert, S. B., Winer, E. & Sottilare, R. A. (2015). The Development of a Testbed to Assess an Intelligent Tutoring System for Teams. In Workshop on Developing a Generalized Intelligent Framework for Tutoring (GIFT): Informing Design through a Community of Practice (p. 9).

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

Hughes, A. M., Gregory, M. E., Joseph, D. L., Sonesh, S. C., Marlow, S. L., Lacerenza, C. N., Benishek, L. E., King, H. B., & Salas, E. (2016). Saving lives: A meta-analysis of team training in healthcare. Journal of Applied Psychology, 101(9), 1266–1304.

McCormack, R.K., Brown, T.A., Orvis, K.L., Perry, S., & Myers, C. (2017). Measuring Team Performance and Coordination in a Mixed Human-Synthetic Team Training Environment. In the Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2017, No.1).

Orvis, K. L., Duchon, A., & DeCostanza, A. (2013, January). Developing Systems-based Performance Measures: A Rational Approach. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2013, No. 1). National Training Systems Association

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

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

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. International Journal of Artificial Intelligence in Education, 1-40.

Wilson, K. A., Salas, E., Priest, H. A., & Andrews, D. (2007). Errors in the heat of battle: Taking a closer look at shared cognition breakdowns through teamwork. Human Factors, 49, 243–256

ABOUT THE AUTHORS

Dr. Robert K. McCormack *is a Principal Mathematician and Lead for the Team and Organizational Performance Capability at Aptima. He has expertise in the areas of unobtrusive measurement, computational linguistics (NLP), machine learning, epidemiological modeling, and human sociocultural modeling and analysis. Dr. McCormack received a Ph.D. and M.S. in Mathematics from Texas Tech University, and a B.A. in Mathematics and Computer Science from Austin College*

Ms. Tara Kilcullen *is a Program and Customer Engagement Lead at Aptima, Inc. In her role, she supports business strategy, planning, and execution, defines market needs for technology requirements, product maturation and successful transition of science and technology (S&T) research programs, and leads defense programs that differentiate Aptima as an industry leader in Modeling, Simulation and Training. Ms. Kilcullen holds B.A.'s from the University of Pittsburgh and an A.S. from Full Sail University as well as several certifications.*

Dr. Anne M. Sinatra *is part of the adaptive training research team within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab, she is lead on the Team Modeling vector and works on the Generalized Intelligent Framework for Tutoring (GIFT) project. Her background is in Human Factors and Cognitive Psychology.*

Dr. Tara Brown *is a Senior Scientist and Lead for the Instructional Strategy and Support Capability at Aptima. She has expertise in unobtrusive measurement, training, and learning in individuals and teams. Dr. Brown holds a Ph.D. and M.A. in Organizational Psychology from Michigan State University and a B.S. in Psychology from Wright State University.*

Dr. Jeffrey M. Beaubien *is Lead Scientist for Aptima's Learning and Training Systems Division. He also serves as Aptima's Institutional Review Board (IRB) chair. Dr. Beaubien has over 15 years of experience designing and evaluating a broad range of training programs—including classroom-based training, interactive multimedia instruction, distance learning, simulation-based training, game-based training, and mobile learning—for clients in the healthcare, commercial aviation, homeland security, and military sectors. Dr. Beaubien holds a PhD in Industrial and Organizational Psychology from George Mason University, an MA in Industrial and Organizational Psychology from the University of New Haven, and a BA in Psychology from the University of Rhode Island.*

ACKNOWLEDGEMENTS

The research reported in this document was performed in connection with contract number W911NF-17-C-0061 with the U.S. Army Contracting Command - Aberdeen Proving Ground (ACC-APG). The views and conclusions contained in this document/presentation are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of ACC-APG, U.S. Army Research Laboratory or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.