

## **Military Team Training Utilizing GIFT**

**Desmond Bonner, Stephen Gilbert, Eliot Winer,  
Michael Dorneich, Anastacia MacAllister,  
Adam Kohl, Kaitlyn Ouversen, Anna Slavina  
Iowa State University  
Ames, Iowa**

**{Dbonner, Gilbert, Ewiner, Dorneich, Anastac,  
Adamkohl, Kmo, Aslavina} @ iastate.edu**

**Anne M. Sinatra  
Army Research Laboratory  
Orlando, Florida  
Anne.m.sinatra.civ@mail.mil**

### **ABSTRACT**

In 2015, the U.S. Army identified intelligent tutoring as a crucial resource for effective training of soldiers. Specifically, team training is essential as military missions are usually team-based and require extensive coordination. Intelligent Tutoring Systems (ITS) review actions taken by the user and provide dynamic instructions to teach subject matter to an individual. A team ITS assesses the performance of the teams' individuals, their overall performance as a team, and the interactions of that team to provide dynamic instructions. While extensive work has been conducted regarding single person ITSs, work regarding team-based ITS is limited. A team ITS is difficult to design as the tutor must account for the actions of multiple individuals and their team interactions. The tutor must teach task skills for completing the objective, and team skills for how a team works to meet the objective.

This paper describes the implementation, development and evaluation of a Team Intelligent Tutoring System for military teams. We faced challenges such as defining the appropriate levels of cognitive load and team communication required to be successful. The goal of the work was to evaluate an ITS's effectiveness in a simple team training scenario, a two-person surveillance task in which participants signaled each other using keystrokes. The scenario was constructed using Virtual Battle Space 2.0 (VBS2), and the tutor was built using the Generalized Framework for Tutoring (GIFT). Sixteen two-person teams were run through the study in one of three feedback conditions (individual feedback, team feedback, or no feedback). Their individual and team performance within the task were assessed. We found that participants in the feedback conditions had fewer extraneous keystrokes in the task than those without feedback.

### **ABOUT THE AUTHORS**

**Desmond Bonner** is a graduate student in the Human-Computer Interaction and Industrial Engineering program at Iowa State University's Virtual Reality Applications Center. His background is in Graphic Design. His work focuses on using games for learning to increase interest in Science, Technology, Engineering, and Mathematics (STEM) fields as well as leadership within small teams.

**Stephen Gilbert, Ph.D.**, is an associate director of the Virtual Reality Applications Center and assistant professor of Industrial and Manufacturing Systems Engineering at Iowa State University. His research interests focus on technology to advance cognition, including interface design, intelligent tutoring systems, and cognitive engineering. He is a member of IEEE and ACM and works closely with industry and federal agencies on research contracts. He is currently lead on a project supporting the U.S. Army Research Laboratory STTC in future training technologies for teams.

**Eliot Winer, Ph.D.**, is an associate director of the Virtual Reality Applications Center and associate professor of Mechanical Engineering and Electrical and Computer Engineering at Iowa State University. He is currently co-leading an effort to develop a next-generation mixed-reality virtual and constructive training environment for the U.S. Army. Dr. Winer has over 15 years of experience working in virtual reality and 3D computer graphics technologies on sponsored projects for the Department of Defense, Air Force Office of Scientific Research, Department of the Army, National Science Foundation, Department of Agriculture, Boeing, and John Deere.

**Michael Dorneich, Ph.D.**, is Associate Professor of Industrial and Manufacturing Systems Engineering and a faculty affiliate of the human computer interaction (HCI) graduate program at Iowa State University. His research interests

focus on creating joint human-machine systems that enable people to be effective in the complex and often stressful environments found in aviation, robotic, learning, and space applications. Dr. Dorneich has over 20 years' experience developing adaptive systems that can provide assistance tailored to the user's current cognitive state, situation, and environment.

**Anastacia MacAllister** is a Ph.D. student in Mechanical Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center. She is working on developing Augmented Reality work instructions for complex assembly and intelligent team tutoring systems.

**Adam Kohl** is a MS student in Mechanical Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center. He is working on developing pattern recognition techniques to enhance n-dimensional data visualization methods.

**Kaitlyn Ouversen** is a graduate student at Iowa State University studying Human-Computer Interaction and Industrial Engineering. She is currently working on user studies involving voting machines for voters with a disability.

**Anna Slavina** is a Ph.D. student at Iowa State University studying Human-Computer Interaction and Psychology. She is currently exploring how technology-based notetaking techniques affect long-term memory.

**Anne M. Sinatra Ph.D.**, is an Adaptive Tutoring Scientist at the Army Research Laboratory's SFC Smith Center. Her background is in Cognitive and Human Factors Psychology. She conducts adaptive training research as a member of the Learning in Intelligent Tutoring Environments (LITE) Lab. She works on the Generalized Intelligent Framework for Tutoring (GIFT) project and is the Team Tutoring research vector lead.

## **Military Team Training Utilizing GIFT**

**Desmond Bonner, Stephen Gilbert, Eliot Winer,  
Michael Dorneich, Anastacia MacAllister,  
Adam Kohl, Kaitlyn Ouversen, Anna Slavina  
Iowa State University  
Ames, Iowa**

**{Dbonner, Gilbert, Ewiner, Dorneich, Anastac,  
Adamkohl, Kmo, Aslavina} @ iastate.edu**

**Anne M. Sinatra  
Army Research Laboratory  
Orlando, Florida  
Anne.m.sinatra.civ@mail.mil**

### **INTRODUCTION**

This paper describes the development, implementation and evaluation of an Intelligent Tutoring System (ITS) designed for teams. An ITS is a software system which provides instruction to a human learner without the intervention of a human trainer beyond setting parameters of assessment. Team ITSs have the capability to provide efficient team training on task skills along with team skills. This training can be scalable and in turn provide complex organizations with cost effective training. For this project, a research team produced a military training focused team tutor and scenario.

The area of team tutoring has grown in significance with the prevalence of collaboration-focused education. Team problem solving is an important commodity within the workforce (De Dreu, & West, 2001). Previously, several individual training tutors have produced a majority of the literature on intelligent tutoring (e.g., Koedinger, Anderson, Hadley, & mark, 1997; Capuano, Marsella, & Salerno, 2000; Corbett, Koedinger, & Anderson, 1997). However, less is known about team intelligent tutors from previous work (Rickel & Johnson, 1998; Singley, Fairweather, & Swerling, 1999; Zachary, et al., 1999; Traum, Rickel, Gratch, & Marsella, 2003). Several difficulties were identified which include individual contribution, interaction, variable assessment, and team skills (Whatley, 2004; Miller, Yin Ioerger, Yen, & Volz, 2000; Schaafstal, Johnstn, & Olser, 2001; Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2014). This project is necessary for several reasons. These include the need for more work on effective team tutors, and low cost military training. Also, by developing this Team ITS, knowledge is added to the existing work providing future implementers with a guide towards building effective tools. To accomplish this, a tutor was built using the program Virtual Battle Space 2.0 (VBS2) and the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is a modular computer-based ITS framework that allows for tutor authoring, instructional management, and evaluation of ITSs (Sottolare, 2014). GIFT also provides users with the ability to develop ITS for individuals within several domains and environments (Sottolare, Brawner, Goldberg & Holden, 2012). Third-party programs are open to use with GIFT with the addition of its gateway module.

It is difficult to create intelligent tutors for teams due to several challenges (Bonner et al 2016). First, it is important to create an effective individual focused ITS, which by itself is a difficult task. This allows for a framework to create the team tutor. Next, it is important to understand the requirements of the team tutor that is being constructed. This includes making sure the task the tutor is assessing is appropriate, identifying which areas will need to be tutored to improve performance, and software with the capability to produce the appropriate level of feedback to learners based upon their performance (Kozlowski & Ilgen, 2006; Bonner et al., 2015a, Walton et al). Additionally, the tutor must be able to account for the different learning skills within the team.

For the team ITS presented in this paper, VBS2 and GIFT were utilized to create the Surveillance Task (Bonner et al., 2015b; Bonner et al., 2016). Within the task, two learners (or players) are responsible for monitoring two adjacent sectors. Their performance is monitored by the tutor, which depending on the setting, provides the learners with team, or individual feedback.

## **BACKGROUND**

Future Team ITS implementation would benefit from having clearly defined authoring strategies. Previous efforts have provided road maps for authoring intelligent tutors effectively, however, our work identifies pitfalls which may not have been adequately addressed (DeShon, Kozlowski, Schmidt, Milner, and Wiechmann, 2004). For tutor authoring, areas of relevant work include team training, feedback, and previous ITS implementations involving individuals and/or teams. Several aspects involving teams include team training, feedback, and previous ITS iterations. For example, effective teams are defined by their structure, type, and roles (Bonner et al, 2015a; Singh, Dong, & Gero, 2012; Salas, Burke, Fowlkes, & Priest, 2004; Johnston, Serfaty, & Freeman, 2003). An ITS should improve performance by producing effective feedback to learners (Walton et al, 2014). Additionally, the tutor needs to be flexible to meet the ever-changing needs of learners.

### **Defining Tutors**

The key question is whether a computer tutor can be as effective as a human tutor. VanLehn (2011) refers to a human tutor as usually an adult expert in a field, who works with a single student. The human tutor can serve as support to the primary instruction format (e.g., a classroom) through delivering remedial or advanced instruction. This is determined by how they perceive the students' performance and learning preferences. This dynamic is viewed as the most effective form of tutoring, which justified more effort being put into a computer-based tutor focusing on an individual (Graesser, Vanlehn, Rose, Jordan, & Harter, 2001).

Computer-based tutors are categorized in one of two ways, Immediate or Dialog based feedback (VanLehn, 2011). They can provide immediate feedback to learners about their performance within a task, such as if they were correct or incorrect. The tutors can then provide feedback in a debriefing. Or the tutors can give feedback throughout the process of the task, such as in a dialog based tutor format where feedback is provided in the form of hints, suggestions, or reminders. The learners can then take the feedback and adjust their performance accordingly to modify their actions (VanLehn, 2011).

### **Previous ITS Strategies**

Some past ITS were assessed to develop a strategy for success with the military team tutor. These include the Advanced Embedded Training System (AETS) and Social Interaction Tutor (Zachary et al, 1999; Kumar et al, 2010). AETS was developed for Naval Air defense training. It assisted human training in assessing team performance. It monitored team members' actions such as keystrokes, speech and eye movements. It then cross-referenced the actions with expectations for ideal individual and team behavior. However, the output of AETS was used by a human trainer, not part of an automated feedback system for the learners.

The Social Interaction Tutor focuses on multiple learners individually (Kumar et al, 2010). It allows multiple learners to interact with a collaborative learning environment within the task of completing a design competition. The tutor acts as a conversational and pedagogical agent, providing feedback in a virtual chatroom. Learners receive immediate feedback, which is effective for a design competition, but would not be effective for a continuous task.

Team Tutors should also be designed to handle interactions between teams. Previous work has indicated the importance of promoting interaction within teams to promote members' sense that their individual successes correlate to the success of their teammates (Singley, Fairweather, & Swerling, 1999). According to DeShon, Kozlowski, Schmidt, Milner, and Wiechmann (2004), when individual tasks, such as reporting opposing forces observed, contribute directly to team performance, e.g., in a larger surveillance mission, an individual may improve the performance of the team while focusing on his or her individual tasks.

In some learning scenarios, computer tutors have been included with learner teams. For example, a study using a hands-on learning experience placed learners in separate roles and provided a tutor with guidelines for each (Rickel & Johnson, 1998). The tutor was also given instructions that some of the roles were interdependent. However, the priority for the tutor was to assess the learners primarily on their individual tasks and provide feedback instead of providing feedback regarding the teams' interactions. In the Surveillance Task described in the current paper, the individual tasks overlap with team tasks but can also be achieved independently such that performing well on individual tasks will not necessarily lead to performing well on team tasks.

## Feedback

Effective tutors are designed with the appropriate type of feedback (Walton et al, 2014). Feedback refers to how an individual or team is given instruction or constructive criticism on their performance with the aim to improve it (Hattie & Timperly, 2007). Several feedback considerations need to be made such as type, timing, location, and modality so that the learner can comprehend progress and performance levels (Hattie & Timperly, 2007; Nadler, 1979). The Surveillance Task compares feedback types associated with teams and individuals to determine the most effective method for future team ITS implementations. While DeShon et al. (2004) note that there have been relatively few studies examining the effects of individual and team feedback on team performance there is a danger to including both types of feedback. Too much task-based feedback during a task might actually be detrimental to performance (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1998). In order to be useful, feedback must be attended and processed, which can lead to decrements in performance if feedback is delivered during a particularly complex task (Osman, 2012). Written feedback requires diverting processing resources from the task at hand to the feedback, which is problematic when one group receives both individual and team feedback while the other only receives team feedback.

## Teamwork & Team Training

A team is defined as “a number of persons associated together in work or activity” (Merriam-Webster, 2017). The research team had previously created a taxonomy describing the considerations necessary for creating a team base tutor with GIFT (Bonner et al, 2015a). A primary consideration was the team structure. Within a team, there may be specific roles team members play. This is important to consider when assessing teams (Salas et al, 2004, Sottolare et al, 2011). For example, a standard five-person fire team consists of a team leader, assault weapon specialist, assistant assault weapon specialist, and two regular soldiers. Each has a role which helps to meet the needs of the team. This type of team is also more flexible than others such as a fire support team that operates a long-range artillery weapon. The fire support team can perform artillery missions, and defensive security missions. However, the fire team can perform different missions such as movement to contact, assault on a fixed position, and fixed reconnaissance. There is also the division of skills to consider. Each team member has a specific skill such as the assault weapon specialists. The team leader must also know how to best utilize each of their team members. This understanding is more effective in experienced teams as the leader is familiar with them through having spent time together.

## The Nine C's of Teamwork

A large portion of our work came from what Salas, Thayer, Bedwell, and Lazzara (2015) called the C's of teamwork. These are nine factors of team behavior and teamwork: Culture, Context, and Composition (which comprise the influencing conditions), Cooperation, Coordination, Cognition, Conflict, Coaching, and Communication (which comprise the core processes) (Salas et al., 2015). The C's determine ways which teamwork measures can be created. This serves the purpose of helping to assess them within our team ITS study in which we utilize *Coordination*, *Communication*, and *Context*. *Coordination* involves the use of team skills or task skills to achieve a goal. *Communication* refers to how teammates decided on their approach and generally interact. *Context* meanwhile refers to the team environment. In the Surveillance Task, participants perform interdependent tasks, which test their *Coordination*. They *Communicate* to meet the goal or adjust their strategy, and exist in a one-time pairing. In future tasks, we will utilize more C's. These are further explored in Bonner et al. (2017 in-press).

## METHODS

### Research Objectives

The objective of this research was to evaluate an ITS's effectiveness for team training. This study also compared two smaller approaches towards feedback for either individuals or teams, which began to address the larger picture of the correct approach for providing feedback to teams. An Institutional Review Board (IRB) at ISU and the Army Research Laboratory approved evaluation of the Surveillance Task Team ITS as described.

### Hypothesis

The hypotheses of this study were as follows:

1. Participants will use feedback provided by the tutor to improve their individual performance, which will also improve their team's performance of the Surveillance Task.
2. Teams receiving feedback at the team level will have higher team scores than the teams receiving tutor feedback at the individual level.

## Participants

The participants of this study are representative of the target age range as they are over the age of 18. They either attended or worked at a large university located in the Midwest with age ranging from 18-35. A total of 32 people (16 teams) participated (22 male, 10 female). Additional teams were recruited, but their data were lost due to technical challenges. No federal employees were in the study.

## Surveillance Task

The Surveillance Task scenario is modelled after a team security operation (Bonner et al, 2015). A pair of distributed participants are located in adjacent offices with desktop computers running the scenario using GIFT, an intercom to communicate, and headsets for scenario audio. Each participant is in control of an avatar located atop a roof in the scenario. They are responsible for their half of a 360-degree environment (180 degree sectors) where opposing force (OPFOR) avatars appear at random and move from one half of the area to the other. The participants must continuously scan their sector. Each participant's screen contains a primary window showing the surveillance zones in VBS2, with a narrow left-side frame containing text-based feedback from the GIFT tutor. See previous work for a more detailed description of the Surveillance Task (Bonner et al., 2015b; Bonner et al., 2016).

Two participants make up one team. They work together to complete the Surveillance Task over four trials lasting five minutes each. To complete the Surveillance Task trials, teams must complete three subtasks: Transfer, Acknowledge, and Identify. A GIFT condition file was written in Java for each subtask to assess player performance. These Java condition files and other information were then combined in GIFT Domain Knowledge Files (DKFs) to continuously evaluate players within the GIFT architecture and to trigger feedback given to players.

*Transfer.* To transfer OPFOR, participants must determine the OPFOR's travel pathway. As they approach the boundaries (notified by the green arrows), the participants press either the 1 or 2 key according to the zone of transfer to notify their team of the impending OPFOR zone transfer. Participants should additionally provide a verbal warning to their partner through the intercom.

*Acknowledge.* Participants acknowledge the impending transfer by pressing the "E" key immediately after they receive feedback in VBS2 from their partner. They then may confirm the transfer with their team member over the intercom.

*Identify.* Once OPFOR are transferred, the teammate's zone to which the OPFOR are moving must press the spacebar key in order to identify the OPFOR once it enters their area. Thus, they now assume responsibility for tracking its movement.

## Independent Variables

There were two independent variables: **Trial Order** (1, 2, 3, 4) and **Feedback** (none, individual, team) as shown in Table 1. The first independent variable (within subjects) is **Trial Order**. Participants conducted four trials where they complete one 5-minute Surveillance Task in each trial. The second independent variable for the study (between subjects) is the **Feedback** received by participants. There were two types of feedback:

- Feedback about an individual task, based on an assessment of that individual task
- Feedback about a team task, based on an assessment of that team task

In the **Individual Feedback condition**, all feedback was provided based on the individual performance of the tasks. For example, a participant in the individual condition would receive feedback from the tutor specifically about how he or she was doing in the individual task, and that participant would be the only one to have access to it ("Player 1, Identify OPFOR"). In the **Team Feedback condition**, both participants would receive feedback about how the team

was doing overall such as “Team, your communication needs work.” For the three tasks, the table below (Table 2) gives some examples of feedback. In the **No Feedback condition**, participants receive no feedback at all.

**Table 1. Feedback by Trial**

	# Teams	Trial # 1	Trial # 2	Trial # 3	Trial # 4
<b>No Feedback</b>	6	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)
<b>Individual Feedback</b>	4	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)
<b>Team Feedback</b>	6	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)	Surv. Task (5 min)

**Table 2. Individual and Team Feedback Examples**

Individual Task	Individual Feedback
<b>Transfer</b>	“It is important to effectively communicate crossings”
<b>Acknowledge</b>	“It is important to confirm at appropriate times”
<b>Identify</b>	“It is important to identify OPFOR as quickly as possible”
Team Task	Team Feedback
<b>Transfer-Ack. pairs</b>	“Report Transferring OPFOR to Team”
	“It is important to confirm at appropriate times. Your team communication needs work”
<b>Identify</b>	“Alpha Team, Identify OPFOR Immediately”

### Dependent Variables / Metrics

The dependent variables include Individual Performance, Team Coordination, Team Performance, and Team Familiarity (see Table 3). Individual Performance was determined by a number of metrics based on the three individual subtasks, Transfer, Acknowledge, and Identify. Transfer is assessed based on the percentage of crossing OPFOR who were transferred. Acknowledge is assessed based on the percentage of crossing OPFOR who were acknowledged. Identify was assessed based on the percentage of crossing OPFOR who were identified, as well as on the timing of the identify, with identification within one second being optimal. Each individual was also rated on Task Errors. Sometimes participants typed additional keystrokes that were not needed, e.g., when one acknowledgment was required, the participant typed the E key twice. That would count as one extra keystroke, and extra keystrokes for all three tasks are counted in Task Errors.

Team Coordination was determined by the success of the three tasks happening in succession, and is a measure of the team’s communication ability. Every OPFOR who is transferred by one teammate should have an acknowledge from the other teammate. The percentage of time that happened was one metric (Transfer-Acknowledge Percent) and the timing of that acknowledgment was another metric (Transfer-Acknowledge Time). Finally, the last coordination metric was based on the percentage of OPFOR for which all three events were present: Transfer, Acknowledge, and Identify (Transfer-Acknowledge-Identify Triples Percent). Team Performance was based on a weighted sum of all of the above measures, averaged across the two teammates.

### Procedure

The Surveillance Task was introduced to participants as part of an experiment. The purpose of the study was presented in the online consent form. Participants were then provided with a pre-survey, and then they were allowed to sign up for an experiment time slot. The study lasts for one 60-minute session. Once participants arrived, they were given an introduction to the study. Next, experimenters randomly designated the participants as “Player 1” and “Player 2.” After meeting their partners briefly, the players entered different offices to complete the remainder of the study. After completing the familiarity survey, the participant designated as “Player 1” underwent eye-tracking calibration. A ten-minute training and practice session followed so that participants could begin to get a feel for the environment as well as the controls. After the training practice, participants were asked to explain the task back to the experimenters to

**Table 3. Dependent Variables/Metrics**

<b>Dependent Variable</b>	<b>Variable Name</b>	<b>Metric</b>	<b>Frequency</b>
<b>Individual Performance</b> Transfer	Transfer Percent	(# Transfer) / (# OPFOR crossed)	Each trial (4x)
Acknowledge	Acknowledge Percent	(# of Ack) / (# of OPFOR crossed)	Each trial (4x)
Identify	Identify Percent	(#IDs) / ( # OPFOR Crossed)	Each trial (4x)
Identify	Cross-ID Time	(ID time – Time Crossed): 5 if < 1 s 4 if < 2 s 3 if < 3 s 2 if < 4 s 1 if < 5 s 0 if 5+ s	Each trial (4x)
Task Errors All Tasks	Extra Keystrokes (typed in error)	(Extra Trans + Extra Ack + Extra ID keys): 5 if 0 extra keystrokes 4 if 1 extra keystroke 3 if 2-3 extra keystrokes 2 if 4-5 extra keystrokes 1 if 6-7 extra keystrokes 0 if 8+ extra keystrokes	Each trial (4x)
Individual Performance All Tasks	Player Performance	Weighted sum of: Transfer Percent Acknowledge Percent Identify Percent Task Errors Cross-ID Time Trans-Ack. Time (–Percent OPFOR transferred after crossing)	Each trial (4x)
<b>Team Coordination</b> Transfer & Acknowledge	Transfer-Acknowledge Percent	(# Transfer-Ack pairs) / (Total # transfers)	Each trial (4x)
<b>Team Coordination</b> Transfer & Acknowledge	Transfer-Acknowledge Time	(Ack time – Trans time): 5 if < 0.5 s 4 if < 1.0 s 3 if < 1.5 s 2 if < 2.0 s 1 if < 2.5 s 0 if 2.5+ s	Each trial (4x)
<b>Team Coordination</b> Transfer, Acknowledge, & Identify	Transfer-Acknowledge-Identify Triples Percent	(# Trans-Ack-ID triples) / (Total # OPFOR crossed)	Each trial (4x)
<b>Team Performance</b> (Task: all tasks)	Team Performance	Weighted sum of players: Average Transfer Percent Average Ack. Percent Average Identify Percent Average Triples Percent Average Task Errors Average Trans-Ack Time Average Cross-ID Time	Each trial (4x)
<b>Team Familiarity</b>	Familiarity	0 if never met teammate 1 if have met teammate	Before Study

confirm that they understood what was supposed to be done. Once training was complete, the first of four trial sessions commenced. Immediately after session trial, participants completed a short post-trial survey. This followed for Trials 2-4. However, after completing Trial 4 and its post-trial survey, participants additionally completed a Post-Experiment Survey and then participated in a debriefing with the experimenters. Once the debriefing was complete, the participants were compensated \$15 each.



Each of the two participants were situated in a room with a desktop computer, wireless headset, and landline speakerphone which served as an intercom for communication. There was an auxiliary laptop for completing the surveys. In Player 1's room, there was also an eye tracker. Each participant was also supervised by one experimenter.

## RESULTS

To describe the results, it is useful to show the data visually. Figure 1 shows Team Performance for each team across its four trials. Teams are also ordered by feedback condition, with Trial 1 at left of each cluster, and Trial 4 at right of each cluster. Note that 1) no team achieved maximum performance of 10.0, 2) most teams improved over time, and 3) some teams performed better than other teams.

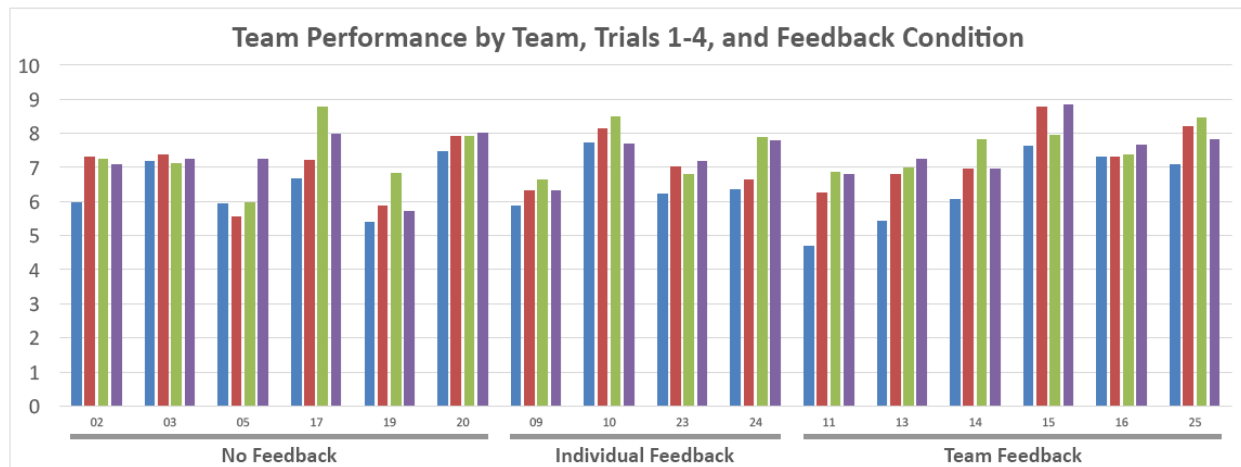


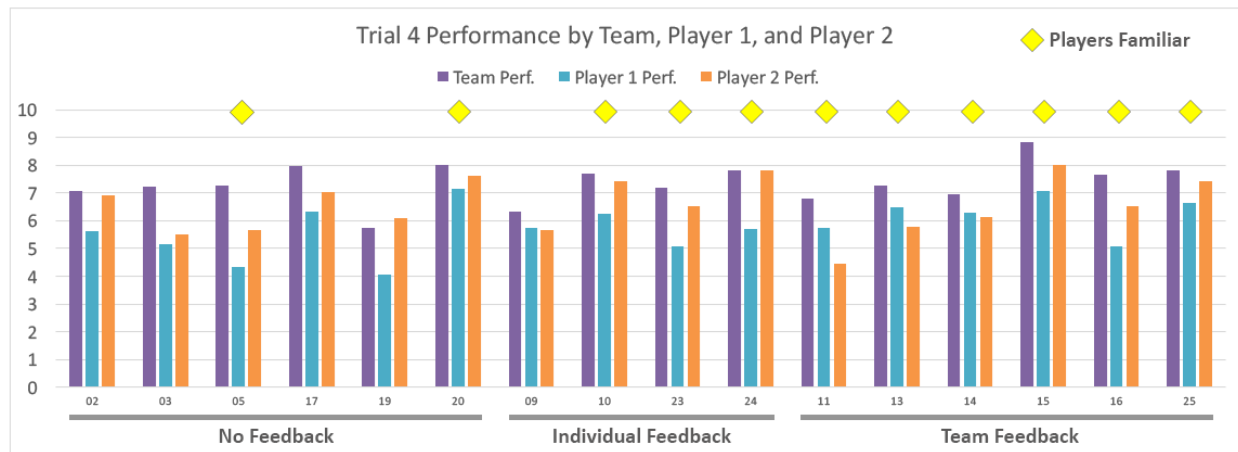
Figure 1. Team Performance by Team, Trials, and Feedback Condition

To evaluate the effect of feedback condition (between-subjects variable) and trial (within-subjects variable) on Team Performance, and whether there was an interaction between feedback and trial, a two-way mixed ANOVA was used. The assumptions for this ANOVA were met appropriately: there were no outliers, as assessed by boxplot; the data were normally distributed, as assessed by Shapiro-Wilk's test of normality ( $p > .05$ ); there was homogeneity of variances ( $p > .05$ ) and covariances ( $p > .05$ ), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively; and Mauchly's test indicated that the assumption of sphericity was met for the two-way interaction. After the analysis, there was no statistically significant interaction between the feedback condition and trial on team performance. There was also no main effect of feedback condition on the team performance of each group. However, the main effect of trial showed a statistically significant difference in mean team score at the different time points,  $F(3, 42) = 14.297, p < .0005$ , partial  $\eta^2 = .505$ . **Trial 1 Team Performance was significantly lower than Trials 2, 3, and 4.** Trials 2, 3, and 4 were not significantly different.

To further investigate whether there might be any effect of feedback condition on individual or team performance, the independent variable for feedback condition was simplified to FeedbackPresent (yes if team or individual feedback, no if no feedback), the effect of FeedbackPresent on each of the dependent variables in Table 3 was examined in only Trial 4, after participants had the most experience with the task and feedback if it was present. In this situation, because trial is no longer an independent variable, and FeedbackPresent has only two levels, an independent samples t-test is appropriate if its assumptions of no outliers, homogeneity of variances, and normality, are met. For several of the variables, normality was not met per Levene's test, and thus the Mann Whitney U test was used instead. Per this test, it was discovered that while most variables were not significantly different, three variables were of interest. The number of error keystrokes that Player 2 typed during Trial 4 while transferring was a mean of 2.5 with no feedback and a mean of 1 with feedback, with a significant difference ( $p = .035, U = 12, Z = -2.105$ ). The number of error keystrokes that Player 1 typed during Trial 4 while acknowledging was a mean of 2 with no feedback and a mean of 1 with feedback, with a significant difference ( $p = .045, U = 13, Z = -2.005$ ). **The number of error keystrokes was significantly lower in the feedback condition than in the no feedback condition.**

Finally, the standard deviation of the team coordination variable Transfer-Acknowledge Time had a mean value of 1.375 seconds for teams with no feedback and mean value of 0.788 for teams with feedback. This difference per Mann Whitney U ( $p = .065$ ,  $U = 13$ ,  $Z = -1.845$ ,  $\eta^2 = .23$ ), was not statistically significant. However, the effect size ( $\eta^2$ ) was large, suggesting that with a larger sample size, the difference might become significant. ***This result suggests that in a future study with more teams, teams with feedback may be able to coordinate the Transfer-Acknowledge tasks more consistently (with lower variance) than teams with no feedback.***

It is also interesting to view the results of team performance plotted alongside each teammate's performance (Figure 2). This figure focuses only on Trial 4, since that trial should represent each team's most practiced performance. Each team's team score and respective Player 1 and Player 2 scores are shown, along with a yellow diamond if the players were familiar (they had met before). It is worth noting that some teams have players who perform similarly (e.g., 03, 20, 09, and 14), while other teams have players with very different performance (e.g., 19, 24, and 16).



**Figure 2. Trial 4 Performance by Team, Player 1, and Player 2.**

We also created a variable called Team Gap based on the magnitude of the difference between player scores. However, neither feedback condition nor trial nor Team Familiarity had any effect on Team Gap. Team Familiarity had no significant impact on any of our dependent variables. It is worth noting that we did not control for familiarity during participant recruiting and therefore had an unbalanced sample across conditions. In the future, this will be corrected.

### Limitations and Assumptions

Even though the study included a military relevant task, we were not able to use military personnel in the study because of the complexity of having an IRB which would include some participants who were students outside ROTC, some students who were part of ROTC, and some non-students, each of which would have required different compensation methods to meet federal guidelines.

### DISCUSSION & CONCLUSION

Our original hypotheses were that 1) participants will use feedback provided by the tutor to improve their individual performance, which will also improve their team's performance of the Surveillance Task, and 2) teams receiving feedback at the team level will have higher team scores than the teams receiving tutor feedback at the individual level. The results of this study supported Hypothesis 1, in that receiving either individual or team feedback assisted participants in decreasing errors in their performance. Also, teams with feedback had approximately half the standard deviation in their Transfer-Acknowledge times by Trial 4 compared with the teams with no feedback, suggesting that feedback aided teams in performing more consistently.

We had hoped for a stronger impact of the feedback on team performance, and to see a different impact of team feedback vs. individual feedback. Hypothesis 2 was not supported in our study, in that the type of feedback (individual

vs. team) did not significantly affect performance. We suggest that the variation in performance was caused less by the feedback type and more by the individual differences among participants, i.e., their team skills. By accounting more for individuals' existing skills in future studies, we hope to remove that factor from the analysis.

Also, while we found no specific effects of teammate familiarity in this study, we do anticipate that the effects of familiarity, if better balanced across larger participant groups, would be more recognizable. In the case of the impact of feedback on performance, however, the variation in individual and team performance appears random enough that it likely does not stem primarily from feedback, and thus, a larger sample size in each feedback group may not help.

Our results did verify that four trials is sufficient to at least minimally master the task, since performance on Trials 2, 3, and 4 were not found to be significantly different, while team performance in Trial 1 was significantly lower than in other trials. Perhaps in future studies, we can find sufficient data with three trials.

Also, we noted above that team member performance seems to vary widely, and that some teams had members who are matched in performance, while others did not. This gap in performance, if studied in the future, will likely have implications for providing feedback differently to teams with a higher gap between members vs. teams with members performing equally well. It will also be interesting to explore in future work how this gap affects the trust among team members.

Lastly, we note that this paper describes the pedagogical impact of the first team tutor designed in GIFT, which demonstrates that it can be done and illustrates some of the challenges that lie ahead for team tutoring in the future.

## **ACKNOWLEDGMENTS**

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

## **REFERENCES**

- Bonner, D., Gilbert, S., Dorneich, M. C., Burke, S., Walton, J., Ray, C., & Winer, E. (2015b, February). Taxonomy of teams, team tasks, and tutors. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (p. 189).
- Bonner, D., Walton, J., Dorneich, M. C., Gilbert, S. B., Winer, E., & Sottolare, R. A. (2015a). The Development of a Testbed to Assess an Intelligent Tutoring System for Teams. In *AIED Workshops*.
- Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., ... & Holub, J. (2016a, September). The Challenges of Building Intelligent Tutoring Systems for Teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1981-1985). SAGE Publications.
- Bonner, D., Slavina, A., MacAllister, A., Holub, J., Gilbert, S., Sinatra, A. M., ... & Winer, E. (2016b). The Hidden Challenges of Team Tutor Development. In *Proceedings of 4th Annual GIFT Users Symposium (GIFTSym4)* (pp. 49-60).
- Capuano, N., Marsella, M., & Salerno, S. (2000, June). ABITS: An agent based Intelligent Tutoring System for distance learning. In *Proceedings of the International Workshop on Adaptive and Intelligent Web-Based Education Systems, ITS*.
- Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. *Handbook of human-computer interaction*, 5, 849-874.
- DeShon, R. P., Kozlowski, S. W., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of applied psychology*, 89(6), 1035.
- De Dreu, C. K., & West, M. A. (2001). Minority dissent and team innovation: the importance of participation in decision making. *Journal of applied Psychology*, 86(6), 1191.
- Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI magazine*, 22(4), 39.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current directions in psychological science*, 7(3), 67-72.

- Koedinger, Kenneth R., et al. "Intelligent tutoring goes to school in the big city." (1997).
- Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological science in the public interest*, 7(3), 77-124.
- Kumar, R., Ai, H., Beuth, J., & Rosé, C. (2010). Socially capable conversational tutors can be effective in collaborative learning situations. In *Intelligent tutoring systems* (pp. 156-164). Springer Berlin/Heidelberg.
- Merriam-Webster. (n.d.). Teamwork. Retrieved May 04, 2017, from <https://www.merriam-webster.com/dictionary/teamwork>
- Miller, M. S., Yin, J., Volz, R. A., Ioerger, T. R., & Yen, J. (2000, June). Training teams with collaborative agents. In *International Conference on Intelligent Tutoring Systems* (pp. 63-72). Springer Berlin Heidelberg.
- Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance*, 23(3), 309-338.
- Osman, M. (2012). The role of reward in dynamic decision making. *Frontiers in Neuroscience*, Vol.6
- Rickel, J., & Johnson, W. L. (1998, May). STEVE (video session): a pedagogical agent for virtual reality. In *Proceedings of the second international conference on Autonomous agents* (pp. 332-333). ACM.
- Salas, E., Burke, C. S., Fowlkes, J. E., & Priest, H. A. (2004). On measuring teamwork skills. *Comprehensive handbook of psychological assessment*, 4, 427-442.
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, 54(4), 599-622
- Schaafstal, A. M., Johnston, J. H., & Oser, R. L. (2001). Training teams for emergency management. *Computers in Human Behavior*, 17(5), 615-626.
- Sottolare, R. A. (2014). Examining opportunities to reduce the time and skill for authoring adaptive intelligent tutoring systems. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (p. 3).
- Singh, V., Dong, A., & Gero, J. S. (2012). Computational studies to understand the role of social learning in team familiarity and its effects on team performance. *CoDesign*, 8(1), 25-41.
- Singley, M. K., Fairweather, P. G., & Swerling, S. (1999, December). Team tutoring systems: Reifying roles in problem solving. In *Proceedings of the 1999 conference on Computer support for collaborative learning* (p. 66). International Society of the Learning Sciences.
- Sottolare, R. A., Holden, H. K., Brawner, K. W., & Goldberg, B. S. (2011). Challenges and emerging concepts in the development of adaptive, computer-based tutoring systems for team training. ARMY RESEARCH LAB ORLANDO FL HUMAN RESEARCH AND ENGINEERING DIRECTORATE.
- Sottolare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT).
- Traum, D., Rickel, J., Gratch, J., & Marsella, S. (2003, July). Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems* (pp. 441-448). ACM.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Walton, J., Dorneich, M. C., Gilbert, S., Bonner, D., Winer, E., & Ray, C. (2014). Modality and timing of team feedback: Implications for GIFT. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (p. 199).
- Whatley, J. (2004). An agent system to support student teams working online. *Journal of information technology education*, 3, 53-63.
- Zachary, W., Cannon-Bowers, J. A., Bilazarian, P., Kreckler, D. K., Lardieri, P. J., & Burns, J. (1998). The Advanced Embedded Training System (AETS): An intelligent embedded tutoring system for tactical team training. *International Journal of Artificial Intelligence in Education (IJAIED)*, 10, 257-277.