9-8-2016

The Challenges of Building Intelligent Tutoring Systems for Teams

Desmond Bonner  
*Iowa State University*, dbonner@iastate.edu

Stephen B. Gilbert  
*Iowa State University*, gilbert@iastate.edu

Michael C. Dorneich  
*Iowa State University*, dorneich@iastate.edu

Eliot Winer  
*Iowa State University*, ewiner@iastate.edu

Anne M. Sinatra  
*United States Army Research Laboratory*

*See next page for additional authors*

Follow this and additional works at: [http://lib.dr.iastate.edu/imse_conf](http://lib.dr.iastate.edu/imse_conf)

Part of the [Computer-Aided Engineering and Design Commons](https://lib.dr.iastate.edu/imse_conf), [Operational Research Commons](https://lib.dr.iastate.edu/imse_conf), [Other Operations Research](https://lib.dr.iastate.edu/imse_conf), [Systems Engineering and Industrial Engineering Commons](https://lib.dr.iastate.edu/imse_conf), and the [Systems Engineering and Multidisciplinary Design Optimization Commons](https://lib.dr.iastate.edu/imse_conf)

**Recommended Citation**

[http://lib.dr.iastate.edu/imse_conf/23](http://lib.dr.iastate.edu/imse_conf/23)

This Conference Proceeding is brought to you for free and open access by the Industrial and Manufacturing Systems Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Industrial and Manufacturing Systems Engineering Conference Proceedings and Posters by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
The Challenges of Building Intelligent Tutoring Systems for Teams

Desmond Bonner¹, Stephen Gilbert¹, Michael C. Dorneich¹, Eliot Winer¹, Anne M. Sinatra², Anna Slavina¹, Anastacia MacAllister¹, Joseph Holub¹
¹Iowa State University
²U.S. Army Research Laboratory

Intelligent Tutoring Systems have been useful for individual instruction and training, but have not been widely created for teams, despite the widespread use of team training and learning in groups. This paper reviews two projects that developed team tutors: the Team Multiple Errands Task (TMET) and the Recon Task developed using the Generalized Intelligent Framework for Tutoring (GIFT). Specifically, this paper 1) analyzes why team tasks have significantly more complexity than an individual task, 2) describes the two team-based platforms for team research, and 3) explores the complexities of team tutor authoring. Results include a recommended process for authoring a team intelligent tutoring system based on our lessons learned that highlights the differences between tutors for individuals and team tutors.

INTRODUCTION

This practice paper describes a two-year period during which a research team at a U.S. state university developed successful Team Intelligent Tutoring Systems (ITSSs) through the efforts of two separate projects. The projects produced a military training and a civilian focused tutor. Previous implementations of individual and team ITSSs are explored. Also, the lessons learned within development are described to create a general set of authoring guidelines for Team ITSS practitioners and researchers exploring team feedback.

An ITSS is a computerized system or software which provides immediate instruction specific to a learner without the involvement of a human trainer. Previously, ITSSs have been successfully utilized in several domains such as education and engineering (Koedinger, Anderson, Hadley, & Mark, 1997; Faria, Silva, Vale, & Marques, 2009). However, they have been designed mostly for use by a single student in one-to-one scenarios, with few being developed for teams (Sottilare, Brawner, Goldberg, & Holden, 2011). Meanwhile, successful human-led team training has been ongoing.

Human-led training can be particularly labor intensive because it requires trainers to be experienced with the subject matter to a point where they can have sufficient knowledge as to provide appropriate instruction, identify and correct trainee mistakes, and specialize the curriculum to suit a trainee’s learning style. This is difficult because time is required for a trainer to acquire a majority of the competencies within a subject field and to effectively train team skills. Also, human trainers are more effective in one-to-one tutoring scenarios (Koedinger et al., 1997).

The development of a team ITSS allows for less reliance on humans’ becoming experienced trainers and utilizes the effectiveness of individual ITSSs (one-to-one) by expanding the tutoring architecture to small groups (Bonner, et al, 2015). In discussing the authorship of an ITSS, it is important to consider two aspects. First, the process of just creating an individual ITSS is difficult. Also, there are different aspects of team training to be considered when creating a team tutor. For instance, the task that teams will be completing has a significant influence on the design of a team tutor. Team tasks consist of several factors such as tasks with time constraint and interdependent tasks which are further described in a taxonomy of team task characteristics (Bonner et al., 2015). It serves as a checklist of key factors to be considered when designing a team ITSS. These tasks are completed within the context of a training environment as to provide an accurate assessment of performance. The research team has been developing a team ITSS test bed for the military that conducts and assess team training. The program Virtual Battlespace 2.0/3.0 (VBS2/3) and the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Goldberg, & Holden, 2012) were used to create the Recon Scenario (or Recon Task). Work for developing the Team ITSSs has also been informed by utilizing aspects of a Multiple Errands Test (MET).

Requirements for a Team ITSS

Three core components are necessary for building a Team ITSS. First, a team task is needed that can be trained and evaluated by a computer system. Team task characteristics include factors such as task interdependence, routine, complexity, time constraints, information exchange, environmental fidelity, task solutions and task type (Bonner et al. 2015). Next, feedback is required for both individuals and the entire team to help improve performance as it is a key component of the learning process (Kozlowski & Ilgen, 2006). Finally, software is needed that can give the learner feedback based on the performance of the learner. Typically, ITSSs focus on only one learner. In a Team ITSS, the task yields not only information from multiple individual learners, but also typically yields information about the interaction between the learners themselves in addition to the tutor. The amount of information increases more than the square of the number of learners rather the additively, which requires a robust system which can handle the data.
RELATED WORK

Relevant research areas include team training, feedback, and past implementations of team ITSs. Previous work explores what different teams consist of and the roles within them (Bonner et al., 2015; Singh, Dong, & Gero, 2012; Salas, Burke, Fowlkes, & Priest, 2004; Johnston, Serfaty, & Freeman, 2003). Several aspects such as team structure and skillset are important to consider as when training teams it is necessary to train for team skills in addition to task skills.

When assessing performance, the purposes and members should be considered (Salas et al., 2004; Sottilare et al., 2011). An important aspect of team training is how feedback differs within the team setting as opposed to when it is specific to an individual (Walton et al., 2015; Nadler, 1979).

Team Training & Feedback

Team Training typically involves either feedback during the task or after-action review (AAR). Feedback is any form of information provided to a receiving person by an outside source (Hattie & Timperley, 2007). These outside sources or agents can come in the form of teachers, peers, environments, and the ITS. Meanwhile, an AAR refers to feedback or a critique of performance within a collective training task immediately following its conclusion. Its goal is to provide a detailed assessment to a unit on individual and collective performance according to the desired outcome (Morrison & Meliza, 1999).

Feedback research suggests that individual or team feedback is the only type to be given in a team setting (Nadler, 1979). Additionally, these types have a degree of influence on how participants’ focus (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004). However, clarification is needed for the amount of feedback and which level leads to the largest increase in performance (Gabelica, Bossche, Van Den, Segers, & Gijselaeirs, 2012). Also, it is important to understand the impact of public feedback. According to Gabelica et al. (2012), individual feedback should have some degree of public feedback characteristics, despite the potential of public embarrassment. It’s not clear how best to consistently choose the best feedback characteristics (e.g., public or private, intensity) and its timing depending on the team task.

Previous Team ITSs

Work from Sottilare et al. (2015) provides a roadmap for team ITSs through their model for team adaptive training and education, but it is also worth examining actual team training scenarios that have been created.

The Advanced Embedded Training System (AETS) was an attempt by the U.S. Navy to apply ITS concepts such as problem-based learning, cognitive diagnosis, and others to team training on ships (Zachary et al., 2007). The primary focus of AETS was air defense, as it required complex interaction to coordinate. An air defense team was tasked with protecting the ship from in-air objects such as fixed-wing air craft, rotorcraft, or missiles. A key finding from this study was that ITS concepts were difficult to apply to teams. AETS did not replace the human trainer but instead acted as a tool to decrease workload of the trainer. To do so it utilized multiple feedback components. Instead of real time feedback, it used an AAR. Real time feedback was too overwhelming for the human trainer to provide useful instruction. Also, if feedback had been directly provided for a team member, the task would need to have temporarily been halted, as the team problem solving process would have been delayed by the single member receiving the feedback.

The current research begins to address several gaps. Teams are assessed by computer tutors in both the Recon and TMET projects. Real-time feedback was chosen as the main form of instruction. An AAR will be done in future work.

DEVELOPMENT OF TEAM TUTORS

This work describes two team tasks that serve as platforms for tutoring. Both are discussed briefly and then we note the similarities between them. Also, we will discuss types of team tasks and team training that are not addressed.

Team Multiple Errands Task (TMET)

The purpose of the original Multiple Errand Task (MET) (Shallice and Burgess, 1991) was used to investigate how successfully patients with prefrontal brain injuries were able to complete different cognitive tasks. These patients were given eight separate tasks within the context of real-world shopping, with a mix of challenging and simple tasks.

The MET has been modified on several occasions. The MET Hospital Version (MET-HV) was designed for patients who could not be observed in a public scenario (Knight, Alderman, & Burgess, 2002). The MET was simplified to the MET Simplified Version (MET-SV) to appeal to wider range of participants that may be encountered within a hospital environment (Alderman, Burgess, Knight, & Henman, 2003). The Virtual MET (VMET) was created by altering the original MET-HV for completion in a virtual environment (Rand, Rukan, Weiss, & Katz, 2009). The authors utilized the MET to evaluate distributed teams, creating the TMET (Walton et al., 2015; Walton, 2015). In exploring a team-based MET, we are able to ask research questions based on how different forms of feedback affects team performance and team trust, how the familiarity of team members affect team performance, and a variety of other team constructs. This effort confirmed that the Team Multiple Errands Task (TMET) could serve as a platform for testing teams’ ability to perform under heavy cognitive load.

The TMET used a 3D first-person viewpoint virtual mall environment with teams of three, each with a separate laptop. In an initial study (Walton, 2015), each team was given eight minutes to complete the task, and the purpose was to explore whether performance in the task would be higher when feedback was delivered publically.

In the TMET, each team member was provided with an individual and team shopping list. The individual lists had six unique items which were only found on respective player’s lists. The team list contained items that players must assign to each other for purchasing. Players were to collect the items as quickly as possible, similarly to the individual MET
guidelines. However, they were provided seven rules that constrained their behavior, some typical, such as "Do not exceed your allotted amount of money" and some more unusual, such as "You can only visit a store once."

The initial study did not produce strong conclusions about feedback modalities, perhaps due to participants not paying close attention to the feedback provided during the task. Discussion with participants within the debriefing revealed that some ignored feedback or did not notice it due to attention dedicated to the experimental task. However, the TMET as a platform for testing team training was successful. TMET provides researchers with a means to test levels of cognitive load in participants. This software was developed using the Unity3D program with standard assets, which lend it to further customization. The individual and team shopping lists items can be altered. The time limits can also be adjusted.

Additionally, the success found in the TMET task could assess performance levels of teams and individuals without ceiling or floor effects. Performance did not randomly change; it appeared consistent or to be improving as sessions progressed. These factors suggest the TMET’s ability to serve as an assessment tool for cognitive performance of teams.

Finally, the TMET platform can also be used as a platform to develop team communication.

According to the properties from the team task taxonomy (Bonner et al., 2015), the TMET shopping task can be solved by adhering to the guidelines and the team members executing pooled interdependent tasks of purchasing items from a list. The task is not routine, as shopping is not typically completed with such strict guidelines. For example, in the virtual environment, participants are not able to enter a store more than once. The task itself can be categorized as a service task as it involves social interaction with a group. The team itself is not dependent on the skills of team members. It has a culture of shared responsibility, as roles are generally identical.

Recon Task

The second team tutor was a reconnaissance task created as a testbed for an ITS designed to train teams in military situations. The task was created using the software Virtual Battlespace Simulator 2 (VBS2) (Bohemia Interactive Simulations, 2011), which has been used for training by the United States Army. Participants completed the Recon Task in a virtual environment where each team member was responsible monitoring a particular zone within the environment (see Figure 1) from an individual laptop in an individual room (see Figure 2). One of the two participant’s eye gaze was assessed via an eye tracker in order to determine if the participant was looking at the feedback provided.

Each zone was bordered by one green pole on one side and two green poles on the other. The zones contained walls arranged in an arc. During the task, enemy combatants referred to as OPFOR (Opposing Forces) emerged at randomly generated times from behind randomly selected walls. Participants’ objective was to identify all of the OPFOR observed in their respective zones. The task was set up for two team members but could be increased to more.

Each member of the team shared in the responsibility of identifying all OPFOR. Each member was responsible for scanning the assigned zone and identifying all OPFOR within it by pressing a key on a keyboard. Each team member was also responsible for communicating with the other team member when an OPFOR approached the boundary between the two zones ("transfer") with a different keystroke (and optional speech via a speakerphone). In this way team members shared information regarding the location of OPFOR and ensured that all movement is accounted for.

Over the course of the task, participants received feedback. The feedback was provided through an adaptive tutor created using the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is an open source, domain-independent framework for creating ITSs. While team tutoring is a goal of the GIFT project, initial experimentation and tutors using GIFT have focused on individual tutoring. In the currently described testbed, the process of authoring a team tutor with GIFT was more complex than generating an individual tutor, as considerations needed to be taken for keeping track of multiple individuals as well as team performance. GIFT interfaces with the VBS2 software, and wrote the participant performance data to log files, which were later analyzed. The feedback was manipulated between teams to determine the efficacy of individual and team feedback. One type of feedback was addressed to individuals within a team and focused on individual performance. The other type of feedback addressed the team and used individual performance from both team members to provide feedback.
about overall team performance based on the weaker
performer in the team.

Performance was measured on how well individual
participants performed in a team context. To be able to keep
track of all OPFOR, participants had to carefully time their
communications so as not to distract their teammates or
detract from their abilities to perform the task. Feedback dealt
with specific components of the task such as timely
identification and timely communication. We were interested
in seeing if there is a difference in prioritization among teams
who receive individual feedback and those who receive team
feedback. This paradigm allowed us to investigate the efficacy
of the two types of feedback on team performance. In this
experiment, participants completed the same task over the
course of four trials. We expected to see improvement in all
teams but at different rates depending on feedback type.

Extensive piloting of the task was required to refine the
process, feedback timing, and logistics. Over the course of 12
pilot sessions we altered the task protocol to best
accommodate the way that our participants understood the
task as well as how they preferred to interact with the software
and each other. Future research will report the final results of
particular studies using this testbed.

Our hope for continued development of the Recon Team
ITS is to introduce more team members and create a hierarchy
within teams. Currently all team members share responsibility
equally for all tasks, but often teams include a leadership role
and specific delegation of tasks to different members. With
this increase in complexity would come a commensurate
increase in tutor complexity, as the ITS would be responsible
for assessing individual performance based on specific
parameters for each role as well as incorporating these
measures into a team performance metric.

LESSONS LEARNED

Previous efforts have yielded lessons learned and insights both
on the building of team tutors and the evaluation of team
tutors. Particularly, insight has been provided about the ways
in which the type of team task that is being trained influences
the type of tutor that is built. Also, we have received insight
about the complexity involved with team tutor authoring.

Building and Evaluating a Team Task for Tutoring

The Recon task and TMET tutors are closely related, as
they were concurrently developed. The tasks both contain
elements of time pressure. This is more evident within the
TMET, as participants are required to continually monitor a
ticking clock, which may add to their cognitive load. In the
Recon task, participants understand that their performance is
determined by how fast they react to the situation. However,
they are not consistently presented with a measure of time.

Similarly, both tasks also require communication. In one
version, the Recon task presents participants with a restricted
amount of communication based on scripts directly associated
with actions. In another version of Recon, as in TMET,
participants are able to openly communicate strategies and
reflect on performance as a team. In both tasks, team members
have identical roles and neither task has participants (non-
confederates) with special skills.

Several key lessons were learned throughout the process.
First, while feedback can be presented, it is up to the learner to
acknowledge it and adjust performance accordingly. Next,
there may be a learning curve that affects variables of the
study as participants learn the task across multiple trials,
which increases the complexity of the statistical analysis.
Ideally, participants should complete the learning curve for the
task before beginning the stage of experimentation in which
the impact of feedback is observed. Or, with a very high
sample size, a baseline "standard" learning curve could be
established and used for comparison when innovative
feedback is provided. Additionally, the implementation of
certain team strategies may affect results, as the strategies are
new to the team members and take time to perfect in the team.

Authoring Process for the Recon Scenario

The authoring process for a team tutoring scenario
involves three main components: the tutoring software, the
users, and the player's environment. All these components
must work together to produce a seamless team tutor. To
author a team tutor, domain specific knowledge is required in
each of the component areas.

In the case of the Recon task, GIFT evaluated the user’s
performance using inputs provided by the user and virtual
environment (VE) data, passed from VBS2. GIFT logged both
the VE and user input information. GIFT also contained the
evaluation logic for producing feedback under given
conditions. Authoring using the GIFT component of the team
tutor required understanding of GIFT’s domain knowledge
files (DKFs), programming condition files using Java. The
DKFs contained the logic behind evaluating performance and
providing feedback. The condition files evaluated the actions
of the participants from the VE scenario.

The user portion of tutor authoring dealt with the team
member’s interaction with the system and each other. This
portion of authoring focused on developing a sound user study
using appropriate research methods.

The VBS2 portion of team tutor authoring dealt with
creation of a virtual environment. Creating this VBS2 recon
scenario required knowledge of VBS2, scripting, and the
geometric assets required to construct the scene.

The knowledge required to author this team tutor ranged
from programming to user study development to
understanding of team dynamics. This wide range of required
competency poses a large adoption barrier.

Much progress has been made integrating GIFT with
VBS2 and other game engines. The current focus on GIFT
development has been on improving functionality and
stability, and the version of GIFT used for the work was
expanded to include tutoring for teams.

Team ITSs / Team Tutors Authoring Recommendations

Based on the experience developing the two team tutors,
we offer authoring guidelines in Figure 3 for the development
of Team ITSs. Those familiar with authoring individual ITSs
will recognize the overall structure, but we have noted the
areas that become different in a team task. Much of a team’s performance depends on interpersonal dynamics and whether team members have previous experience with each other, so it is more critical with a team tutor than with an individual tutor to know as much about the learners as possible from the start.

How will task characteristics determine training (see taxonomy)?
Will you tutor both kinds? What do your learners know at the start?
What skill indicators can technology observe? Teammate interactions?
These errors will most likely trigger feedback.
Will feedback actually improve performance? With repetition?
Based on individual or team behavior? Give to individual or team?
With teams, can be a combinatoric issue.
Feedback too long? Too frequent? Too distracting?

Figure 3: Authoring Guidelines

SUMMARY

Team ITSs are an emerging and potentially powerful tool for training. However, developing an effective team tutor is not an easy task. Previous implementations such as AETS addressed the difficulty of providing feedback to teams but did not successfully implement automated team training.

Through the creation of two team tutors we have identified the requirements necessary to create effective team tutors and documented lessons learned for authoring them. The main contribution is the authoring guidelines, as they serve as a resource for ITS and team feedback researchers to use for the development of their own team tutors.

ACKNOWLEDGEMENTS

The work was completed with support from a U.S. Army Research Laboratory- Human Research & Engineering Directorate (ARL-HRED). Statements and opinions expressed in this text do not necessarily reflect the position or the policy of the United Stated Government, and no official endorsement should be inferred. Also, this work would not be possible without contributions from Joan H. Johnston.

REFERENCES