Iowa State University
Digital Repository

Industrial and Manufacturing Systems Engineering Publications

Industrial and Manufacturing Systems Engineering

2018

Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach

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

Michael Dorneich
Iowa State University, dorneich@iastate.edu

Jamiahus Walton
Iowa State University, jwalton@iastate.edu

Eliot Winer
Iowa State University, ewiner@iastate.edu

Follow this and additional works at: https://lib.dr.iastate.edu/imse_pubs

Part of the Ergonomics Commons, Graphics and Human Computer Interfaces Commons, Operational Research Commons, and the Systems Engineering Commons

The complete bibliographic information for this item can be found at https://lib.dr.iastate.edu/imse_pubs/193. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Book Chapter 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 Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach

Abstract

This chapter describes five disciplinary domains of research or lenses that contribute to the design of a team tutor. We focus on four significant challenges in developing Intelligent Team Tutoring Systems (ITTSs), and explore how the five lenses can offer guidance for these challenges. The four challenges arise in the design of team member interactions, performance metrics and skill development, feedback, and tutor authoring. The five lenses or research domains that we apply to these four challenges are Tutor Engineering, Learning Sciences, Science of Teams, Data Analyst, and Human–Computer Interaction. This matrix of applications from each perspective offers a framework to guide designers in creating ITTSs.

Keywords

Intelligent team tutoring system, intelligent tutoring system, training, teamwork, task work, interdisciplinary research

Disciplines

Ergonomics | Graphics and Human Computer Interfaces | Operational Research | Systems Engineering

Comments

This is a manuscript of a chapter published as Gilbert, Stephen B., Michael C. Dorneich, Jamiahus Walton, and Eliot Winer. "Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach." In: *Building Intelligent Tutoring Systems for Teams: What Matters (*Joan Johnston, Robert Sottilare, Anne M. Sinatra, and C. Shawn Burke, eds.). Research on Managing Groups and Teams, Volume 19. Bingley: Emerald Publishing Limited (2018): 247-277. DOI: 10.1108/S1534-085620180000019014. Posted with permission.

This book chapter is available at Iowa State University Digital Repository: https://lib.dr.iastate.edu/imse_pubs/193

Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach

Stephen B. Gilbert, Ph.D, Michael Dorneich, Ph.D., Jamiahus Walton, M.S., Eliot Winer, Ph.D.

Iowa State University, Ames, IA

Correspondence concerning this article should be addressed to Stephen Gilbert, Iowa State University, 1620 Howe Hall, 537 Bissell Rd, Ames, IA 50011-2274. gilbert@iastate.edu

Abstract

This chapter describes five disciplinary domains of research or lenses that contribute to the design of a team tutor. We focus on four significant challenges in developing Intelligent Team Tutoring Systems (ITTSs), and explore how the five lenses can offer guidance for these challenges. The four challenges arise in the design of team member interactions, performance metrics and skill development, feedback, and tutor authoring. The five lenses, or research domains that we apply to these four challenges are: Tutor Engineering, Learning Sciences, Science of Teams, Data Analyst, and Human-Computer Interaction. This matrix of applications from each perspective offers a framework to guide designers in creating ITTSs.

Keywords: intelligent team tutoring system, intelligent tutoring system, training, teamwork, task work, interdisciplinary research

INTRODUCTION

Intelligent Tutoring Systems (ITSs) have been created for a variety of domains such as the military (Zachary et al., 1999), intelligent computer-assisted language learning (Gamper & Knapp, 2002) and education (Bradáč & Kostolányová, 2016). While they have been successful in providing personalized individual instruction in a variety of domains (e.g., programming, algebra, physics, and on-the-job training) (Anderson, 1989; Arroyo-Figueroa, Hernandez, & Sucar, 2006; Koedinger, 1997; VanLehn, van de Sande, Shelby, & Gershman, 2010), the complexity and difficulty of building an ITS has been well-documented (Murray, Blessing, & Ainsworth, 2003). Given the combinatorics of team member interactions and the need for a team tutor to track those interactions it is expected that building an intelligent team tutoring system (ITTS) is more complex. In addition, a team tutor must decide whether to give feedback to specific individuals or the entire team. Our goal in this chapter is to provide lessons learned from our experience in developing three ITTSs to give future developers an understanding of the challenges they will need to resolve. We briefly describe the ITTSs we developed and then introduce the four main challenges we encountered; more details can be found elsewhere in Gilbert et al. (2017). We then introduce and discuss how an interdisciplinary perspective, through the lenses of five different research domains, can enable developers to address each challenge.

THREE INTELLIGENT TEAM TUTORING SYSTEMS

Over the past several years, the authors and their colleagues developed and conducted research with the Surveillance Task (two-person team, both same role), the Surveillance with Sniper Task (three-person team, one sniper and two spotter roles), and the Team Multiple Errands Task (three-person team, all the same role).

The Surveillance Task is described in more detail in Bonner et al. (2016; 2017) and Gilbert et al. (2017). Each of two team members play the role of a spotter atop a building in the center of a small town and must report to each other when people on the ground below move out of one member's zone into the other member's zone, e.g., "Two people entering your zone at the pole." The receiving team member must acknowledge that communication and then note when he or she sees them enter the zone, e.g., "Acknowledged…. Ok, I see two people." Each team member is sitting at his or her own laptop in a room with an open audio communication channel shared with the other teammate. When they speak to communicate they must also press a key corresponding to what they are saying, i.e., typing 1 to indicate a crossing near the pole, E to acknowledge, and Spacebar to indicate seeing someone enter the zone. Team members receive textual on-screen feedback from the tutor if they miss someone crossing, do not communicate well with the teammate, etc. This ITTS was developed with the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Goldberg, & Holden, 2012) for the tutoring engine and Virtual Battlespace 2 (Bohemia Interaction Simulations, 2011) for the scenario.

The Surveillance with Sniper Task (SwS) extended the Surveillance Task using GIFT and VBS2 and is described in more detail in Bonner et al. (2017). Both spotters from the Surveillance Task interact with a sniper who sits in a high tower with a view of the entire town and has high-powered binoculars. The spotter's job is to alert the sniper of potential threats running on the ground and the sniper identifies whether they are either a civilian, or a threat level 1, or a threat level 2. The team members communicate on an open audio channel and signal their communications to the tutor with keystrokes as in the Surveillance Task. Each member receives feedback pertaining to their specific task roles, as well as for the more general requirement of good communication. Depending on the feedback mode set in the ITTS, all team members may receive the same feedback, or it may be personalized to individuals.

The Team Multiple Errands Task (TMET) is described in more detail in Walton et al. (2015) and Walton, Gilbert, Winer, Dorneich, and Bonner (2015). The original Multiple Errands Task (MET) (Shallice & Burgess, 1991) has been used to identify cognitive deficits in patients and consists essentially of a shopping trip with a time limit and a several additional constraints such as, "You may purchase only one thing from a store." The TMET was adapted from the MET and is a three-person team of friends purchasing supplies for a mutual friend's surprise party. Each member is given an individual shopping list and there is a team shopping list with items unassigned to specific team members which they must divide up. They enter a virtual shopping mall using their respective laptops and attempt to complete the lists as quickly as possible while staying within budget. Team members communicate via an open audio channel. They receive textual on-screen feedback from the tutor such as, "Remember that you cannot buy more than one thing per store" and "You only have 3 minutes left." This ITTS was developed with the Unity scenario and tutoring engine.

CHALLENGES

The primary challenges to our ITTS development arose from team member interactions, performance metrics and skill development, feedback design, and tutor authoring. They are similar to individual ITSs challenges, however, specific nuances related to training teams make addressing these challenges crucial to the success of an ITTS.

The team member interactions challenge stems from the task of tracking and identifying team member actions and interactions with other team members. If team interactions occur via an intermediary user interface, for example, between distributed team members (e.g., computer-supported collaborative work), they can be more easily tracked than if all members are in the same room and communicating with each other via voice. In addition, this challenge includes the dynamics that arise from different team member roles and the interdependence of team member tasks. For example, the team might have a captain and multiple followers (e.g., in a sports team or military patrol), or the team may be a collection of individuals with specific skills (e.g., an emergency room response team). Furthermore, the tracking granularity and choosing what to track can be a challenge, such as determining which trackable behaviors will map to performance constructs that must be measured, and, if one's goal is to assess whether one team is better than another (see overlap with the performance metrics challenge) and a team’s performance will be measured in successful communications, determining the required granularity of team member communications that must be tracked.

The performance metrics and skill development challenge relates to the evaluation of teams and team members while performing their team task. Most team performance outcome measures are for task skills such as time to completion or error rate. However, most trainers need to improve their team’s ability to work as a team (i.e., teamwork) which involves communicating well, understanding each other's roles, and performing smoothly. The metrics used to evaluate teamwork can vary according to the team task context. One of the most difficult aspects of this challenge is mapping the behaviors that one can measure during the team activity to the desired skills and performance metrics. For example, if the team completes a training scenario six times are the members sufficiently trained in the corresponding skills or learning outcomes? Another challenge is that team studies typically do not yield independent data as is required by the more standard statistical approaches such as ANOVA. For example, on a team of Alice and Bob, Alice's performance is not independent of Bob's, and their team's performance on Trial 4 is not independent from their performance on Trial 1. These two types of dependence, across time and across team members, typically add complexity to statistical analysis, leading to mixed designs.

The feedback design challenge involves deciding what the tutor should tell them, when it should tell them, how it should communicate within the software, and when to give feedback to individuals versus the whole team. This step is what Van Lehn (2006) calls the inner loop of an ITS which is observing what learners are doing on individual steps of a larger task and giving them appropriate feedback to be successful. Feedback might appear frequently in the form of textual messages that appear on the screen, or perhaps infrequently in the form of vibrations on one's mobile device. Ideally, the tutor knows something about each student and can personalize the frequency and tone of the feedback for each learner. A variety of researchers have analyzed the best content for feedback, e.g., how long to wait until revealing the answer, whether to be strategic or tactical, and the impact of different affective approaches (McKendree, 1990; Roll, Aleven, McLaren, & Koedinger, 2011; Shute, 2008; Yang & Dorneich, 2016). However, there has been little research on how to decide when intelligent team tutors should offer feedback to individuals vs. the entire team, though this issue has been researched to some extent in the context of sports teams (DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Stokes, Luiselli, Reed, & Fleming, 2010).

The tutor authoring challenge in designing a team tutor is determining how to implement the vision that the tutor author has in his or her head. Does the implementation require programming, or perhaps the completion of a series of on-screen dialog boxes with a wizard-based system that asks questions of the author? Some tutor authoring systems such as the Cognitive Tutor Authoring Tools (CTAT) (Koedinger, Aleven, Heffernan, McLaren, & Hockenberry, 2004) allow authoring by demonstration; the expert simply demonstrates the correct procedure for accomplishing the task. However, even in this easy-to-use system, the process becomes more complicated when the expert wants to "teach" the tutor what feedback to give when the student makes mistakes, and how to recognize those mistakes. This process usually requires conditional logic, e.g., If the student does X, say Y. This logic can be difficult for non-programmers to create. Also, debugging or testing can be difficult and the tutor author may wonder: Have I tested all the possible cases that a student might encounter?

With a team tutor, these issues become further complicated by the interactions between multiple team members who may have specific roles. A conditional rule may be, for example, if one is learning to be a surgeon: “If someone hands you something, acknowledge by saying its name." That use of "someone" in the rule is essentially a variable and opens the possibility that the rule may be invoked if any of the learners on the team perform that action. Also, conditionals may depend on multiple team members, e.g., "If a team member does X and two or more other team members do Y, then say Z." This type of condition requires Boolean logic that spans team members not specified directly, i.e., the second part of the predicate requires that any number of team members greater than two perform an action, not counting the team member noted in the first clause the predicate. The use of constructions such as "someone," "two or more," and "other" make team conditions significantly more complicated to author and to debug and test.

THE FIVE LENSES: RESEARCH DOMAINS CRITICAL TO ITTS DESIGN

To address these challenges we adopted a multidisciplinary perspective based on the following five research domains: Tutor Engineering, Learning Sciences, Science of teams, Data Analyst, and Human-Computer Interaction.

Tutor Engineering

Though there are ITSs for various domains each has the same four basic components: some representation of the task, student or learner representation (student model), teaching instructions (feedback) that follow the completion of one or two components, and a domain model (Hartley & Sleeman, 1973). Engineering challenges arise when building an ITTS designed to instruct distributed or co-located teams when considering these four components (Sottilare, Holden, Brawner, & Goldberg, 2011).

The inherent distributed nature of teams creates design and engineering challenges specific to ITTSs. The goal is to create an experience for the team receiving the tutoring instruction that will improve the team’s performance. Sottilare et al. (2017) conducted a meta-analysis exploring how to effectively design adaptive instructions for teams. According to the results, the antecedents for team performance are collective efficacy, cohesion, communication, and leadership. The antecedents for team learning are trust, cohesion, conflict, and conflict management. This conclusion suggests that the ITTS must have the ability to capture, interpret, and respond to these behavioral markers. This means that the ITTS needs to effectively communicate the knowledge and information gathered about a team. A human tutor has the ability to communicate their knowledge about the error committed by the student while allowing them to self-correct the error committed (Merrill, 1992). The logical conclusion would be to design an ITTS to communicate in a way that is like a human tutor. According to Shute and Psotka (1994), however, the goal of an ITS is to communicate knowledge effectively, not to communicate knowledge identical to human instructors. This conclusion suggests that an innovative ITTS designer could consider instructional methods for teams that go beyond the capabilities of a human team trainer.

Learning Sciences

Learning Sciences (LS) is an interdisciplinary research area that is based in psychology, cognitive sciences, and education (Kolodner, 2004; Sawyer, 2014). A branch of the LS that can provide insight into how teams learn is Computer-Supported Collaborative Learning (CSCL), which focuses on how people collaboratively learn with the help of technology (Stahl, Koschmann, & Suthers, 2006). Also, there has been a growing interest in a teaching method that has been around since the 1980s known as Team-Based Learning (TBL) (Michaelsen, Knight, & Fink, 2002). Team-Based Learning is different from the traditional lecture style class and requires the students to be more active, working through assignments in teams during class. Existing LS research on TBL could provide valuable insight into not only how individuals work as a team, but also how they can learn as a team. The core of TBL focuses on team formation, accountability, feedback, and applications that promote learning and team development (Ofstad & Brunner, 2013). Learning more about these core principles has the potential to provide critical insight on how to develop an ITTS that effectively promotes learning within a team.

Science of Teams

This lens represents the large body of research on how team members interact, particularly from the perspective of team cognition (Salas & Fiore, 2004), and team mental models (Cannon-Bowers & Salas, 2001). It is worth noting that this research area is separate from the study of how scientific teams conduct research, which is often designated "team science" (Fiore, 2008).

A component that hinders the development of ITTSs is the team representation, since team structures can vary widely (Bonner et al., 2015). Studies have sought to determine which team characteristics, among many, are accurate indicators of effective teams. Campion, Medsker, and Higgs (1993) argue that job design, interdependence, composition, context, and process are related to the effectiveness of a team and are generalizable (Campion, Papper, & Medsker, 1996). Mickan and Rodger (2000) determined that the themes for characteristics tied to effective team are Organizational structure, individual contribution, and team processes. Bannister, Wickenheiser, and Keegan (2014) claim purpose, roles and skills, and openness are key elements to effective teams. Salas, Sims, and Burke (2005) concluded that important aspects of teamwork are team leadership, mutual performance modeling, backup behavior, adaptability, and team orientation. Effective teams also have the ability to undergo phenomena such as team adaptation (Burke, Stagl, Salas, Pierce, & Kendall, 2006). Team theories and research provide valuable insight into how teams function and indicate an ITTS must have the ability to evaluate these characteristics to determine team effectiveness.

Data Analysis

The data analyst wants to know what happened during the tutoring, whether the tutoring led to conceptual change and improved skills, which tasks posed the greatest challenge, and what the most effective feedback messages were. This lens draws on a range of backgrounds, from traditional psychological research paradigms to Educational Data Mining (EDM) and Learning Analytics (LA), which are areas of data mining focused on automated methods to analyze educational data (Winne & Baker, 2013). The insights gained from these automated methods have the potential to improve instructions delivered by a tutoring system. For example, an ITS called Pinyin (Kowalski, Zhang, & Gordon, 2014) was built to help teach students how to write spoken Chinese phrases in Pinyin. This tutor utilized large amounts of data collected on the different types of errors students made when attempting this task. The authors were able to implement machine learning techniques to discover the most difficult part of a task and develop a model for student performance. A similar approach allowed Carnegie Learning, Inc. to significantly reduce the number of parameters required to model a student (Ritter et al., 2009). These machine learning techniques are critical to developing an effective ITTS because it will generate more than twice the data of an ITS.

The traditional statistical research lens that assesses risks of bias in experimental design and the likelihood that results are by chance is critical to extracting information about teams that will enable an ITTS to be a more effective. The challenge for the data analysis lens in this context is always "What can be measured?" and "What can be concluded from those data?"

It is difficult to measure team performance (Burke et al., 2006) because teamwork has multiple levels and require different measurements to study it (Salas, Rosen, Burke, Nicholson, & Howse, 2007). The next logical step is to determine which metrics are relevant to the area of focus. One challenge in determining relevant metrics to measure team performance is that teams change over time. One could assume that if a team changes over time, then the ITTS instructing a team must have the ability to adapt over time as well (see Johnston et al., this volume).

Human-Computer Interaction

Human-Computer Interaction (HCI) is an interdisciplinary field that continues to evolve with the goal of encouraging interactions that allow both humans and computers to perform tasks efficiently. Long and Dowell (1989) identified craft discipline, applied scientific discipline, and engineering discipline as three conceptualizations of HCI. The preferred HCI concept would depend on how efficiency is approached. However, Carroll (2010) argued that the scope presented by Long and Dowell is too narrow and did not address non-work-related tasks. Instead, Carroll noted that HCI should be viewed as a “meta-discipline” with a community centered on the idea of usability and user-centered design.

One aspect of user-centered design is the user's cognitive workload. With ITTSs, the question arises as to whether the additional workload of communicating with a tutoring agent can be valuable enough to offset the workload of performing the team task without the agent. Van der Meij (2013), for example, found that developing an agent focused on motivating students to generate principle-based explanations resulted in higher performance scores during training.

ADDRESSING THE CHALLENGES WITH THE FIVE LENSES

In this section we explore how the scientific and practical evidence from the five disciplinary lenses can be used to address the four challenges described earlier in the chapter.

Team Member Interactions (TMI)

Tutor Engineering (TMI)

To address team member interactions, this lens focuses on what individuals' actions and multi-person interactions need to be tracked, and how to track them. The granularity of learner actions of interest can vary widely. In a Cognitive Tutor by Carnegie Learning, Inc., for example, almost every user action can be used to help assess what a learner is thinking and what the user might click next. At the other end of this spectrum is the TMET tutor where learners take many actions moving around the shopping mall that are not tracked because the designer only wanted to know which stores they visited.

There is also the issue of what actions can realistically be tracked. Team communication can be quite difficult to parse computationally especially if people talk over one another. Sociometric badges have been used to measure communication dynamics based on utterance frequency and duration without actually recording team member speech (Pentland, 2012). Also, body language is a challenge to track. For example, an ITTS designer might want to give feedback to team members who do not look each other in the eye, but unless each member is wearing eye-tracking glasses collecting that behavior is difficult.

Lastly, a key element of tracking is whether patterns formed over time are important to track within a particular team task. If the designer wishes to give feedback such as, "You are improving rapidly on communicating with your teammates," then data must be logged over time, and that pattern identified. The software architecture must support such pattern recognition over a history of actions among multiple team members.

Learning Sciences (TMI)

The Learning Sciences have several theoretical frameworks to address the challenge of team member interactions. Team-Based Learning (TBL) is a pedagogy that focuses on assigning a problem of interest to multiple small teams in parallel within a classroom (Michaelsen et al., 2002). Teams are given significant problems that cannot be solved individually, and feedback is given frequently and just-in-time. Regarding team-member interactions, TBL emphasizes the value of real-time peer feedback with the team, between the team and the instructor, and between teams. TBL tries to create situations where students are teaching students when addressing the challenges of group quizzes and activities. A good team tutor learning scenario should try to provide real-time peer collaboration as an affordance. Peer instruction is a related theory that suggests the value of placing team members in settings in which they can learn from each other (Crouch & Mazur, 2001).

Computer Supported Collaborative Learning (CSCL) and collaborative learning (Roschelle, Suthers, & Grover, 2014) focus on enabling the social interactions and group cognition that are required to work well together. In this view, the ideal team tutor could monitor how learners find common ground, how they move from peripheral members of the team's community to core members, and how the team becomes a community of practice by establishing its own team norms and terminology related to the learning and tasks at hand. Tracking team member motivations, level of engagement, and sense of accountability to others are all important factors in collaborative learning, though are logistically difficult to track.

Science of Teams (TMI)

The team behaviors that one wants to track vary widely as noted in the Bonner et al. (2015) taxonomy of teams and team tasks. Team members may know each other or not; members may have different roles or the same role; and these are just a couple of the characteristics that may vary among teams. The structure of team tasks also varies significantly in the interdependency of team member subtasks. For example, an assembly line worker early in the line heavily influences the work of others downstream, in contrast, farm workers picking fruit in parallel hardly overlap.

As noted in the section introduced earlier in this chapter, researchers such as Salas et al. (2015) found numerous constructs that affect team performance such as trust, cohesion, communication, and conflict management. Sottilare et al. (2017) recommended mapping specific team behaviors (markers) to these constructs which will be help to guide developers in which behaviors to track.

Data Analysis (TMI)

There are essentially three forms of data to analyze. First are the real-time lower granular behavioral data that needs to be analyzed in order to provide tutoring, labeled the "inner loop" per VanLehn (2006). Next is performance data collected over time that enables the tutor to choose appropriate tasks to continue to challenge the learners or VanLehn's "outer loop," also described at times as "stealth assessment" (Shute, 2011). Finally, there is the behavioral data about the entire team tutoring experience that can be collected and analyzed across multiple teams by a researcher or educational data analytics expert to evaluate the longer-term process of learning, the learning effectiveness of tasks, or usability challenges of the tutoring software. At each of these three levels it is often difficult to choose what specific observable actions to record and analyze for a team. As noted above, Sottilaire et al. (2017) have attempted to map specific team member behavioral markers onto team constructs such as communication and trust. Another approach labeled the Human Performance Markup Language was developed by Stacy and Freeman (2016). Questions of particular interest that arise with team tutoring but not with individual tutoring include: "Does this team interaction fit a recognizable pattern or profile?" and "How do the team members contribute to the performance of the team?" Related to the first question, with an ITTS it is likely that an additional "team loop" could be considered, along with VanLehn's inner and outer loops, in which behaviors by the group of learners are compared with a model of team interactions to identify the particular strategies a given team is using.

Human-Computer Interaction (TMI)

Computer-Supported Cooperative Work (CSCW) is an interdisciplinary field that studies how people work within groups and how technology can be used to support team member coordination and collaboration (Grudin, 1994). CSCW refers to activities facilitated by computers in which multiple people are working toward a common goal. The collaboration occurs in two dimensions: time (synchronous, asynchronous) and location (co-located and distributed). CSCW tools aim to improve or manage coordination of activities and individuals (Bannon & Schmidt, 1989), social presence or degree of visibility of participants in the collaboration (Rice, 1993), media richness or the amount of information available (Wernick, 1998), information sharing that identifies what information is known to whom (McGrath, 1984), and shared state information reflected in the current state of system data (Farley, 1998).

Group interactions are typically in support of creation of a work product (artifact). Team members develop a shared understanding of the work through two modes of communication: direct person to person communication, or communication through changes to the artifact. This gives rise to three types of CSCW tools: meetings and decision support systems to support the development of shared understanding, computer-mediated communication tools to support the direct communication between participants, and shared applications and artifacts to control and make visible changes to shared work objects (Dix, Finaly, Abowd, & Beale, 2001). When users make changes to an artifact in a shared application, feedback from the application lets users see the effects of their changes. Feed-through takes place when the collaborators can also see that changes, and enables communication between team members through the artifact. Application of CSCW principles to ITTSs would require that team tutors be open to all aspects of collaboration, between a group of learners, between the tutor and the learners, and through any shared artifacts. The team tutor system does not need to automate every aspect of the communication and shared work, but should support the cooperative work as a whole.

Performance Metrics and Skill Development (PMSD)

Tutor Engineering (PMSD)

An ITTS should be able to collect all four forms of data described above: inner loop data, team loop data, outer loop data, and across-teams research analytics data. The two main challenges in this process are acquiring the data, and mapping it to desired outcome variables with appropriate filters. The logistics of acquiring the data can be a significant challenge. If team member voice communications are important to track, can each member be miked? As the audio signals arrive to the tutor, can it separate utterances and parse them? If a variety of external data sources are being integrated within the ITTS (e.g., audio or physiological signals), then time synchronization is an important issue to resolve. Each data source has its own time stamps, typically, and the tutor needs to know the time point at which all sources began logging simultaneously. Essentially, the ITTS needs an analogue of the clapperboard used in movie making to synchronize the picture and sound.

Because tracking team interactions is so difficult many team training scenarios take place in a software environment such as a game engine so that all interactions can be more easily tracked. To track learner actions and world state in a game engine, the ITTS will need an API to communicate with the game engine. Using this API can pose a vocabulary or granularity mapping challenge. For example, if the author decides to encode a condition for feedback such as, "If the learner enters Building 4, then…", but the game engine does not have an event triggered by building entry, and instead just offers the map location of each player, then interstitial code needs to be written that polls whether player locations are inside buildings and sends appropriate signals to the tutor. This challenge also arises in traditional ITSs (Gilbert, Blessing, & Blankenship, 2009; Ritter & Koedinger, 1996), but with ITTSs team member interactions likely need to be tracked, and the game engine probably will not have specific logged events for those, requiring further custom software development. In the opposite situation the game engine may log many variables about the state of the scenario that are not of interest. Filtering out unnecessary data may require software development as well.

The other engineering challenge in processing performance metrics is converting the observable to data to metrics that are useful for tutor evaluation and reporting to other stakeholders. MacAllister et al. (2017) describe a "Metric Manager" that was used with the Surveillance Task to create metrics from logged data based on the needs of the research team. Ideally a user interface for authoring metrics would be present, allowing the data analyst or tutor author to easily create metrics such as "PercentOnTime = # of prompt arrivals / total arrivals." An advanced ITTS authoring system in the future may have a machine learning pattern-recognition system that automatically finds patterns in the data to report to the researcher or trainer.

Learning Sciences (PMSD)

The learning scientist will want to know how the learners' actions in the team scenario can be used to measure learning. This is a classic problem in simulation-based training or game-based learning: how many times does a learner need to play to meet the learning objectives? And how can we know that meeting the learning objectives in the simulation or game will mean appropriate transfer to the real world? Valerie Shute and colleagues describe stealth assessment as a method of embedding assessments into the simulation, and gauging learning according the choices the player makes in the environment (Schwartz & Arena, 2013; Shute, 2011; Shute, Ventura, Bauer, & Zapata-Rivera, 2009).

With teams and other group work, there is the added question of whether all team members have met the learning goals, or just some of the members, pulling the others over the finish line with them. Collaborative learning theory, for example, suggests that it is important to measure not only task skills and team skills, but also each team member's level of membership in a specific team: to what extent was each member a core member of the team community, vs. a peripheral member? It could be that a team member has relatively high team skills, but the dynamics of a particular team and team task left the member as a relative outsider to the team community. More generally, collaborative learning theory suggests that the ITTS use metrics that measure team members' motivation for working together, level of mutual engagement and joint attention, and degree of self-reflection on the teamwork itself (Roschelle et al., 2014).

Science of Teams (PMSD)

The performance metrics of interest depend on the focus of the training facilitated by the ITTS. For example, Swain and Mills (2003) focused on the difference between expert teams, where members have worked together extensively or recently, and novice teams, where members have never previously worked together. The authors found that expert teams communicate implicitly more than novice teams by using observational measures and a self-report questionnaire.

The subject of individual differences could benefit from further exploration into how they affect team efficiency. Though some researchers argue that individual differences do not have a significant influence on functional structures (Hollenbeck et al., 2002), other suggest they influence team efficiency. For example, Neuman and Wright (1999) argue that personality traits of individuals should be considered when making member selection for a team, and Dorn and Dustdar (2010) suggest personality traits can lead to more optimized teams.

Team performance is a multilayer construct that needs many different metrics (Salas et al., 2007) and it is more appropriate to identify the core process that are critical to team performance. For example, Salas et al. (2015) identified six core processes that are important to consider when studying teamwork: cooperation, conflict, coordination, communication, coaching, and cognition. They also described composition, culture, and context as three "influencing conditions" of the core processes, therefore, metrics should be selected that will give insight into both processes and conditions.

Data Analysis (PMSD)

Statistical analysis approaches for team performance metrics is a significant challenge because classic tests such as analysis of variance that depend on independent and identically distributed samples will not be effective. In teams, the performance of each individual is dependent on other team members, and analyzing the same team's performance over time, across multiple trials, leads to dependence between trials. The data analyst must therefore use other analytical approaches such as repeated measures and mixed experimental designs (e.g., a between-subjects factor of team-oriented vs. non-team-oriented feedback combined with a within-subjects factor of trial number if the team repeats the task over time). Have the statistical power to test results with these approaches requires larger sample sizes, and sample sizes in team studies can be difficult to achieve because of the logistics of recruiting multiple simultaneous participants and the higher cost per trial if participants are paid.

There are several approaches to measuring the skills of individual members and the team. Cognitive Tutors developed by Carnegie Learning, Inc. have typically used a Bayesian approach to modeling learner skills, with priors based on typical experience, as well as the probability that a given successful action is a lucky guess and that a given incorrect action is just a slip rather than an indicator of conceptual misunderstanding (Corbett, 1992). More recently, Microsoft developed the TrueSkill™ measure, a Bayesian approach to estimating the skill of video game players for matchmaking appropriate players (Herbrich, Minka, & Graepel, 2007). Others have attempted to generalize the TrueSkill™ algorithm to teams of variable size (Nikolenko & Sirotkin, 2010). In addition, since there has been a long history of research on what makes an effective team, a variety of researchers having measures of team skills (e.g., Loughry, Ohland, & Woehr, 2014; Neuman & Wright, 1999; Salas, Burke, Fowlkes, & Priest, 2004).

Another goal for the data analyst is the creation of a data dashboard to support several purposes. If multiple teams are learning from the ITTS, the analyst, trainer, researcher, or supervisor will want to understand how learners are progressing. Choosing appropriate measures to display, and with appropriate update frequency, can be challenging. Simply plotting proportions of communication utterances, for example, can be helpful in determining who is speaking most, but those numbers do not reveal whether a high-communicating person is dominating the conversation negatively, or facilitating positively, asking each member for their opinions. Similarly, if one of the goals of the dashboard is to show a team's progress through a scenario, it is important to show data at an appropriate level of granularity. The analyst may not care, for example, that the team members have taken 412 steps through the virtual environment of the scenario, but would care that to see how long it took the team to reach each of six milestones in problem solving.

Human-Computer Interaction (PMSD)

From the HCI lens, the two critical issues with designing performance metrics are the level of granularity (if any) the tutor uses to tell learners what it knows about their state, and whether the performance metrics accurately measure the skills of the learner (are valid). The granularity issue results from the basic question of whether to place a score indicator on the screen of the ITTS user interface while learning. If the answer is yes, then the question arises of which score: just a team score, just an individual score, or perhaps a team score and all team members' scores. This question may be answered in part by the type of team task at hand; some do not lend themselves to scores, but might instead use milestone achievements to indicate progress. Or, as in Carnegie Learning Inc.'s Cognitive Tutors, a learner's progress might be indicated by a "skillometer," a set of achievement indicators on each of the learning objectives of the current task. These achievement indicators play a related role to feedback, in that both feedback and progress indicators help the learner answer the question, "How am I doing?" Malacria et al. (2013) propose a more general framework for understanding skillometers as any form of reflective interface that helps users reflect on their own behavior and improve their performance with a software tool.

In a team setting, being aware of one's role and current cognitive load means that that team can be rated highly on the Team Cognition construct (one of Salas et al.'s nine C's (2015), as well as on situation awareness (Endsley, 1995). For example, Dorneich et al. (2017) demonstrated that tool to help pilots and co-pilots understand each other's cognitive load enabled them to offer each other backup behavior and achieve their goals more efficiently. Thus, some appropriate user interface is likely needed in the ITTS to aid team members in knowing other members' activities and cognitive state.

The second issue raised by the HCI perspective for performance metrics is whether they truly measure what is desired. In a team simulation of a surgical appendectomy, for example, a score based solely on how quickly the surgery is completed may reflect an important factor in real surgeries, and it would be easy to measure, but the team's performance rating is not likely complete without a measure of patient recovery, which cannot be assessed until after the surgery. Perhaps a system that estimated the probability of healthy recovery and lack of complications could be included to give a real-time metric for patient safety. In another example, in the analysis of the Surveillance Task an individual performance metric was defined as a weighted sum of performance scores from each of three subtasks, less the number of errors. An overall team performance metric was chosen to be a weighted sum of the average individual performances of all players as well as several additional measures of performance on specific team subtasks. These formulas raise the question, however, of how the weights were set on the two weighted sums. In this analysis, the weights were chosen based on the task design, but they were not statistically validated with the performances of the learners in the Surveillance Task. Ideally, the teams which most efficiently meet the specified goals of the task should score highest on a team performance metric. Also, it should be the case that the team performance score is a function of the individual performance scores and an additional team skills factor (the team is more than the sum of its parts). A validation process would ensure that these requirements hold. Thus, just as a validated test accurately predicts performance in the skills it is intended to measure, a robust ITTS should have team and individual performance metrics that are validated.

Feedback Design (FD)

Tutor Engineering (FD)

In traditional ITSs, feedback has been tied to If…Then conditions that trigger the feedback statement if certain conditions in the system are true, e.g., a student has just taken an action which the system recognizes as the result of a misconception. When these conditions are grouped together they are often deemed an "expert model" because the conditions are usually based on the wisdom of a subject matter expert. When developing an ITTS this process becomes more complex in several ways. First, multiple expert models will be required if team members have more than one role, e.g., in an emergency room scenario there will be an expert model for the nurse, the doctor, and the anesthesiologist. In addition, a separate model should monitor the team as a whole that gives advice on how anesthesiologist-nurse-doctor teams should work well together. Feedback conditions on working with other job roles could possibly be embedded in each of the individual job role models, but they also might be separate depending on how many other ITTSs need to be built using that same mix of team roles. Lastly, there would likely be an expert model on basic team skills that apply to this team but also other teams with other roles, focusing on communication, trust, and etc. Thus, with this example, an individual ITS has one expert model, but a three-person team ITTS has possibly five expert models.

Next, consider the case of multiple team members that have the same role, e.g., customer service agents in a call center. Theoretically, this ITTS needs just the one task expert model for a customer service agent. This approach might work if the tutor is context free, that is, it has no memory of previous states and simply issues feedback based on whatever the current state is at the current time. For example, Cognitive Tutors do not strongly consider the history of the learner's actions (Blessing, Gilbert, Ourada, & Ritter, 2009). However, given that team performance often depends on interpersonal dynamics it is useful to have an additional layer of abstraction that can filter feedback based on previously provided feedback, feedback frequency, and learner profile (some people might like frequent reminders, for example, while others might find them irritating). In this way, the expert model for customer service agent can serve all learners, but do so via this personalized layer that tracks prior feedback transactions. This approach also assumes that there is a method in the tutor software architecture for distinguishing similar learner actions that arrive at the expert model (X_1 came from Alice and X_2 came from Bob…) and for routing potentially identical feedback messages back to the appropriate learner ("Alice, please remember to…"). Lastly, the expert models need to coordinate with each other. For example, the conditions in one model may reference the state of the other model, e.g., "If the nurse has been given Feedback A, and the state of the scenario is X, then give the doctor Feedback B," or "If more than half of the customer service learners have received Feedback P in the past 15 minutes, then…"This coordination requires a centralized module to regulate the overall amount of feedback that any one learner is receiving.

Learning Sciences (FD)

Learning Sciences research recommends several important factors with respect to feedback that improve learning and performance in teams. First is how feedback is provided to teams.

Feedback may be directed to individuals or to an entire team. Smith (1972) concluded that individual reinforcement feedback produced more satisfaction with the task than group reinforcement feedback. An ITTS should have a feature that gives feedback to the individual members of teams and to the team as a whole. The second factor is timing of the feedback. In some training contexts, individual and group feedback is given at the end of a task with an after action review so that trainees' focus on the task is not interrupted. However, an ITTS can provide real-time feedback if it can be done without affecting team performance.

The third factor is feedback modality. Research suggests that visual and vocal feedback benefits performance the most (Smith & Ward, 2006; Stokes et al., 2010). Research reviews of instructional feedback in the learning sciences field (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Mory, 2004) suggest multiple ideal characteristics of feedback, as well as multiple methods of providing it.

Science of Teams (FD)

The Salas et al. (2005) framework of the five teamwork components that are required to complete any team task (team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation) can influence feedback design.

Though some research concludes that team leadership is not important in most situations (Fransen, Weinberger, & Kirschner, 2013), others contend that it is an important contribution to team effectiveness (Zaccaro, Rittman, & Marks, 2001). Team leadership facilitates team problem solving by enabling shared mental models, coordination, and motivation (Salas et al., 2005). Team leaders should receive feedback that shows how well they are facilitating the team. The feedback given to the team and the timing of that feedback would change if the focus of the task at hand changes (e.g., if the team already had the knowledge to complete the task and was focused on efficiency).

Team member need to monitor other members' work while performing their own tasks. The monitoring includes ensuring that progress is on schedule, everything is functioning as expected, and team members are following procedures (McIntyre & Salas, 1995). It is difficult to measure mutual performance monitoring as it may not be unobtrusively observable (Salas et al., 2005). Feedback pertaining to the mutual performance monitoring of a team would most likely be focused on the team's process (Walton et al., 2014). Also, it is difficult to determine if members of a team exhibit successful mutual performance monitoring if no problem ever arises during a task. Thus, studying a team's performance often requires multiple trials.

When a teammate recognizes a need to provide resources and task-related efforts to another member, he or she is backing up the lack of performance of the teammate. For instance, a team member may recognize a problem with the distribution of workload within the team and adjust accordingly (Porter et al., 2003). There are different ways that members of a team can provide backup behavior. For example, members of a team can provide verbal feedback and coaching to help improve performance (Marks, Mathieu, & Zaccaro, 2001). Members can also go beyond that by assisting a teammate in performing a task when overload is detected (Marks et al., 2001; Salas et al., 2005). If feedback pertaining to backup behavior is to help team members identify and act when other members need help, then feedback should be given in real time (Walton et al., 2014). However, if the task at hand is to give the team a chance to practice their skills, then feedback should be delayed until the end of the task.

Teams adapt to continuously changing tasks utilizing the knowledge, skills, and attitudes to recognize deviations from anticipated actions and adjust actions accordingly (Priest, Burke, Munim, & Salas, 2002). An ITTS should be able to compare the actions of a team during a task to the expected action in order to understand adaptability (Walton et al., 2014).

An individual's satisfaction with their own effort and performance can be improved by a team orientation attitude (Salas et al., 2005). Team orientation can facilitate overall performance (Driskell & Salas, 1992; Eby & Dobbins, 1997), and influence team cooperation behaviors (Eby & Dobbins, 1997). An ITTS should be able to give feedback pertaining to team orientation.

Data Analysis (FD)

The data analyst will want to explore whether the feedback is effective with that question divided into several components based on the many characteristics of feedback mentioned above. Context is the most complex issue. It may be, for example, that a longer feedback phrase such as, "Remember to acknowledge your teammates whenever they speak to you," is more effective at the beginning of a task scenario when there is less cognitive load, but later in the scenario, as the action accelerates, learners have time only to attend to a short alert such as, "Acknowledge!" Thus, the data analyst who wants to know whether a given piece of feedback is effective (whether it be a spoken phrase, an indicator light, or an audio alert) must be sure to log all possible system state variables that might relate to its effectiveness, as well as state and trait variables of the players, i.e., information about their previous team skill and task skill levels as well as their current performance state and mood. Players with high cognitive load may perceive feedback differently than those with less load. As Rosalind Picard discussed in *Affective Computing* (1997), and Nass et al. (2005) explored with varied emotions of a car voice, people perform better when the tone of an agent like a tutor matches the tone of the learner.

Assuming all the appropriate data are gathered to assess the impact of each element of feedback, then timing and frequency become an important consideration. If the team members are wearing physiological sensors, for example, it is critical to know the latency of a physiological signal in order to conclude which of many events in the scenario might have caused a spike. Also, time-based data of all the team members' actions, communications, scenario events, feedback items given, and etc., lend themselves easily to correlational analysis, but not always to causal analysis. To know whether a given piece of feedback actually led to better team communication, for example, the analyst needs enough data to locate instances of team communication that occurred both with and without the feedback item, ideally in multiple teams, to compare the control group with the teams that received feedback.

Human-Computer Interaction (FD)

A variety of efforts in the human factors literature analyze the design of user alerts (Sorkin & Woods, 1985), particularly in such human safety and life-critical domains as healthcare (Russ et al., 2014), cockpit design (Thomas & Rantanen, 2006), and nuclear power plant control room design (Kinkade & Anderson, 1984). Some alerts distinguish between prompts to encourage a specific behavior (Herrmann & Nierhoff, 2017); cues to let users know the status of the system (Byrne, 2008); and decision aids to help users make a decision under pressure or high cognitive load (Glover, Prawitt, & Spilker, 1997; Todd & Benbasat, 1994). These human factors approaches offer a systems view of feedback design to the ITS community which has tended to focus more on the instructional design of feedback (e.g. Murray, 1999; Roll et al., 2011).

Tutor Authoring (TA)

Tutor Engineering (TA)

An ITTS must keep track of each step involved in a task, the progress of each team member in completing the steps, the order in which the steps must be completed, and any interdependence between steps. The structure of the database that records this information is critical. Also, if team members have specific roles, the tutor must track the responsibility of each team member to each task and model the relationships between teammates (Rickel & Johnson, 1998; Salas et al., 2004). Creating an ITTS requires at least three new modules: an expert model for teaching team skills, a pedagogical model that can decide when to give feedback to specific members vs. the whole team, and a team model to record the current team skills of the team. Also, there will be a new individual learner model for each individual.

Intelligent team tutors must adapt behavior in real-time to best support learning. Adaptive systems have four general categories of modification: function allocation, task scheduling, interaction style, and content (Feigh, Dorneich, & Hayes, 2012). Traditional ITS design has focused on the cognitive aspects of learning, such as assessing student content knowledge to trigger tutor feedback (Roll et al., 2011; Wood & Wood, 1999; Zakharov, Mitrovic, & Johnston, 2008). Strategies for adapting the tutoring experience include changing task difficulty (Harley, Lajoie, Frasson, & Hall, 2015), adjusting timing and difficulty of assessments (Arroyo et al., 2014), or providing additional examples and hints (Chaffar, Derbali, & Frasson, 2009; Woolf et al., 2009). Affect-aware tutoring considers the emotional state of leaners when deciding what actions to take. Given the interpersonal aspects of teams, detection of and adaptation to affect may play a role in an ITTS. Emotion-focused actions include providing empathetic responses (D'mello & Graesser, 2012; Mao & Li, 2009), mirroring the learner emotions (Picard et al., 2004; Zakharov et al., 2008), or providing behavioral prompts (D’Mello & Graesser, 2014). More recently, there has been research exploring the viability of changing the interaction style of ITSs when providing feedback, without changing the content of the feedback (Yang & Dorneich, 2016). The ideal ITTS authoring software architecture would enable authors to specify these forms of adaptation.

Learning Sciences (TA)

The Learning by Teaching theory (Biswas, Leelawong, Schwartz, Vye, & The Teachable Agents Group at Vanderbilt, 2005) and as well as Case-Based Reasoning (Kolodner, 1992) suggest that the process of formalizing knowledge into a particular representation can change the way experts think about their own expertise. That is, the authoring process makes explicit what has previously been only implicit for the expert. Ideally, the authoring tools used do not force undesired conceptual change or representing the knowledge in a way that is not natural for the expert.

Apprenticeship learning is a pedagogy based on theories of cognitive apprenticeship (Collins, Brown, & Newman, 1987; Hoppe, 1993) and situated learning (Lave & Wenger, 1991). In this approach, situated learning occurs when apprentices are active participants in an activity, within a team and usually with an expert. Apprentices' process of learning moves them from peripheral participation to full participation in the authentic activities of a community of practice, while the expert "fades" from engagement of the activity as the apprentices gain competency. By situating authentic practice within a problem-based learning paradigm, students learn as they engage in meaningful activities in pursuit of project goals. In settings where collaboration is an important and natural mechanism for learning and instruction, teams of learners can acquire valuable team skills. Apprentice learning (Lajoie & Alan, 1992) can be the basis of an ITTS, where the tutor can facilitate the interactions, or act as the expert (e.g., Dorneich & Jones, 2001). An ITTS authoring tool must support creating ITTSs using this approach.

Science of Teams (TA)

As we have discussed earlier in this chapter an ITTS has to not only train individuals in completing a task but also teach team skills (e.g., communication, coordination) and other team constructs. The ITTS author will need to create the expert model for the team skills required in the team task at hand. However, because team skills typically apply across multiple tasks, it is useful if the expert model can be created in a way that promotes re-use across multiple ITTSs with only minor adjustments. A good authoring tool for the expert model of team skills would provide transparency and easy control over the elements of the model that can be re-used.

Also, the author of the ITTS might well be an expert at the task skills involved in the task being tutored, but be less of an expert at coaching team skills. Thus, as far as possible, the authoring tool should prompt the author to consider such issues, e.g., "Since you have indicated that your team has four people, each with a different role, you might include the team skill module on interdisciplinary communication." An ITTS authoring system should adapt to the author's level of knowledge about teams.

Data Analysis (TA)

From the data analyst perspective, the team tutor authoring process is a usability challenge. The data analyst might want decide who the most efficient ITTS author is among a group of authors, for example, or use clickstream data in the authoring software to evaluate the software itself. Researchers at Google and Stanford, for example, predicted many of the severe usability errors by counting undo and erase events in Google SketchUp (Akers, Simpson, Jeffries, & Winograd, 2009). A variety of analyses of the usability of ITS authoring tools are relevant to the ITTSs. These efforts include examining whether computational thinking (Wing, 2008) was required (Blessing & Gilbert, 2008; Gilbert, Blessing, & Kodavali, 2009), analyzing the time spent in each screen of the authoring tool (Gilbert, Blessing, & Guo, 2015), and using Green and Petre's cognitive dimensions framework (1996) to compare authoring systems. A survey of the authoring process within five different ITS authoring tools gives more detail into the process

(Blessing et al., 2015). These examples might allow the data analyst to choose appropriate measures for establishing the effectiveness of the authoring tool. If the data analyst wishes to compare the quality of the tutors themselves with each other, one simple approach is to use rubrics as described in (Martin, Mitrovic, & Suraweera, 2007). Although no software tools are currently designed for authoring ITTSs, many of the approaches described above can be extended to evaluating such software as it emerges.

Human-Computer Interaction (TA)

Traditionally a maxim of HCI is to minimize the expectation-execution gap (Norman, 1988) in the user interface. That means the ITTS author's mental model of the domain should be easily encoded into the ITTS using an interface that fits naturally to the needs of that model. However, authoring an ITTS typically requires computational thinking, and not every team training expert excels at that. To make the authoring experience as natural as possible, the UI needs to use the same constructs that the training expert uses, e.g., team members, skills, and personality traits, rather than nodes, conditions, properties, and other technical terms.

Using an HCI perspective, Gilbert, et al. (2017) describe ten cognitive activities within authoring an ITTS, the following five being an abbreviated list. The first is describing the team task, team member roles, and team context, e.g., "This is an assembly line task with three operators and one supervisor, and there are several subtasks for each job role…" Several user interface approaches are available for this task. A wizard interface such as those used in popular home software packages to complete annual tax forms and asks the expert a series of questions that will branch into more questions depending on the answer to the previous question. An ontology editor, such as Protégé (Noy et al., 2003), that offers a more computationally powerful approach, but also requires more computational thinking. In the assembly line example, the author might specify parent objects, their child objects, and their properties, e.g., an assembly line has stations, and stations have tools. An assembly line also has a product, and products have subassemblies, and subassemblies have parts. Subassemblies have a property called "required tools." Finally, a method that combines visual programming (Green & Petre, 1996) and the programming by demonstration method (Aleven, Sewall, McLaren, & Koedinger, 2006; Cypher & Halbert, 1993) can be used. In this approach, used mostly for procedural team tasks, the author might create a flowchart bubble for each step of the procedure that the team will follow.

The second, third, and four major authoring activities, described in more detail in the section on performance metrics and feedback design are: create the rules for feedback, establish performance measures, and create the feedback and tutor pedagogy.

The fifth activity is previewing the learner experience. Team tasks are often complex, transitioning from one world-state to another in an unpredictable fashion. Thus, just like the designer of a museum exhibit, in which components must often create an experience that makes sense when explored in any sequence, the author must be confident that the learning experience created will work smoothly no matter the method in which the task unfolds. This activity can be a challenge with an individual tutor, but is more so with an ITTS, when the outcomes may depend on team members' personalities and approach to working together. Ideally, a UI would first give the author feedback if there were any unresolved paths through the experience, e.g., "Author: There are two situations in which the student will do something wrong but receive no feedback. They are …" Once all paths are resolved, the author could request a preview with different team members personas, e.g., with one team member who does everything right, and one member who does things wrong by omitting actions, and one team member who does things wrong by choosing the wrong actions. Perhaps those team members would also have team skill profiles, e.g., two are good communicators and the third is not. And the team members might have different roles in the task, each of which might merit its own form of preview. This preview experience is key for providing the author an impression of what kind of team tutor he or she has created. Is the tutor friendly? Naggy? Encouraging? Vague? Direct? If the author is not satisfied with the tutor, he or she will return to the stage of editing the feedback.

CONCLUSION

In this chapter we described a multidisciplinary perspective to mitigate the challenges in developing ITTSs. As software agent intelligence increases, the authors recommend adding another lens on human-agent teaming and human-robot interaction. While these terms, along with human-agent interaction, human-robot collaboration have different nuances the authors recommend that ITTS designers study this domain to leverage the relevant research. Concepts from this field such as function (or task) allocation (Sheridan, 2000), levels of automation (Johnson et al., 2011; Parasuraman, Sheridan, & Wickens, 2000), adaptive agent automation based on human state (Feigh et al., 2012; Lohani et al., 2017), and behavior modeling (Silverman, Johns, Cornwell, & O'Brien, 2006), for example, will be important as agents become co-working team members rather than simply tutors.

Acknowledgments

The work was completed with support from the 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 States Government, and no official endorsement should be inferred.

References

Akers, D., Simpson, M., Jeffries, R., & Winograd, T. (2009). *Undo and erase events as indicators of usability problems.* Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Aleven, V., Sewall, J., McLaren, B. M., & Koedinger, K. R. (2006). *Rapid Authoring of Intelligent Tutors for Real-World and Experimental Use.* Paper presented at the Sixth International Conference on Advanced Learning Technologies.

Anderson, J. R., Conrad, F. G., & Corbett, A. T. (1989). Skill acquisition and the LISP tutor. *Cognitive Science, 13*, 467–505.

Arroyo, I., Woolf, B. P., Burelson, W., Muldner, K., Rai, D., & Tai, M. (2014). A multimedia adaptive tutoring system for mathematics that addresses cognition, metacognition and affect. *International Journal of Artificial Intelligence in Education, 24*(4), 387-426.

Arroyo-Figueroa, G., Hernandez, Y., & Sucar, E. (2006). Intelligent Environment for Training of Power Systems Operators. In B. Gabrys, R. J. Howlett & L. C. Jain (Eds.), (pp. 943-950). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bannister, S. L., Wickenheiser, H. M., & Keegan, D. A. (2014). Key Elements of Highly Effective Teams. *Pediatrics, 133*(2), 184-186. doi: 10.1542/peds.2013-3734

Bannon, L. J., & Schmidt, K. (1989). CSCW-four characters in search of a context *DAIMI Report Series* (Vol. 289).

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by Teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392. doi: 10.1080/08839510590910200

Blessing, S., & Gilbert, S. B. (2008). *Evaluating an Authoring Tool for Model-Tracing Intelligent Tutoring Systems.* Paper presented at the 9th International Conference on Intelligent Tutoring Systems.

Blessing, S. B., Aleven, V., Gilbert, S. B., Heffernan, N. T., Matsuda, N., & Mitrovic, A. (2015). Authoring Example-based Tutors for Procedural Tasks. In R. Sottilare, A. Graesser, X. Hu & K. Brawner (Eds.), *Design Recommendations for Intelligent Tutoring Systems* (Vol. 3, pp. 71-93). Orlando, FL: U.S. Army Research Laboratory.

Blessing, S. B., Gilbert, S. B., Ourada, S., & Ritter, S. (2009). Authoring model-tracing cognitive tutors. *International Journal for Artificial Intelligence in Education.*

Bohemia Interaction Simulations (Producer). (2011). Virtual Battlespace 2. Retrieved from http://products.bisimulations.com/products/vbs2

Bonner, D., Gilbert, S., Dorneich, M. C., Burke, S., Walton, J., Ray, C., et al. (2015). Taxonomy of Teams, Team Tasks, and Tutors. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym2)* (pp. 189-198).

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., et al. (2016). *The Challenges of Building Intelligent Tutoring Systems for Teams*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting. http://journals.sagepub.com/doi/10.1177/1541931213601451

Bonner, D., Gilbert, S., Winer, E., Dorneich, M., MacAllister, A., Kohl, A., et al. (2017). *Military Team Training Utilizing GIFT.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Bradáč, V., & Kostolányová, K. (2016). Intelligent Tutoring Systems. In G. Vincenti, A. Bucciero, M. Helfert & M. Glowatz (Eds.), *E-Learning, E-Education, and Online Training (eLEOT)* (pp. 71-78): Springer International Publishing.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*(6), 1189-1207. doi: 10.1037/0021-9010.91.6.1189

Byrne, M. D. (2008). *Preventing Postcompletion errors: How much cue is enough?* Paper presented at the Proceedings of the Cognitive Science Society.

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations Between Work Group Characteristics and Effectiveness: Implications for Designing Effective Work Groups. *Personnel Psychology, 46*(4), 823-847. doi: 10.1111/j.1744-6570.1993.tb01571.x

Campion, M. A., Papper, E. M., & Medsker, G. J. (1996). Relations Between Work Group Characteristics and Effectiveness: A replication and Extension. *Personnel Psychology, 49*(2), 429-452. doi: 10.1111/j.1744-6570.1996.tb01806.x

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*(2), 195-202.
Carroll, J. M. (2010). Conceptualizing a possible discipline of human–computer interaction. *Interacting with Computers, 22*(1), 3-12. doi: 10.1016/j.intcom.2009.11.008
Chaffar, S., Derbali, L., & Frasson, C. (2009). Inducing positive emotional state in Intelligent Tutoring Systems. In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (pp. 716-718): IOS Press.
Collins, A., Brown, J. S., & Newman, S. E. (1987). Cognitive Apprenticeship: Teaching the Craft of Reading, Writing, and Mathematics. Hillsdale, NJ.
Corbett, A. T., & Anderson, J. R. (1992). *Student modeling and mastery learning in a computer-based programming tutor.* Paper presented at the International Conference on Intelligent Tutoring Systems, Montreal, Canada.
Crouch, C. H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69*(9), 970-977.
Cypher, A., & Halbert, D. C. (1993). *Watch what I do: programming by demonstration*: MIT press.
D'mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(4), 23.
D’Mello, S. K., & Graesser, A. C. (2014). Feeling, thinking, and computing with affect-aware learning. In R. A. Calvo, S. K. D’Mello, J. Gratch & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 419-434).
DeShon, R. P., Kozlowski, S. W. J., 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. *The Journal of applied psychology, 89*(6), 1035-1056. doi: 10.1037/0021-9010.89.6.1035
Dix, A., Finaly, J., Abowd, G., & Beale, R. (2001). *Human Computer Interaction, 2003* (3rd ed.): Pearson Prentice Hall.
Dorn, C., & Dustdar, S. (2010). Composing Near-Optimal Expert Teams: A Trade-Off between Skills and Connectivity. In *On the Move to Meaningful Internet Systems: OTM 2010* (pp. 472-489).
Dorneich, M., Passinger, B., Hamblin, C., Keinrath, C., Vašek, J., Whitlow, S. D., et al. (2017). Utilizing Open-loop Cognitive State Feedback to Drive Dynamic Task Sharing. *Frontiers in Neuroscience (Special Issue: New Advances in Computational Neuroergonomics and Adaptive Human-Automation Systems), 11*(144).
Dorneich, M. C., & Jones, P. M. (2001). The UIUC Virtual Spectrometer: A Java-Based Collaborative Learning Environment. *Journal of Engineering Education, 90*(4), 713-720. doi: 10.1002/j.2168-9830.2001.tb00663.x
Driskell, J. E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 34*(3), 277-288.
Eby, L. T., & Dobbins, G. H. (1997). Collectivistic orientation in teams: An individual and group-level analysis. *Journal of Organizational Behavior, 18*, 275-295.
Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 65-84.
Farley, J. (1998). *Java distributed computing*. Cambridge: O'Reilly.
Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. *Human Factors, 54*(6), 1008-1024.
Fiore, S. M. (2008). Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research, 39*(3), 251-277.
Fransen, J., Weinberger, A., & Kirschner, P. A. (2013). Team effectiveness and team development in CSCL. *Educational psychologist, 48*(1), 9-24.
Gamper, J., & Knapp, J. (2002). A Review of Intelligent CALL Systems. *Computer Assisted Language Learning, 15*(5), 329-342.
Gilbert, S., Blessing, S., & Guo, E. (2015). Authoring Effective Embedded Tutors: An Overview of the Extensible Problem Specific Tutor (xPST) System. *International Journal of Artificial Intelligence in Education, 25*(3), 428-454.
Gilbert, S. B., Blessing, S. B., & Blankenship, L. (2009). *The Accidental Tutor: Overlaying an Intelligent Tutor on an Existing User Interface.* Paper presented at the CHI '09 Extended Abstracts on Human Factors in Computing Systems.

Gilbert, S. B., Blessing, S. B., & Kodavali, S. (2009). *The Extensible Problem-Specific Tutor (xPST): Evaluation of an API for Tutoring on Existing Interfaces.* Paper presented at the 14th International Conference on Artificial Intelligence in Education.

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., et al. (2017). Creating a Team Tutor Using GIFT. *International Journal of Artificial Intelligence in Education*. doi: 10.1007/s40593-017-0151-2

Glover, S. M., Prawitt, D. F., & Spilker, B. C. (1997). The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes, 72*(2), 232-255.

Green, T. R. G., & Petre, M. (1996). Usability Analysis of Visual Programming environments: A ‘cognitive dimensions’ framework. *Journal of Visual Languages and Computing, 7*, 131-174.

Grudin, J. (1994). Computer-supported cooperative work: History and focus. *Computer, 27*(5), 19-26. doi: 10.1109/2.291294

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2015). An integrated emotion-aware framework for intelligent tutoring systems. In C. Conati, N. Heffernan, A. Mitrovic & M. Verdejo (Eds.), *Artificial Intelligence in Education* (pp. 616-619): Springer.

Hartley, J. R., & Sleeman, D. H. (1973). Towards More Intelligent Teaching Systems. *International Journal of Man-Machine Studies*, *5*(2), 215-236.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Herbrich, R., Minka, T., & Graepel, T. (2007). *TrueSkill™: a Bayesian skill rating system.* Paper presented at the Advances in neural information processing systems.

Herrmann, T., & Nierhoff, J. (2017). *Prompting–A Feature of General Relevance in HCI-Supported Task Workflows.* Paper presented at the International Conference on Human-Computer Interaction.

Hollenbeck, J. R., Moon, H., Ellis, A. P. J., West, B. J., Ilgen, D. R., Sheppard, L., et al. (2002). Structural contingency theory and individual differences: Examination of external and internal person-team fit. *Journal of Applied Psychology, 87*(3), 599-606. doi: 10.1037//0021-9010.87.3.599

Hoppe, H. U. (1993). Cognitive Apprenticeship - The Emperor's New Method? A Polemical Reaction to the Debate on Situated Cognition and Cognitive Apprenticeship. *Journal of Artificial Intelligence in Education, 4*(1), 49-54.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Hoffman, R. R., Jonker, C., van Riemsdijk, B., et al. (2011). Beyond cooperative robotics: The central role of interdependence in coactive design. *IEEE Intelligent Systems, 26*(3), 81-88.

Kinkade, R., & Anderson, J. (1984). Human factors guide for nuclear power plant control room development. Final report: Essex Corp., San Diego, CA (USA).

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284. doi: 10.1037/0033-2909.119.2.254

Koedinger, K. R., Aleven, V., Heffernan, N., McLaren, B., & Hockenberry, M. (2004). Opening the Door to Non-programmers: Authoring Intelligent Tutor Behavior by Demonstration. In J. C. Lester, R. M. Vicari & F. Paraguaçu (Eds.), *Intelligent Tutoring Systems* (Vol. 3220, pp. 7-10): Springer Berlin / Heidelberg.

Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal for Artificial Intelligence in Education*, *8*, 30-43.

Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review*, *6*(1), 3-34.

Kolodner, J. L. (2004). The Learning Sciences: Past, Present, and Future. *Educational Technology: The Magazine for Managers of Change in Education*, *44*(3), 37-42.

Kowalski, J., Zhang, X., & Gordon, G. (2014). Statistical Modeling of Student Performance to Improve Chinese Dictation Skills with an Intelligent Tutor. *JEDM - Journal of Educational Data Mining*, *6*(1), 3-27.

Lajoie, S. P., & Alan, L. (1992). Apprenticeship training in the workplace: Computer-coached practice environment as a new form of apprenticeship. In M. J. Farr & J. Psotka (Eds.), (pp. 15-36). New York: Taylor and Francis.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*: Cambridge University Press.

Lohani, M., Stokes, C., Dashan, N., McCoy, M., Bailey, C. A., & Rivers, S. E. (2017). A Framework for Human-Agent Social Systems: The Role of Non-technical Factors in Operation Success. In *Advances in Human Factors in Robots and Unmanned Systems* (pp. 137-148): Springer.

Long, J., & Dowell, J. (1989). Conceptions of the discipline of HCI: craft, applied science and engineering. In A. Sutcliffe & L. Macaulay (Eds.), (pp. 9-32): Cambridge University Press.

Loughry, M. L., Ohland, M. W., & Woehr, D. J. (2014). Assessing teamwork skills for assurance of learning using CATME team tools. *Journal of Marketing Education, 36*(1), 5-19.

MacAllister, A., Kohl, A., Gilbert, S., Winer, E., Dorneich, M., Bonner, D., et al. (2017, Apr 25-27). *Analysis of Team Tutoring Training Data.* Paper presented at the MODSIM World 2017, VA Beach.

Malacria, S., Scarr, J., Cockburn, A., Gutwin, C., & Grossman, T. (2013). *Skillometers: reflective widgets that motivate and help users to improve performance.* Paper presented at the Proceedings of the 26th annual ACM symposium on User interface software and technology.

Mao, X., & Li, Z. (2009). *Implementing emotion-based user-aware e-learning.* Paper presented at the CHI'09 Extended Abstracts on Human Factors in Computing Systems.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of management review, 26*(3), 356-376.

Martin, B., Mitrovic, A., & Suraweera, P. (2007). Domain modelling with ontology: A case study. In A. Cristea & R. M. Carro (Eds.), *Proceedings of the 5th Int. Workshop on Authoring of Adaptive and Adaptable Hypermedia, User Modeling*.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14): Prentice-Hall Englewood Cliffs, NJ.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45). San Franscisco: Jossey-Bass.

McKendree, J. (1990). Effective Feedback Content for Tutoring Complex Skills. *Human-Computer Interaction, 5*(4), 381-413. doi: 10.1207/s15327051hci0504_2

Merrill, D. C., Reiser, B.J., Ranney, M., & Gregory, J. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences*.

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2002). *Team-based learning: A transformative use of small groups*: Greenwood publishing group.

Mickan, S., & Rodger, S. (2000). Characteristics of effective teams: a literature review. *Australian Health Review, 23*(3), 201-208. doi: http://dx.doi.org/10.1071/AH000201

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (Vol. 2, pp. 745-784).

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education, 10*(1), 98-129.

Murray, T., Blessing, S., & Ainsworth, S. (Eds.). (2003). *Authoring Tools for Advanced Technology Learning Environments: Toward Cost-effective Adaptive, Interactive, and Intelligent Educational Software*. Norwell, MA: Kluwer Academic Publishers.

Nass, C., Jonsson, I.-M., Harris, H., Reaves, B., Endo, J., Brave, S., et al. (2005). *Improving automotive safety by pairing driver emotion and car voice emotion.* Paper presented at the CHI'05 Extended Abstracts on Human Factors in Computing Systems.

Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*(3), 376-389. doi: 10.1037/0021-9010.84.3.376

Nikolenko, S. I., & Sirotkin, A. V. (2010). *Extensions of the TrueSkillTM rating system.* Paper presented at the Proceedings of the 9th International Conference on Applications of Fuzzy Systems and Soft Computing.

Norman, D. (1988). *The design of everyday things*. New York: Basic Books.

Noy, N. F., Crubézy, M., Fergerson, R. W., Knublauch, H., Tu, S. W., Vendetti, J., et al. (2003). *Protege-2000: an open-source ontology-development and knowledge-acquisition environment.* Paper presented at the AMIA Annu Symp Proc.

Ofstad, W., & Brunner, L. J. (2013). Team-based learning in pharmacy education. *American journal of pharmaceutical education, 77*(4), 70.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30*(3), 286-297.

Pentland, A. (2012). The new science of building great teams. *Harvard Business Review, 90*(4), 60-69.

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.

Picard, R. W., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., et al. (2004). Affective learning—a manifesto. *BT technology journal, 22*(4), 253-269.

Porter, C. O., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P., West, B. J., & Moon, H. (2003). Backing up behaviors in teams: the role of personality and legitimacy of need. *Journal of Applied Psychology, 88*(3), 391.

Priest, H. A., Burke, C. S., Munim, D., & Salas, E. (2002). Understanding team adaptability: Initial theoretical and practical considerations. *Proceedings of the human factors and ergonomics society annual meeting, 46*(3), 561-565.

Rice, R. (1993). Media appropriateness: Using social presence theory to compare traditional and new organisational media. Cited in Stacey, E.(2002) Social presence online: Networking learners at a distance: Networking the Learner: Computers in Education. Boston, MA: Kluwer Academic Press.

Rickel, J., & Johnson, L. (1998). *Animated pedagogical agents for team training.* Paper presented at the Proceedings of the ITS Workshop on Pedagogical Agents.

Ritter, S., Harris, T. K., Nixon, T., Dickison, D., Murray, R. C., & Towle, B. (2009). Reducing the Knowledge Tracing Space. *International Working Group on Educational Data Mining*.

Ritter, S., & Koedinger, K. R. (1996). An architecture for plug-in tutor agents. *Journal of Artificial Intelligence in Education, 7*(3), 315-347.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267-280.

Roschelle, J., Suthers, D., & Grover, S. (2014). CIRCL Primer: Collaborative Learning, from http://circlcenter.org/collaborative-learning/

Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., et al. (2014). Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *Journal of the American Medical Informatics Association, 21*(e2), e287-e296.

Salas, E., Burke, C. S., Fowlkes, J. E., & Priest, H. A. (2004). On measuring teamwork skills. In *Comprehensive handbook of psychological assessment* (Vol. 4, pp. 427-442).

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for Enhancing Team Cognition in Complex Environments: The Power of Team Performance Diagnosis. *Aviation, Space, and Environmental Medicine, 78*(5), B77-B85.

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. doi: 10.1002/hrm.21628

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small group research, 36*(5), 555-599.

Salas, E. E., & Fiore, S. M. (2004). *Team cognition: Understanding the factors that drive process and performance*: American Psychological Association.

Sawyer, R. K. (2014). The New Science of Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 1-18): Cambridge University Press.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice-based assessments for the digital age*. Cambridge: MIT Press.

Shallice, T., & Burgess, P. W. (1991). Deficits in Strategy Application Following Frontal Lobe Damage in Man. *Brain, 114*, 727-741.

Sheridan, T. B. (2000). Function allocation: algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies, 52*(2), 203-216.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction, 55*(2), 503-524.

Shute, V. J., & Psotka, J. (1994). Intelligent tutoring systems: Past, present, future. *Technical Report AL/HR-TP-1994-0005, USAF, Armstrong Laboratory*.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects*, 295-321.

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: part I: enabling science with PMFserv. *Presence: Teleoperators and Virtual Environments, 15*(2), 139-162.
Smith, K. H. (1972). Changes in group structure through individual and group feedback. *Journal of Personality and Social Psychology, 24*(3), 425.
Smith, S. L., & Ward, P. (2006). Behavioral interventions to improve performance in collegiate football. *Journal of applied behavior analysis, 39*(3), 385-391.
Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human-computer interaction, 1*(1), 49-75.
Sottilare, R., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT).
Sottilare, R. A., Holden, H., Brawner, K., & Goldberg, B. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *U.S. Army Research Laboratory, Human Research and Engineering Directorate, Orlando, Florida*.
Sottilare, R. A., Shawn Burke, C., 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*. doi: 10.1007/s40593-017-0146-z
Stacy, W., & Freeman, J. (2016). Training objective packages: enhancing the effectiveness of experiential training. *Theoretical Issues in Ergonomics Science, 17*(2), 149-168.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In (pp. 409-426). Cambridge, UK.
Stokes, J. V., Luiselli, J. K., Reed, D. D., & Fleming, R. K. (2010). Behavioral coaching to improve offensive line pass-blocking skills of high school football athletes. *Journal of applied behavior analysis, 43*(3), 463-472. doi: 10.1901/jaba.2010.43-463
Swain, K., & Mills, V. (2003). Implicit communication in novice and expert teams: Defence Science and Technology Organisation Salisbury (Australia) Systems Sciences Lab.
Thomas, L. C., & Rantanen, E. M. (2006). Human factors issues in implementation of advanced aviation technologies: A case of false alerts and cockpit displays of traffic information. *Theoretical Issues in Ergonomics Science, 7*(5), 501-523.
Todd, P. A., & Benbasat, I. (1994). The influence of decision aids on choice strategies under conditions of high cognitive load. *IEEE Transactions on Systems, Man, and Cybernetics, 24*(4), 537-547.
van der Meij, H. (2013). Do Pedagogical Agents Enhance Software Training? *Human–Computer Interaction, 28*(6), 518-547. doi: 10.1080/07370024.2013.789348
VanLehn, K. (2006). The behavior of tutoring systems. *International journal of artificial intelligence in education, 16*(3), 227-265.
VanLehn, K., van de Sande, B., Shelby, R., & Gershman, S. (2010). The Andes Physics Tutoring System: An Experiment in Freedom. In R. Nkambou, J. Bourdeau & R. Mizoguchi (Eds.), *Advances in Intelligent Tutoring Systems* (Vol. 308, pp. 421-443): Springer Berlin Heidelberg.
Walton, J., Bonner, D., Walker, K., Mater, S., Dorneich, M., Gilbert, S., et al. (2015). *The Team Multiple Errands Test: A Platform to Evaluate Distributed Teams.* Paper presented at the Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing.
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)* (pp. 199-207).
Walton, J., Gilbert, S., Winer, E., Dorneich, M., & Bonner, D. (2015). *Evaluating Distributed Teams with the Team Multiple Errands Test.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.
Wernick, M. (1998). *Evaluating collaborative technology: towards a framework and an instrument.* University of Illinois at Urbana-Champaign.
Wing, J. M. (2008). Computational Thinking and Thinking about Computing. *Philosophical Transactions of the Royal Society, 366*, 3717-3725.
Winne, P. H., & Baker, R. S. J. D. (2013). The Potentials of Educational Data Mining for Researching Metacognition, Motivation and Self-Regulated Learning. *Journal of Educational Data Mining, 5*(1).

Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education, 33*(2), 153-169.

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology, 4*(3-4), 129-164.

Yang, E., & Dorneich, M. C. (2016). *Evaluation of Etiquette Strategies to Adapt Feedback In Affect-Aware Tutoring.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Zaccaro, S. J., Rittman, A. L., & Marks, M. A. (2001). Team leadership. *The Leadership Quarterly, 12*(4), 451-483.

Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. *International Journal for Artificial Intelligence in Education*, *10*, 257-277.

Zakharov, K., Mitrovic, A., & Johnston, L. (2008). *Towards emotionally-intelligent pedagogical agents.* Paper presented at the Intelligent Tutoring Systems.