



ARL-TR-8893 • JAN 2020



Toward a Scientifically Rooted Design Architecture of Team Process and Performance Modeling in Adaptive, Team- Based Intelligent Tutoring Systems

**by Robert Sottolare, Shawn Burke, Anne M Sinatra,
Joan Johnston, and Eduardo Salas**

Approved for public release; distribution is unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



Toward a Scientifically Rooted Design Architecture of Team Process and Performance Modeling in Adaptive, Team-Based Intelligent Tutoring Systems

Robert Sottolare
Soar Technology, Inc.

Shawn Burke
Institute for Simulation and Training, University of Central Florida

Anne M Sinatra and Joan Johnston
*US Army Combat Capabilities Development Command, Soldier Center,
Simulation and Training Technology Center*

Eduardo Salas
Rice University

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) January 2020		2. REPORT TYPE Technical Report		3. DATES COVERED November 2013–December 2016	
4. TITLE AND SUBTITLE Toward a Scientifically Rooted Design Architecture of Team Process and Performance Modeling in Adaptive, Team-Based Intelligent Tutoring Systems			5a. CONTRACT NUMBER W911NF-13-2-0006		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Robert Sottolare, Shawn Burke, Anne M Sinatra, Joan Johnston, and Eduardo Salas			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) US Army CCDC Soldier Center, Simulation and Training Technology Center* ATTN: FCDD-SCT-CE 12423 Research Parkway, Orlando, FL 32826			8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-8893		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES * The work detailed in this report was performed by the US Army Research Laboratory Human Research and Engineering Directorate. As of this report's publication, the Simulation and Training Technology Center is part of the US Army Combat Capabilities Development Command Soldier Center.					
14. ABSTRACT The goal of this research was to support the development of a practical architecture for the computer-based tutoring of teams. This report examines the relationship of various team behaviors, attitudes, and cognition as antecedents to successful team performance, learning, and performance during adaptive instruction guided by Intelligent Tutoring Systems (ITSs). Adaptive instruction is a training or educational experience tailored by artificially intelligent, computer-based tutors with the goal of optimizing learner outcomes (e.g., knowledge and skill acquisition, performance, enhanced retention, accelerated learning, or transfer of skills from instructional environments to work environments). Teamwork focuses on the coordination, cooperation, and communication among individuals to achieve a shared goal. For ITSs to optimally tailor team instruction, tutors must have key insights about team members, roles, and team objectives. To aid the modeling of teams, we examined the literature to evaluate the relationship of teamwork behaviors, attitudes, and cognition with team outcomes (learning and performance) as part of a large-scale meta-analysis of the ITS, team training, and team-performance literature. The goal is to develop and apply findings of this meta-analysis to guide instructional decisions by the Generalized Intelligent Framework for Tutoring, an open-source architecture for authoring, delivering, managing, and evaluating adaptive instruction.					
15. SUBJECT TERMS adaptive instruction, team tutoring, intelligent tutoring systems, team modeling, Generalized Intelligent Framework for Tutoring, GIFT					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 119	19a. NAME OF RESPONSIBLE PERSON Robert Sottolare
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) 407-208-3007

Contents

List of Figures	vi
List of Tables	vi
Acknowledgments	viii
1. Introduction	1
1.1 Team Taskwork, Teamwork, and Collaborative Learning	2
1.2 Evolving Toward a Team Tutor	3
1.3 Envisioning GIFT as a Team Tutor	4
2. A Guiding Framework for Tutoring Teams	7
3. Conceptualizing Team Effectiveness	9
3.1 Team Performance	10
3.2 Team Learning	11
3.3 Team Satisfaction	12
3.4 Team Viability	13
4. Meta-Analysis of Effective Team Constructs	13
4.1 Results	14
4.2 Behaviors Examined	14
4.2.1 Team Performance	14
4.2.2 Team Learning	18
4.2.3 Team Satisfaction	20
4.2.4 Team Viability	22
5. Effective Team Constructs Revisited	24
5.1 Team Performance	24
5.2 Team Learning	25
5.3 Team Satisfaction	26
5.4 Team Viability	27
5.5 Discussion	28

6.	Dynamic Assessment	30
6.1	Methodology	31
6.2	Marker Development	32
6.2.1	Team Psychological Safety	32
6.2.2	Trust Markers	33
6.2.3	Collective Efficacy	35
6.2.4	Cohesion	37
6.2.5	Conflict Markers	39
6.2.6	Team Communication	41
6.3	A Way Forward: The Testing of the Proposed Markers	43
6.3.1	Gaining Evidence of Content Validity	44
6.3.2	Gaining Evidence toward Construct Validity	44
6.4	Team Indexing and Aggregation	45
6.4.1	Conceptually Developing Team Constructs	46
6.4.2	Refining Measure and Item Referents	47
6.4.3	Selecting Agreement Indices	47
6.4.4	Choosing an Aggregation Technique	49
6.5	Discussion	49
7.	Feedback	50
7.1	Team-Feedback Review	51
7.2	Is Team Feedback Always the Right Solution?	52
7.3	Best Practices for Team Debrief	53
7.3.1	Team-Member Characteristics	53
7.3.2	Feedback Design	56
7.3.3	Work Environment	59
7.4	Implementation in Team-Tutoring Environment: Critical Question	63
7.5	How to Establish Credibility in the Automated Feedback Tool?	63
7.6	Does the Type of Task Influence How Team Feedback Is Implemented?	64
7.7	Is One Source of Feedback Delivery Ever Enough?	64
7.8	Is More Always Better?	65
8.	Workshop Summary	66
8.1	Breakout Session 1: Team Competency	67

8.2	Breakout Session 2: Team Pedagogy	69
8.3	Breakout Group 3: Team Assessment	70
8.4	Breakout Group 4: Team Feedback	73
9.	Conclusions	74
10.	References	76
	Appendix. Phase 2: Detailed Article Coding	96
	List of Symbols, Abbreviations, and Acronyms	108
	Distribution List	109

List of Figures

Fig. 1	Enhanced team-tutoring model in the GIFT	7
Fig. 2	Organizing a framework for thinking about team processes and instruction within an ITS	8
Fig. 3	Ontology for team performance based on meta-analytic structural equation modeling results	18
Fig. 4	Ontology for team learning based on meta-analytic structural equation modeling	20
Fig. 5	Ontology for team satisfaction based on meta-analytic structural equation modeling.....	22
Fig. 6	Ontology for team viability based on meta-analytic structural equation modeling	23
Fig. 7	Team performance-state models that should be a primary assessment when team performance is the targeted outcome.....	25
Fig. 8	Team learning-state models that should be a primary assessment when team learning is the targeted outcome	26
Fig. 9	Team satisfaction-state models that should be a primary assessment when team satisfaction is the targeted outcome.....	27
Fig. 10	Team viability-state models that should be a primary assessment when team viability is the targeted outcome	28

List of Tables

Table 1	MASEM results	14
Table 2	Behavioral contributions to team performance.....	15
Table 3	Attitudinal contributions to team performance: effect of cooperation	16
Table 4	Cognitive contributions to team performance.....	16
Table 5	Team-performance correlation matrix	17
Table 6	Behavioral contributions to team learning.....	19
Table 7	Attitudinal contributions to team learning: effect of cooperation.....	19
Table 8	Behavioral contributions to team satisfaction.....	21
Table 9	Attitudinal contributions to team satisfaction: effect of cooperation..	21
Table 10	Cognitive contributions to team satisfaction.....	22
Table 11	Key team constructs based on meta-analytic results.....	31
Table 12	Overview of process used to identify behavioral markers.....	32

Table 13	Potential markers of collective efficacy.....	35
Table 14	Potential markers of team cohesion	38
Table 15	Potential markers for team conflict.....	40
Table 16	Potential markers for team conflict management	40
Table 17	Communication markers' components	41
Table 18	Potential markers of team communication	42

Acknowledgments

Portions of this report were previously published in a journal article.¹ The information in the journal article was expanded, and the current report is a full account of the work conducted during the overall project and provides context for the progress throughout.

The authors would like to thank Elyse Burmester and Martin Kufus for their assistance with the editing of this report.

¹ Sottolare RA, Burke CS, Salas E, Sinatra AM, Johnston JH, Gilbert SB. Designing adaptive instruction for teams: a meta-analysis. *Inter J Art Intel Ed*. 2018; 28(2): 225–264.

1. Introduction

The team in all of its forms (e.g., crews, squads, platoons, and battalions) is the common organizational element describing a group of people working together to accomplish some goal or set of goals. Given the critical importance of military teams and their associated missions, it is also essential that these teams be adequately prepared for their missions through a rigorous and thorough training regimen. Training technology in the form of intelligent tutoring systems (ITSs) has been shown to be significantly more effective than traditional training methods (e.g., classroom training), but ITSs have primarily been applied to tasks for individual learners, not teams, and in a limited set of nonmilitary domains (e.g., mathematics, physics, and software programming).

To be a useful tool for team training, ITSs must be adapted to model team attitudes, behaviors, and cognition, which are much more complex than for individual learners.

The goal of the research described herein is the development of a practical, efficient architecture for the computer-based tutoring of teams. This report examines the relationship between team behaviors as precursors or antecedents for successful team performance and team learning (e.g., collaborative learning) during adaptive instruction as guided by ITSs, a subset of artificially intelligent technologies known as adaptive instructional systems (AISs).*

Sottolare and Brawner (2018) defined AISs as “computer-based systems that guide learning experiences by tailoring instruction and recommendations based on the goals, needs, and preferences of each learner in the context of domain learning objectives”. AISs guide and adapt training or educational experiences using artificially intelligent, computer-based logic to modify content, change instructional pace, and adjust feedback and actions by the tutor. Examples of AIS technologies include ITSs, intelligent mentors, instructional recommender systems, personal assistants for learning, and other intelligent instructional media.

The goal of adaptive instruction is to optimize learner outcomes (e.g., knowledge and skill acquisition, performance, enhanced retention, accelerated learning, or transfer of skills from instructional environments to work environments). Given this goal and the previously stated development goal of creating a practical, efficient architecture for computer-based tutoring of teams, we now move on to

* The work detailed in this report was performed by the US Army Research Laboratory Human Research and Engineering Directorate. As of this report’s publication, the Simulation and Training Technology Center is part of the US Army Combat Capabilities Development Command Soldier Center.

discussing teams, adaptive instruction for teams, and the Generalized Intelligent Framework for Tutoring (GIFT) as a tool for team tutoring.

1.1 Team Taskwork, Teamwork, and Collaborative Learning

We begin by distinguishing between team **taskwork** (the actions of team members related to team goals during the performance of an assigned task or mission) and **teamwork** (the attitudes, behaviors, and cognition of individual team members as they interact with other team members during the performance of an assigned task or mission), and **collaborative learning** (team members cooperating to acquire specific knowledge and skills).

Team taskwork focuses on developing proficiency in task domains that are required for a specific duty as part of one's job or assigned role (Salas 2015) and is generally considered to be a domain-dependent learning activity. Team taskwork often involves team training or working together to become more proficient at accomplishing a task. Team training, instructing teams to perform interdependent tasks, is often confused with the concept of teambuilding or teamwork, which focuses on improving how teams coordinate, share information, and share responsibilities (van Berlo 1997).

Teamwork includes “coordination, cooperation, and communication among individuals to achieve a shared goal” (Salas 2015, p. 5) where the attitudes, behaviors, and cognition within the team are largely independent of the taskwork. However, the performance of the team may influence team member attitudes, behaviors, and cognition so teamwork and taskwork are not totally independent. Teamwork activities may include teambuilding whose goal is to strengthen the coordination, cooperation, communication, coaching, conflict management, cohesion, and collective efficacy of the group (Salas 2015). Teamwork is a necessary prerequisite to satisfactory taskwork performance (van Berlo 1997).

Collaborative learning, also referred to as cooperative learning, is “a situation in which two or more people learn or attempt to learn something together” (Dillenbourg 1999, p. 1). Cooperative learning “can be defined as a structured form of group work where students pursue common goals while being assessed individually” (Prince 2004, p. 223) and reinforces active participation and active learning (van Berlo 1997) and thereby enhances engagement (Prince 2004). Collaborative learning generally focuses on a learning goal and is primarily domain dependent and includes computer-supported collaborative learning activities that may be delivered via an AIS like an ITS.

Teamwork, taskwork, collaborative learning, team learning, and team performance appear to have sufficient coverage within the general training and education literature. While the teamwork literature has received ample scholarly attention (Cannon-Bowers and Bowers 2011), much less is known about the relationship of core team attitudes, behaviors, and cognition, and their influences on teamwork and thereby team outcomes (e.g., learning, performance, satisfaction, and viability). Most notably there have been few domain-independent approaches to the development of teamwork models for computer-based tutoring.

1.2 Evolving Toward a Team Tutor

Since 2009, the use of computer-based tutoring programs for learning has seen a renewed interest in the social sciences, particularly in terms of using such systems for the training and development of teams. Sottolare (2009) put forward the notion that the maturation of machine perception of trainee affect (e.g., mood and emotions) is critically important to optimizing learning for both individuals and teams in embedded simulations where direct observation of trainees on operational platforms (e.g., tanks or aircraft) may not be reliable or even possible at times.

Sottolare (2010) posed the idea of developing ITSs for distributed team training. Sottolare et al. (2011) observed that scientists at that point were beginning to ponder the application of intelligent tutors and coaches in more challenging team-training domains where human tutors are either unavailable or impractical. Soon after Sottolare (2012) considered the development of an ontology for generalized architecture for individual learners that was realized as GIFT (Sottolare et al. 2012a; Sottolare et al. 2013; Sottolare et al. 2017) extended the GIFT ontology to understand the interactions between learner(s) and ITS functions in what became known as the learning-effect chain and later the learning-effect model (LEM). The LEM was gradually extended to include a model for team interaction during the tutoring process (Sottolare et al. 2018).

Understanding that the LEM is a model and not a functional team tutor, we began to examine the relationship of various attitudes, behaviors, and cognition that contribute to successful teamwork and examined the relationship of successful teamwork antecedents to team learning, performance, satisfaction, and viability. This report empirically examines these relationships through the lens of a meta-analytic methodology and provides additional material related to the team meta-analysis published by Sottolare et al. (2018). In essence, it updates the discussion about what is needed to enable computer-based tutoring programs to make effective instructional decisions during adaptive team-training experiences.

Before jumping into a discussion of the meta-analytic process, we would like to describe GIFT, its functional elements and processes, and how the results of a meta-analytic process could be used to enable GIFT to support adaptive instruction for teams.

1.3 Envisioning GIFT as a Team Tutor

There are many capabilities involved in the development of an effective ITS for teams. Whether the focus of the tutor is on taskwork, teamwork, or collaborative learning, each system is specially crafted around a particular context or content area and is not easily adapted for multiple purposes. However, researchers began to push for an ITS architecture to support adaptive instruction for both individuals and teams that could be used for multiple training domains (e.g., situations, scenarios, and problem sets). Such a push resulted in GIFT (Sottolare 2012; Sottolare and Goldberg 2012; Sottolare et al. 2012a; Sottolare et al. 2012b; Sottolare et al. 2017), which is driven by extensive prior research on ITSs and is designed to be adaptable for a range of individual instructional domains while also being adapted to support team-instructional domains.

GIFT, like other AISs, has four primary components that interact to provide a guided experience to learners (Elson-Cook 1993; Nkambou et al. 2010; Graesser et al. 2018; Psotka and Mutter 1988; Sleeman and Brown 1982; VanLehn 2006; Woolf 2009): the domain model, the student model, the instructional or tutoring model, and the user-interface model. GIFT similarly adopts this four-part component model, but with slightly different corresponding labels (domain module, learner module, pedagogical module, and tutor–user interface) and with the addition of the sensor module, which captures physiological and behavioral data about the learner and can be viewed as an expansion of the user interface.

“The *domain model* contains the set of skills, knowledge, and strategies/tactics of the topic being tutored. It normally contains the ideal expert knowledge and also the bugs, mal-rules, and misconceptions that students periodically exhibit” (Sottolare et al. 2018, p. 4). Expert knowledge is sometimes referred to as domain content. GIFT’s domain model contains a domain knowledge file that includes the concepts to be learned, the associated assessments, and the measures to support an evidence-centered assessment process—all without a human tutor in the instructional loop.

“The *learner model* consists of the cognitive, affective, motivational, and other psychological states that evolve during the course of learning. Since learner performance is primarily tracked in the domain model, the learner model is often

viewed as an overlay (subset) of the domain model, which changes over the course of tutoring” (Sottolare et al. 2018, p. 4).

The learner model discussed above is a real-time model that changes during instruction. A *long-term learner model* (LTLM) is also required to track the development of competence in various domains and bridge the gap between instructional experiences. The historical data that are represented in the LTLM can be in many forms. Learner data can be specific to an experience and represented in a learning-management system (e.g., Moodle, Blackboard, Canvas, or edX) or it can be the record of outcomes from an experience (e.g., experience application program interface or xAPI). Together, all of these historical data form a basis for evaluating the competence of the learner (i.e., novice, journeyman, or expert) in the next domain of instruction.

“The *tutor model* (also known as the pedagogical model or the instructional model) takes the domain and learner models as input and selects tutoring strategies, steps, and actions on what the tutor should do next in the exchange” (Sottolare et al. 2018, p. 4). In the design of mixed-initiative systems, learners may also take initiative by taking actions, asking questions, or requesting help (Aleven et al. 2006; Rus and Graesser 2009). GIFT currently has very limited mixed-initiative capabilities.

“The *user interface* interprets the learner’s contributions through various input media (speech, typing, clicking) and produces output in different media (text, diagrams, animations, agents)” (Sottolare et al. 2018, p. 4). Some common user interface features include 1) natural language interaction (Graesser 2016; Johnson and Lester 2016; Nye et al. 2014); 2) speech recognition (D’Mello et al. 2010; Litman 2012); and 3) sensing of learner emotions (Baker et al. 2010; D’Mello and Graesser 2010; Goldberg et al. 2011). To date, GIFT’s primary interface is text-based with some oral presentation by virtual humans.

Like the individual learner model, the *team model* consists of a set of states (e.g., learning and performance) relevant to the assessment of taskwork and teamwork and its antecedents. The team model must be able to track progress toward team learning objectives for training or collaborative learning goals, but also has the added complexity of monitoring teamwork antecedents (e.g., coordination, communication, coaching, and conflict management) that moderate or influence team and individual learning and performance (Sottolare et al. 2017). Team states are generally much more difficult to assess than individual learner models due to their larger number of individuals, their interactions, and the complexity in determining their shared cognition (shared mental model).

The designers of an AIS must make decisions on each of the various major components in order to create an enhanced learning experience through

well-grounded pedagogical strategies (optimal plans for action by the tutor) that are selected based on learner states and traits (learner-centric focus) and that are delivered to the learner as instructional tactics, which include optimal actions by the tutor that may take the form of specific feedback, questions, prompts, pumps, or directions to the learner(s), scaffolding or support (e.g., hints or requests to reflect), or modifications to the environment (e.g., increased/decreased complexity). Instructional tactics are chosen based on the previously selected strategies and instructional context (the conditions of the training environment at the time of the instructional decision). What has been described above is the *LEM for individuals and teams* (Sottolare 2012; Fletcher and Sottolare 2013; Sottolare et al. 2013; Sottolare et al. 2018), which is the basis for the GIFT architecture.

While GIFT and the LEM provide a starting point for a more advanced and flexible intelligent tutoring process, additional research is necessary to fully realize the potential of GIFT, especially as a team or collective tutoring system. Specifically, while initial team models have been theorized, further work is needed to identify a complete design architecture, including behavioral markers and metrics to support assessments during tutoring and that can be used to model team processes and performance in GIFT. This design architecture must be rooted not only in principles of intelligent tutoring, but also be based upon the science behind teamwork, team performance, and collaborative learning. Thus, the research described herein addresses this issue by providing a set of critical steps toward the development and validation of a design architecture of team process and performance modeling, which are specific to the needs of GIFT. In other words, the focus of this research is on assessing various aspects of teamwork as they contribute to team learning and performance.

Given the need for a comprehensive design architecture based on the science of team process and performance modeling that will inform the development of GIFT, we proposed two major objectives. The first produced a refined and detailed version of the proposed team models and the content of each. This was accomplished in two phases, the first being a synthesis of the existing science to inform the refinement of the team models and the second being the actual refinement and finalization of a team design architecture that is built upon the resulting best practices and principles derived from the initial review. The second objective was to use the revised set of models/design architecture to develop a set of behavioral markers and metrics that could be embedded within the GIFT to accurately assess teamwork processes and performance throughout tutoring. This report documents the resultant products of this effort.

2. A Guiding Framework for Tutoring Teams

As we began to think about the goals of this research and how outputs contributed to the different pieces of the GIFT architecture, we reviewed Sottolare et al. (2011) and examined the enhanced team-tutoring model that was presented in that article. This model took the initial team-state models introduced by Sottolare (2010) and mapped the interactions among distributed autonomous training systems to produce an example of the components that would be needed to maintain a common picture across the system with respect to team performance and learning (see Fig. 1).

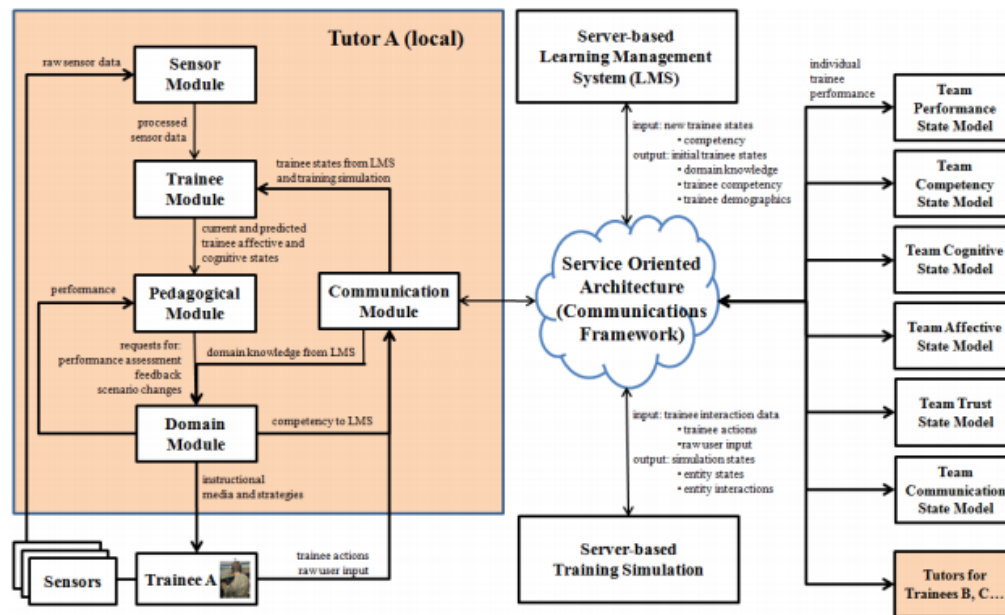


Fig. 1 Enhanced team-tutoring model in the GIFT (Sottolare et al. 2011, p. 6)

We used the framework in Fig. 1 as a starting point along with examining the broader literature on intelligent tutoring systems and began to think about what this research might contribute to the primary modules that comprise an ITS. This thought experiment led to the creation of an organizing framework for teams (Fig. 2) whereby we began to overlay what is known regarding the science of team training on top of portions of the enhanced team-tutoring model (Sottolare et al. 2011). We began with the fact that as with any training system, the initial drivers of any team-training effort are the learning objectives—in this case, team-learning objectives (referred to as team concepts in GIFT). These objectives can be focused on the development of taskwork or teamwork skills as both are tracks that need to be mastered in order for teams to perform effectively (Salas et al. 1992). The learning objectives then drive the content, which is pulled from the pedagogical and domain modules. Within the pedagogical module we would expect to see things

such as instructional strategies as well as some baseline scenarios/content. Within the domain module, we would expect to see specific tactics and measures to support learning assessments.

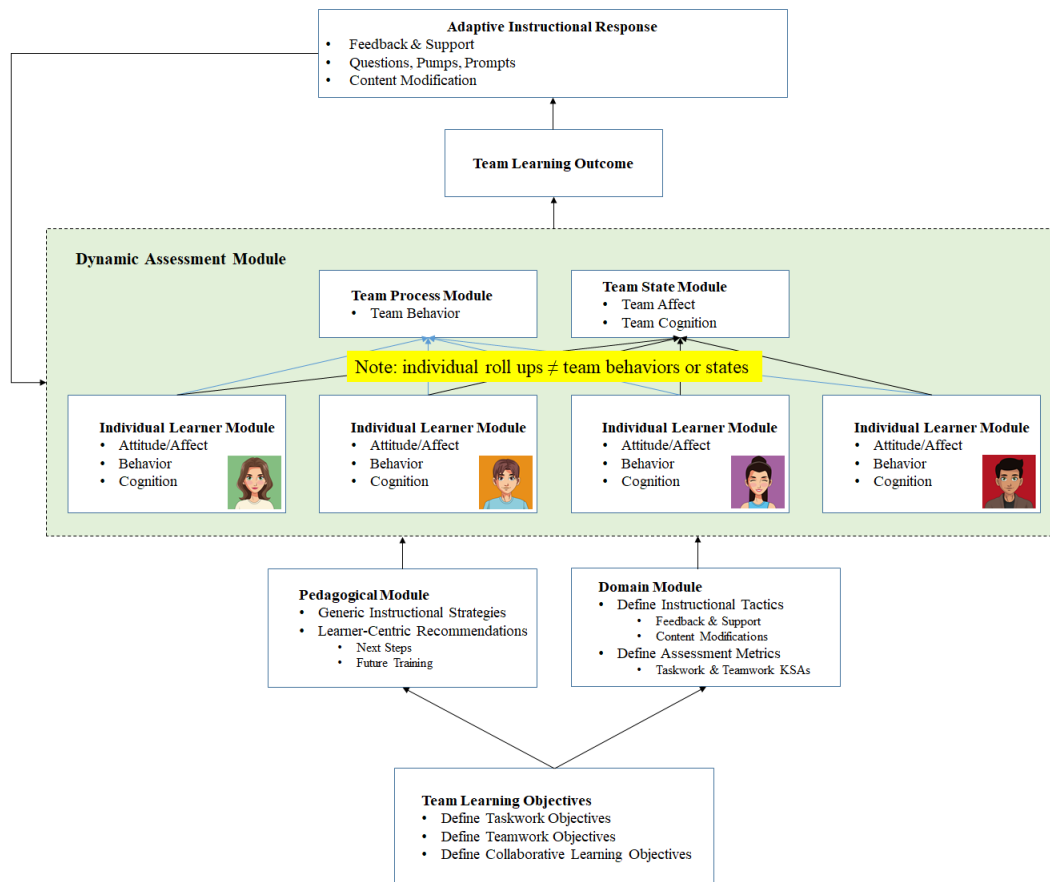


Fig. 2 Organizing a framework for thinking about team processes and instruction within an ITS

Potentially contained within the pedagogical module would be the guidance pertaining to an event-based approach to training as well as best practices regarding instructional strategies that have been shown to have efficacy for particular team knowledge, skills, abilities, and other characteristics (KSAOs). While important, the efforts on the current project did not focus on this aspect. However, some knowledge regarding the above was presented by members of the team at an Interservice/Industry Training, Simulation and Education Conference and later at a hosted workshop.

Another relevant focus area is the competency model (also referred to as an expert model or ideal student model). It is within this model we would expect to see taskwork, teamwork, and collaborative learning KSAOs. We expect some of the outcomes of our research will serve to inform this model through the identification

of those KSAOs that are related to key team outcomes of interest (i.e., team learning, team performance, team viability, and team satisfaction). These results could then be used to more fully expand the content of the team-state models initially proposed by Sottolare (2010). The pedagogical module, domain, and competency modules are particularly important in that they not only need to be able to communicate with each other so that relevant KSAOs are matched to scenarios and potential instructional strategies, but so this information feeds our dynamic assessment module (Fig. 2).

Information contained within the dynamic assessment module would pertain to particular measures, metrics, and methodologies that can be used to assess the KSAs that are being targeted in training. It is within this module that much of the work that has been produced as deliverables for the current project may be the most relevant. Within this module we propose that there are potentially three types of relevant models: 1) an individual state model that reflects the relevant individual attitudes, behaviors, and cognition, 2) a team-process model reflecting team behavior, and 3) a team-state model that reflects team affect and cognition.

Our goal was the systematic development of a set of behavioral markers for team constructs, guidance pertaining to the process of systematically developing such markers, a high-level plan for testing the viability of these markers, and guidance pertaining to the indexing and aggregation of team constructs. The results of the dynamic assessment were then tied to specific team-learning outcomes (which should map back onto the team-learning objectives). Portions of this research have highlighted team outcomes that are the most prevalent in team training and development as well as the particular team processes and emergent states that are related to key outcomes.

The last stage in this model is the feedback, which is based on the information obtained during the dynamic assessment as tied to team-learning outcomes. Some of the deliverables from this project are also relevant with respect to feedback; in particular, a best-practices paper on feedback delivery in teams that heavily emphasizes what is known regarding team debriefs.

3. Conceptualizing Team Effectiveness

There has been much work conducted to examine those factors that contribute to effective team outcomes. Past reviews have noted the variety of models and frameworks that have been proposed to delineate the factors impacting team outcomes. An examination of key models within the literature and areas of integration to devise a more practitioner-friendly framework formed the foundation of our initial review. Using this framework as a starting point, we examined the

team literature with respect to identifying those factors that contribute to effective teams.

The following sections put forth some guidance for facilitating each of the four specific team outcomes we investigated: team performance, team viability, team satisfaction, and team learning. These outcomes were chosen as they most often reflect team effectiveness within the conceptual models that appear in the literature. Team learning, while not always traditionally conceptualized as a team-effectiveness outcome, was included as the team's ability to learn and adapt, which is a hallmark of effectiveness.

3.1 Team Performance

In examining the literature on team performance, results indicated this is where the most effort is being put forth by researchers. Team performance in this sense refers to the evaluation of the outcomes of team processes relative to some set of criteria: a judgment of how well the results of teamwork meet some set of relatively objective or subjective standards. Results indicate that the following team behaviors explain significant amounts of variance in team performance: organizational citizenship behaviors (OCBs), communication, coordination, mutual support, reflexivity, conflict, and leadership. Team behaviors were found to have a positive relationship with team performance and explain small to moderate amounts of variance. The exception to this was conflict, which has a significant negative relationship. Team behaviors explained up to 25% of the variance in performance, with most behaviors explaining 10%–15%.

In addition to team behaviors, team attitudes were found to have a positive relationship with team performance and explain small to moderate amounts of variance. Attitudes (i.e., trust, collective efficacy, psychological safety, and cohesion) explained significant portions of variance in team performance ranging from 11% to 18%. Finally, team cognition was also found to explain significant amounts of variance in team performance. In this vein, team mental models and transactive memory systems were the most often studied, but situation awareness also emerged as a significant cognitive factor associated with team performance. These cognitive variables explained between 4% and 37% of the variance in team performance. Therefore, it is suggested that if one is interested in facilitating team performance, the following constructs should be trained and assessed:

- communication
- coordination
- reflexivity

- mutual support
- conflict
- coaching/leadership
- OCBs
- trust
- collective efficacy
- cohesion
- psychological safety
- team mental models
- transactive memory
- situation awareness

Emerging work also suggests the importance of action and transition processes (Marks et al. 2001), portions of which are captured in some of the prior constructs.

3.2 Team Learning

With respect to the second indicator of team effectiveness, team learning, our results suggest this is an area where further research is needed in order to have confidence in the initial trends coming out of the meta-analysis. Communication, coordination, reflexivity, conflict, and conflict management were the behaviors that our analysis indicated explain significant amounts of variance in team learning. While this set of behaviors accounted for between 6% and 52% of the variance in team learning, with conflict management accounting for the most, many of the results are based on two or fewer effects (i.e., reflexivity, conflict, or conflict management). However, there are fairly strong theoretical arguments that can be made for the constructs listed in Section 3.1 being significantly related to team learning. Coordination may be one of the weaker arguments in that team learning may itself facilitate team coordination.

Attitudes indicative of cooperation significantly contributed to explaining variance in team learning. For the most part these relationships were stronger than those seen for the behavioral constructs identified above, ranging from 24% to 48%. Trust, psychological safety, and cohesion were all significantly related to team learning. Unexpected was the finding that collective efficacy was not significantly related. However, this result was based on only one effect size so it should be interpreted

with extreme caution. No empirical studies on team cognition met the criteria for inclusion in the quantitative results. However, it might be argued that transactive memory systems might be linked to team learning as that system would alert team members where to look for key knowledge within the team as well as who is in need of specific new incoming knowledge. Based on the previously stated findings, those interested in facilitating team learning should train, assess, and provide feedback on communication, coordination, reflexivity, trust, psychological safety, and cohesion. Based on past theoretical work, the conflict, conflict management, and transactive memory systems might also need to be measured from an exploratory standpoint even though there were not enough studies in the current period for stable relationships to emerge.

3.3 Team Satisfaction

The last two team outcomes that were examined were more attitudinal in nature: team satisfaction and team viability. With respect to team satisfaction our review indicated that a few quantitative studies had examined this team outcome. Team behaviors explained between 11% and 41% of the variance in team satisfaction, with most falling in the 18%–25% range. Behaviors that explained significant amounts of variance in team satisfaction, and had a large enough sample size to be expected to be stable, were communication, coordination, mutual support, reflexivity, conflict (negative), leadership, and conflict management. Less studied with regard to team satisfaction were team attitudes indicative of cooperation and team cognition. While both team mental models and transactive memory systems were found to explain positive amounts of variance in team satisfaction—up to 69%—the number of effects the results are based on are small and thus should be interpreted with caution. Finally, the following constructs indicative of cooperation explained significant amounts of positive variance: trust, collective efficacy, and cohesion. Psychological safety also explained a significant 23% of the variance in team satisfaction but the result was, again, only based on one effect size and thus should not be relied upon at this point. Based on the overarching findings, we recommend the following to be trained and assessed if seeking to gauge team satisfaction: communication, coordination, mutual support, reflexivity, conflict (negative), leadership, conflict management, trust, collective efficacy, cohesion, and team mental models. Further empirical work is needed to investigate the roles of psychological safety and team cognition with respect to team satisfaction.

3.4 Team Viability

With respect to the final outcome, team viability, results indicated this was the least-studied team outcome. Team behaviors contributed between 4% and 20% of the variance in team viability indicating those variables that were examined contributed significantly. Specifically, the following behaviors were deemed important: communication, coordination, mutual support ($k = 1$, interpret with caution), and conflict. With respect to cooperative states the following were found to be important: collective efficacy, psychological safety, and cohesion. While all were found to explain significant amounts of variance (28%–62%), all were based on a small number of effect sizes. Therefore, the role of the specific cooperative components in team viability should be interpreted with caution; however, collapsing across cooperative constructs indicated that as a set they explained a significant 15% of the variance in team viability. In terms of cognition, transactive memory systems were the only cognitive construct examined and, while explaining a significant portion of variance (61%), it is based on only one effect size. In general, if seeking to facilitate team viability the following constructs are important: communication, coordination, conflict, and collective efficacy. Others that may be examined in an exploratory manner due to a small number of effect sizes currently existing are mutual support, conflict management, cohesion, and psychological safety.

4. Meta-Analysis of Effective Team Constructs

Given the need for a comprehensive design architecture based on the science of team process and performance modeling that will inform the development of the GIFT, we continued our research on effective team constructs. First, we compiled empirical work on teams to synthesize the literature. For this, we conducted meta-analyses on the relationships among team behaviors, attitudes, and cognition and team outcomes to produce a refined and detailed version of team models. Second, we investigated these processes and emergent states simultaneously to compare and contrast their relationship to each team outcome within the same framework. Specifically, we performed four meta-analytic structural equation modeling (MASEM) procedures to assess relationships to team performance, learning, satisfaction, and viability. More importantly, these analyses contributed to literature on teamwork by providing an overarching model for each team outcome. This serves as a theoretical and practical understanding of teamwork within complex, diverse contexts by providing a more accurate nomological network, identifying gaps within the literature, and allowing practical insight for highlighting the simultaneous importance of team constructs.

4.1 Results

As traditional meta-analyses were conducted, as well as MASEM, we report the results in the following manner. First, the results of the series of meta-analyses that examined the empirical link among specific team behaviors, attitudes, and cognition with team outcomes (e.g., performance, learning, satisfaction, and viability) are reported. Within this first set of meta-analyses each key construct was examined in isolation with respect to the reported team outcomes (i.e., constructs were not allowed to covary with one another). This was done to refine the initial set of team-state models proposed by Sottilare et al. (2011). Second, an ontology was built for each of those outcomes to assess the influence of these antecedents in a more integrated manner. This second method of analysis was conducted with hopes at providing insight into the relative contribution that each construct included in the set had in a particular team outcome (e.g., behaviors, attitudes, and cognitions were allowed to covary to see each one's relative contribution to each of the team outcomes). Findings indicated that the MASEM models all show adequate fit (see Table 1), but we expected such fit to become more robust with the inclusion of other team behaviors, attitudes, and cognitions. We will now summarize the findings for each of the outcomes in more detail.

Table 1 MASEM results

Ontology	χ^2	df	RMSEA (90% CI)	TLI (NNFI)	CFI	SRMR
Team performance ^a	4.306	3	0.019 (0.000; 0.056)	0.994	0.999	0.008
Team learning ^b	4.740	3	0.040 (0.000; 0.107)	0.975	0.996	0.019
Team satisfaction ^c	10.721	3	0.081 (0.033; 0.136)	0.905	0.981	0.031
Team viability ^d	2.080	1	0.056 (0.000; 0.166)	0.988	0.999	0.007

RMSEA: root mean square of approximation; CI: confidence interval; TLI: Tucker–Lewis index; NNFI: non-normed fit index; CFI: comparative fit index; SRMR: standardized root mean square residual

^aNumber (N) of individuals evaluated = 1173, ^bN = 350, ^cN = 396, ^dN = 346

4.2 Behaviors Examined

4.2.1 Team Performance

According to our meta-analytic findings, team behaviors explain up to 42% of the variance in team performance. Considering the importance of distinguishing specific behaviors, we highlight action processes and OCBs as the most important (see variance accounted in Table 2). A number of studies were found in this area of research, giving us the confidence regarding the moderate explanatory power of team behaviors to team performance (Table 2).

Table 2 Behavioral contributions to team performance

Outcome	K^a	N^b	Ro	Rho^c	Variance accounted
Communication	64	4183	0.31	0.36 ^d	13%
Coordination	49	3805	0.26	0.30 ^d	9%
Mutual support	11	881	0.35	0.40 ^d	16%
Reflexivity	9	721	0.33	0.38 ^d	14%
Monitoring	8	482	0.08	0.09	1%
Conflict	31	2062	-0.14	-0.17 ^d	3%
Transition processes	4	235	0.25	0.30 ^d	9%
Action processes	4	345	0.53	0.65 ^d	42%
Interpersonal processes	2	143	0.19	0.23 ^d	5%
Leadership	46	3839	0.29	0.33 ^d	11%
OCBs	4	313	0.46	0.55 ^d	30%
Conflict management	10	902	0.23	0.27 ^d	7%

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Similar to the degree of research that has examined the relationship between team behaviors and team performance, a wealth of research exists linking attitudes to team performance. The most common researched attitudes in relation to team performance were trust, collective efficacy, and cohesion. Results indicate that collective efficacy and psychological safety explain the most variance in team performance, 20% and 17% respectively (Table 3). This is followed by trust and cohesion, which also explain significant amounts of variance (9% and 15% respectively). Justice was also examined, but it only had a K of 1 and was not significantly related to team performance.

Table 3 Attitudinal contributions to team performance: effect of cooperation

Outcome	K^a	N^b	Ro	Rho^c	Variance accounted
Overall	98	7587	0.38	0.44 ^d	19%
Trust	23	1617	0.26	0.30 ^d	9%
Collective efficacy	27	2657	0.39	0.45 ^d	20%
Psychological safety	7	644	0.33	0.41 ^d	17%
Cohesion	23	1540	0.34	0.39 ^d	15%
Justice	1	170	0.08	0.09	1%

^a K is the number of effect studies analyzed for this outcome

^b N is the number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Team cognition also explained significant variance in team performance. Transactive memory systems and shared mental models accounted for 20% and 10% of the variance in team performance, respectively (Table 4). Surprisingly, but with a caveat, situation awareness was the one construct to show the largest relationship with performance. However, this conclusion is based on small sample size that calls for further investigation in order to strengthen confidence in the findings.

Table 4 Cognitive contributions to team performance

Outcome	K^a	N^b	Ro	Rho^c	Variance accounted
Overall	39	2984	0.32	0.38 ^d	14%
Team mental models	21	1478	0.26	0.32 ^d	10%
Transactive memory systems	14	1206	0.39	0.45 ^d	20%
Situation awareness	2	126	0.56	0.63 ^d	40%

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Table 5 shows the available intercorrelations across the team constructs that influence performance.

Table 5 Team-performance correlation matrix

	1	2	3	4	5	6	7
1. Communication							
<i>k</i>	...	...	...	...	...	...	...
N							
2. Coordination	0.46						
<i>k</i>	15	...	...	...	...	...	...
N	1123						
3. Conflict	-0.19	0.35					
<i>k</i>	10	3	...	...	...	...	...
N	627	262					
4. Leadership	0.31	0.45	0.10				
<i>k</i>	11	11	2	...	...	...	...
N	789	947	36				
5. Trust	0.79	0.03	-0.74	0.68			
<i>k</i>	5	6	1	6	...	...	...
N	186	556	67	409			
6. Collective efficacy	0.45	0.49	-0.43	0.43	0.73		
<i>k</i>	1	3	2	6	2	...	...
N	102	341	113	774	240		
7. Cohesion	0.64	0.52	-0.41	0.57	0.81	0.68	
<i>k</i>	5	8	6	5	1	7	...
N	472	719	415	324	36	382	
8. Performance	0.36	0.30	-0.17	0.33	0.30	0.45	0.39
<i>k</i>	64	49	31	46	23	27	23
N	4183	3805	2062	3839	1617	2657	1540

k = number of independent samples.

These results are compared with those of the MASEM, which allowed constructs to covary naturally. Findings from the MASEM indicated the model in Fig. 2 showed adequate fit (CFI = 0.999, RMSEA = 0.019; see Table 1).

Figure 3 shows the integrative model that highlights the importance of communication ($\beta = 0.10$), leadership ($\beta = 0.09$), collective efficacy ($\beta = 0.15$), and cohesion ($\beta = 0.08$). However, it also shows how most of the variance found in coordination, conflict, and trust influences performance through other mechanisms. Last but not least, the lack of intercorrelations present in the database does not allow for the inclusion of reflexivity, mutual support, OCBs, or any of the cognitive variables in the MASEM analysis.

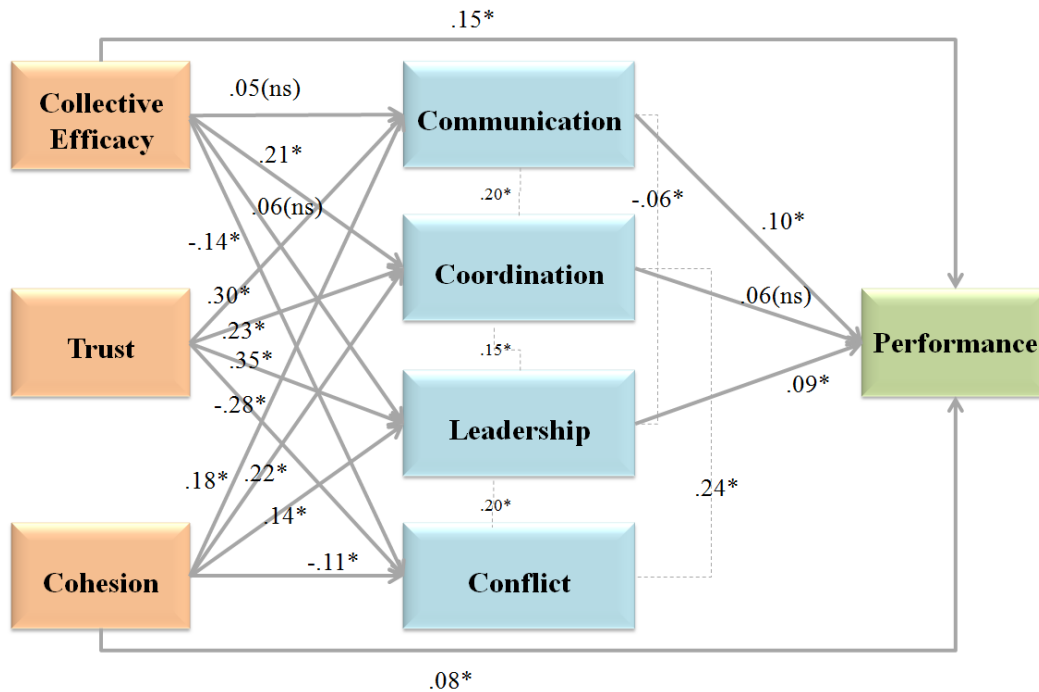


Fig. 3 Ontology for team performance based on meta-analytic structural equation modeling results

4.2.2 Team Learning

According to our meta-analytic results, team behaviors account for 7–36% of the variance in team learning (Table 6). Specifically, conflict and conflict management appear as important antecedents, but the amount of studies available in this area limits the interpretation of such finding. Communication and reflexivity appear as main antecedents accounting for 28% and 15% of the variance, respectively. Additionally, these variables have a higher number of included studies thereby producing more confidence in results. Earlier work also suggests coaching/leadership can play a significant role in learning (e.g., Hackman and Wageman 2005; Edmondson et al. 2001).

Table 6 Behavioral contributions to team learning

Outcome	K ^a	N ^b	Ro	Rho ^c	Variance accounted
Overall	6	505	0.51	0.60 ^d	36%
Communication	3	316	0.44	0.53 ^d	28%
Coordination	3	343	0.22	0.26 ^d	7%
Reflexivity	2	119	0.33	0.39 ^d	15%
Conflict	2	209	-0.37	-0.44 ^d	19%
Transition processes	1	78	0.19	0.23	5%
Action processes	1	78	0.27	0.33	11%
Conflict management	1	224	0.59	0.70	49%

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Not surprisingly, attitudes reflective of cooperation account for 37% of the variance in team learning (Table 7). The role of psychological safety (74%), cohesion (44%), and trust (27%) becomes evident, but the issue of number of studies remains. The findings regarding team learning show a promising avenue that calls for future research to strengthen the confidence in the findings. There were no identified studies examining the relationship between team cognition and team learning.

Table 7 Attitudinal contributions to team learning: effect of cooperation

Outcome	K ^a	N ^b	Ro	Rho ^c	Variance accounted
Overall	5	523	0.52	0.61 ^d	37%
Trust	4	307	0.45	0.52 ^d	27%
Collective efficacy	1	78	0.15	0.18	3%
Psychological safety	1	178	0.69	0.86	74%
Cohesion	1	78	0.56	0.66	44%
Justice	NA	NA	NA	NA	NA

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

A slightly different picture emerges with the MASEM results where all antecedents are allowed to covary naturally. The model below showed adequate fit (CFI = 0.996, RMSEA = 0.040; see Table 1). Figure 4 shows the integrative model that highlights the importance of trust ($\beta = 0.11$), cohesion ($\beta = 0.18$), conflict ($\beta =$

−0.12), and conflict management ($\beta = 0.27$). However, it also shows how most of the variance that team behaviors (e.g., coordination and communication) account for with respect to learning occurs through other mechanisms. This highlights the importance of emergent states as antecedents when the goal is to improve learning. Moreover, the lack of intercorrelations among the data in our data set limits the model in regards to the inclusion of reflexivity, psychological safety, leadership, or any of the cognition variables even though prior work has suggested these appear to be relevant.

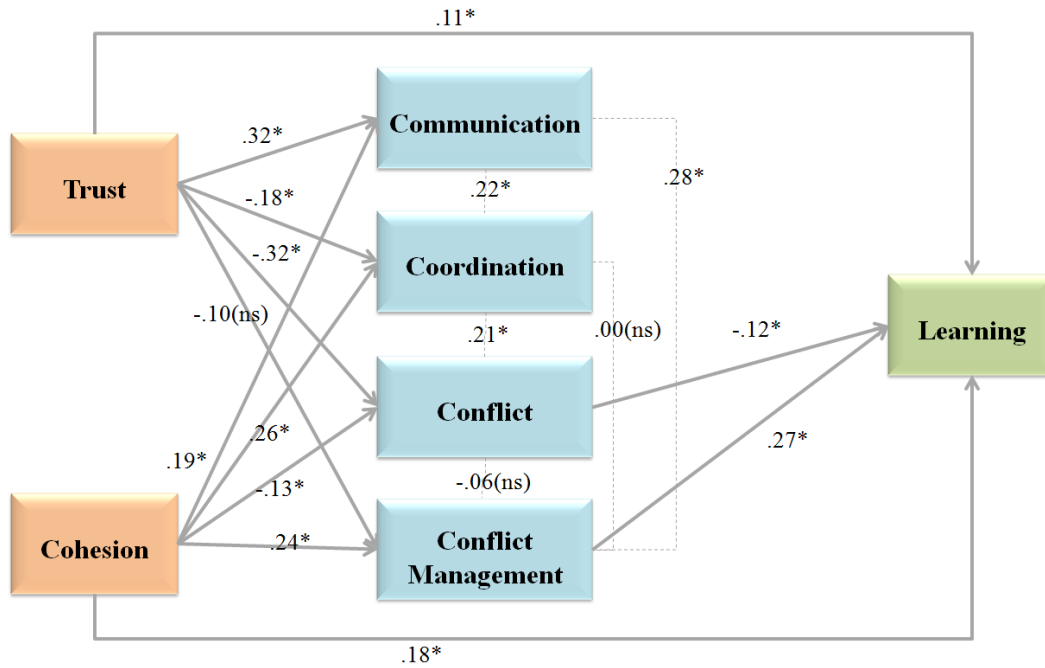


Fig. 4 Ontology for team learning based on meta-analytic structural equation modeling

4.2.3 Team Satisfaction

Results of our meta-analysis supported the frequency with which this relationship was examined as behaviors were most often examined as antecedents to team satisfaction. Team behaviors accounted for 44% of the variance, highlighting leadership, reflexivity, conflict, coordination, and conflict management (28%–42%; see Table 8). It is important to point out that different aspects of leadership account for different amounts of variance in satisfaction. Additionally, mutual support and conflict accounted for moderate amounts of variance.

Table 8 Behavioral contributions to team satisfaction

Outcome	K^a	N^b	Ro	Rho^c	Variance accounted
Overall	23	1336	0.56	0.66 ^d	44%
Communication	8	512	0.39	0.47 ^d	22%
Coordination	8	451	0.50	0.58 ^d	34%
Mutual support	2	106	0.33	0.38 ^d	14%
Reflexivity	3	234	0.54	0.62 ^d	38%
Conflict	6	351	−0.45	−0.53 ^d	28%
Leadership	4	346	0.57	0.65 ^d	42%
Conflict management	3	145	0.38	0.45 ^d	20%

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Less often examined were the role that attitudes or cognition may occupy in team satisfaction. In regard to attitudes, it was trust, collective efficacy, and cohesion that accounted for the most variance in team satisfaction (42%–77%; see Table 9). Even though those amounts of variance were significant, the number of studies were small.

Table 9 Attitudinal contributions to team satisfaction: effect of cooperation

Outcome	K^a	N^b	Ro	Rho^c	Variance accounted
Overall	14	825	0.70	0.82 ^d	67%
Trust	2	91	0.76	0.88 ^d	77%
Collective efficacy	3	162	0.51	0.59 ^d	35%
Psychological safety	1	48	0.36	0.44	19%
Cohesion	4	248	0.57	0.65 ^d	42%

^a K: number of effect studies analyzed for this outcome

^b N: number of individuals evaluated in this outcome

^c Rho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^d Statistically significant; confidence interval excludes zero

Similarly, when making conclusions about team cognition and team satisfaction, the 23% that team cognition as a group accounted for is limited to no situational awareness studies and one or two studies including either transactive memory and team mental model constructs (Table 10).

Table 10 Cognitive contributions to team satisfaction

Outcome	K ^a	N ^b	Ro	Rho ^c	Variance accounted
Overall	4	257	0.40	0.48 ^d	23%
Team mental models	2	105	0.55	0.68 ^d	46%
Transactive memory systems	1	64	0.66	0.78	61%
Situation awareness	NA	NA	NA	NA	NA

^aK: number of effect studies analyzed for this outcome

^bN: number of individuals evaluated in this outcome

^cRho is the corrected effect size, Ro is the uncorrected effect size; both weighted for sample size

^dStatistically significant; confidence interval excludes zero

Once again we compare these results with those obtained in conducting a MASEM. The integrative model put forth in the MASEM showed adequate fit (CFI = 0.981, RMSEA = 0.081; see Table 1). Figure 5 suggests that trust ($\beta = 0.30$), coordination ($\beta = 0.28$), leadership ($\beta = 0.17$), and conflict ($\beta = -0.21$) play a major role in the extent to which members enjoyed being a part of the team. However, it also highlights the lack of information regarding team cognition and attitudes beyond trust.

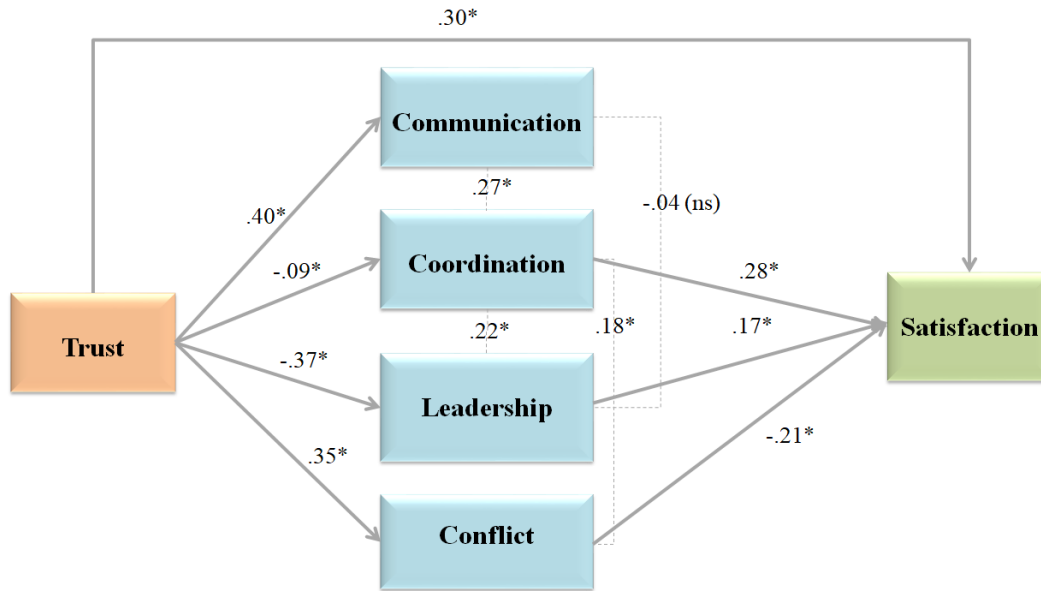


Fig. 5 Ontology for team satisfaction based on meta-analytic structural equation modeling

4.2.4 Team Viability

Surprisingly, team behaviors—as a group—account for little variance in team viability. The construct that accounted for the most variance (i.e., 19%) is conflict, followed by mutual support and coordination. Attitudes, on the other hand, often

showed moderate to high explanatory variance of team viability, ranging from 29% to 72%. Collective efficacy was the most important attitude, followed by cohesion. In regard to cognition, transactive memory systems were the only variable linked to team viability in the literature. When considering this team outcome, almost every effect had to be looked at with the caveat of very low samples, mostly relying on one study for each relationship. Therefore, these results should be viewed with extreme caution and point to a need in the research.

Conducting the MASEM on team viability was a bit challenging due to the few studies identified. The model that was able to be tested showed adequate fit (CFI = 0.999, RMSEA = 0.056; see Table 1). Figure 6 illustrates the results. Collective efficacy can be seen to explain large amounts of variance in team viability ($\beta = 0.81$), followed by cohesion ($\beta = 0.64$). Team behaviors showed important but contradictory effects. Specifically, communication ($\beta = -0.44$), coordination ($\beta = -0.25$), and conflict ($\beta = 0.18$) showed a moderate relationship, but with the inverse sign. This can indicate that most of the variance from these team behaviors is being included in attitudinal constructs and unique variance may actually be detrimental. Even though this is a fair possibility, the small number of studies in each of those links warrants caution.

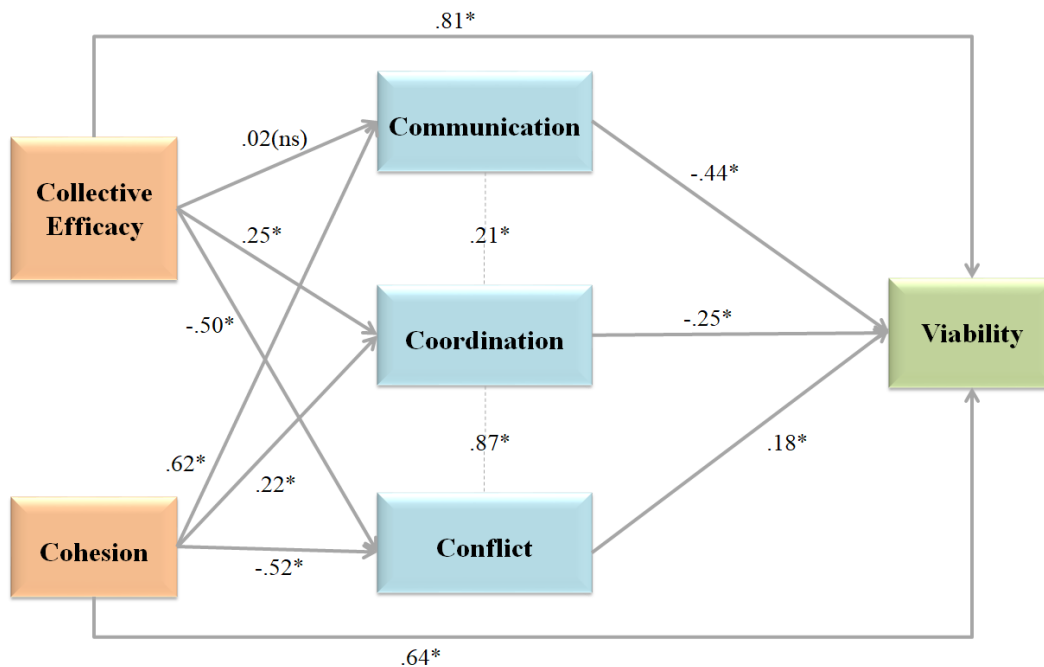


Fig. 6 Ontology for team viability based on meta-analytic structural equation modeling

5. Effective Team Constructs Revisited

One of the primary goals of this project was to begin to further refine the initial state models (i.e., team performance, team competency, team cognitive, team affective, team trust, and team communication) proposed by Sottolare et al. (2011) based on the scientific and qualitative literature. Thereby, one of the initial steps in this project was to examine the initial definitions put forth and map those against definitions within the team literature. The work within this section primarily reflects work conducted within the first and second years of the project. This information is perhaps most relevant to the competency module and the assessment module.

This initial first step was conducted to facilitate a shared mental model regarding the type of content that would be examined within the various models and, in some cases, serve to further expand the definition based on common construct boundaries within the broader team literature. This initial step also facilitated the team's ability to understand how the results of the meta-analytic work to follow (investigating the relationship between key team constructs and team outcomes) would map back onto the team-state models and thereby serving to provide more depth to the information that was currently contained in those models.

In conducting our meta-analytic efforts, we expanded performance beyond what was originally depicted in the Sottolare et al. (2011) team performance-state model to include other team outcomes that are often examined as a portion of team performance or effectiveness. Specifically, we examined key team constructs in relation to the team outcomes of team performance, team learning, team viability, and team satisfaction. Therefore, in talking about the state models we have four sets of results, one for each team outcome of interest. Within each of the four models, we further delineate team affective, behavioral, and cognitive state models.

5.1 Team Performance

The revised state models for team performance appear in Fig. 7. With respect to team performance, the results of the meta-analysis and MASEM suggest the team-affect state model should include such constructs as trust, collective efficacy, cohesion, and psychological safety. The team-behavioral state model contains the following constructs: communication, coordination, reflexivity, mutual support, conflict, leadership, and OCBs. Finally, the cognitive-state model contains team mental models, transactive memory systems, and situation awareness.

In interpreting the information in Fig. 7, items with a "1" superscript at the end indicate constructs that showed significant direct linkages to team performance in

both the traditional meta-analysis as well as the MASEM (i.e., collective efficacy, cohesion, communication, and leadership). Conversely, items with a “2” superscript were shown to be significantly related to team performance in both the traditional and MASEM; however, their relationship to team performance in the MASEM was indirect (i.e., trust and conflict). Psychological safety, coordination, reflexivity, mutual support, OCBs, team mental models, transactive memory systems, and situation awareness were significantly related to team performance in the traditional meta-analysis, but did not have significant relationships to team performance in the MASEM. In the case of psychological safety, reflexivity, mutual support, OCBs, and all of the identified cognitive constructs, their absence in the MASEM was because they dropped out of the ontology due to a lack of studies examining their relationship with the remaining constructs in the ontology.

Team Affect State Models	Team Behavior State Models	Team Cognitive State Models
• Trust² • Collective efficacy¹ • Cohesion¹ • Psychological safety	• Communication¹ • Coordination • Reflexivity • Mutual support • Conflict² • Leadership¹ • OCBs	• Team mental models • Transactive memory systems • Situation awareness (?)

Fig. 7 Team performance-state models that should be a primary assessment when team performance is the targeted outcome

5.2 Team Learning

The revised state models for team learning appear in Fig. 8. With respect to team learning, the results of the meta-analysis and MASEM suggest the team-affect state model should include such constructs as trust, conflict management, cohesion, and psychological safety. The team behavioral-state model contains the following constructs: communication, coordination, reflexivity, conflict, and (as suggested by prior work) leadership. Finally, the relationship between team learning and shared cognition (i.e., shared mental models, transactive memory systems, and shared situation awareness) was not explicitly examined within the time frame covered by the meta-analysis, so we have not identified specific cognitive constructs within the cognitive-state model. This is an area where future research is needed.

As with team performance, those constructs in Fig. 8 that have “1” superscripts indicate constructs that showed significant direct linkages to team learning in both the traditional meta-analysis as well as the MASEM (i.e., trust, conflict management, cohesion, and conflict). Conversely, items with a “2” superscript were shown to be significantly related to team learning in both the traditional and

MASEM; however, their relationship to team performance in the MASEM was indirect (i.e., communication and coordination). Further, psychological safety, reflexivity, and leadership were significantly related to team learning in the traditional meta-analysis but did not have significant relationships to team learning in the MASEM. In all cases their absence in the MASEM was because they dropped out of the ontology due to a lack of studies that examined their relationship with the remaining constructs in the ontology. Finally, question marks appear beside three of the constructs (conflict management, cohesion, and psychological safety) in the team affect-state model due to the fact that, while results indicated significant relationships with team learning, the finding was based on a small K (number of studies). Therefore, this indicates additional areas where research should be conducted to strengthen confidence in the findings.

Team Affect State Models	Team Behavior State Models	Team Cognitive State Models
• Trust¹ • Conflict management¹ (?) • Cohesion¹ (?) • Psychological safety (?)	• Communication² • Coordination² • Reflexivity • Conflict¹ • Leadership (prior work would suggest)	• ?????

Fig. 8 Team learning-state models that should be a primary assessment when team learning is the targeted outcome

5.3 Team Satisfaction

The revised state models for team satisfaction appear in Fig. 9. With respect to team satisfaction, the results of the meta-analysis and MASEM suggest the team affect-state model should include such constructs as collective efficacy, cohesion, trust, and psychological safety. The team behavioral-state model contains the following constructs: communication, coordination, reflexivity, mutual support, conflict, conflict management, and leadership. Finally, the cognitive-state model has team mental models and transactive memory systems. The lack of shared situation awareness appearing within the cognitive state model is due to a lack of studies found within the time frame covered by the meta-analysis that examined this cognitive construct in relation to team satisfaction.

As with the state models discussed earlier in the paper, those constructs in Fig. 9 that contain a “1” superscript indicate constructs that showed significant direct linkages to team satisfaction in both the traditional meta-analysis as well as the MASEM (i.e., trust, coordination, conflict, and leadership). Conversely, the item

with a “2” superscript was shown to be significantly related to team satisfaction in both the traditional and MASEM; however, its relationship to team performance in the MASEM was indirect (i.e., communication). Further, collective efficacy, cohesion, psychological safety, reflexivity, mutual support, conflict management, team mental models, and transactive memory systems were significantly related to team satisfaction in the traditional meta-analysis but did not have significant relationships to team satisfaction in the MASEM. In all cases their absence in the MASEM was because they dropped out of the ontology due to a lack of studies that examined their relationship with the remaining constructs in the ontology.

Finally, question marks appear beside two of the constructs (trust and safety) in the team affect-state model; this is due to the fact that while results indicated significant relationships with team satisfaction, the finding was based on a small K. Therefore, this indicates additional areas where research should be conducted to strengthen confidence in the findings.

Team Affect State Models	Team Behavioral State Models	Team Cognitive State Models
• Collective Efficacy • Cohesion • Trust¹ (?) • Psychological Safety (?)	• Communication² • Coordination¹ • Reflexivity • Mutual Support • Conflict¹ • Leadership¹ • Conflict Management	• Team Mental Models (?) • Transactive Memory Systems (?)

Fig. 9 Team satisfaction-state models that should be a primary assessment when team satisfaction is the targeted outcome

5.4 Team Viability

The revised state models for team viability appear in Fig. 10. With respect to team viability, the results of the meta-analysis and MASEM would suggest that the team affect-state model should include such constructs as collective efficacy, cohesion, and psychological safety. The team behavioral-state model contains conflict and coordination. Finally, the cognitive-state model includes transactive memory systems. Within the time frame covered by the meta-analysis, team viability was one of the least researched outcomes of those which were examined. Therefore, many of the affective, behavioral, and cognitive variables that appear in the state models for the other team outcomes (team performance, team learning, and team satisfaction) were not examined with respect to team viability. In interpreting the

findings, we suggest not viewing the absence of a particular construct as evidence of its nonimportance, but more in light of areas of research that need to be further investigated.

As with the state models discussed earlier in this report, those constructs in Fig. 10 that have a “1” superscript indicate constructs that showed significant direct linkages to team viability in both the traditional meta-analysis as well as the MASEM (i.e., collective efficacy, cohesion, conflict, coordination, and communication). All of the constructs in Figure 10 were significantly related to team viability in the traditional meta-analysis, but did not have significant relationships to team learning in the MASEM. In all cases their absence in the MASEM was because they dropped out of the ontology due to a lack of studies that examined their relationship with the remaining constructs in the ontology. With respect to team viability, even those constructs included in the state models are based on a small K (number of studies). Therefore, this indicates additional areas where research should be conducted to strengthen confidence in the findings.

Team Affect State Models	Team Behavioral State Models	Team Cognitive State Models
• Collective Efficacy¹ • Cohesion¹ • Psychological Safety	• Conflict¹ • Communication¹ • Coordination¹	• Transactive Memory Systems

Fig. 10 Team viability-state models that should be a primary assessment when team viability is the targeted outcome

5.5 Discussion

This work reflects the initial steps toward determining those constructs that would be important to include within an intelligent team tutor. Some of the constructs identified herein (i.e., leadership, conflict, collective efficacy, cohesion, and trust) have been shown to be important across different team outcomes. Similar to previous meta-analytic findings that task-focused leadership and person-focused behaviors were related to team effectiveness and team productivity (Burke et al. 2006), we found leadership to be positively related to performance. Furthermore, our results also show evidence for assertions such that when team members evaluate their team leaders positively, they are both more cohesive and satisfied (Gomes et al. 2011; Wolfram and Mohr 2009). Furthermore, results indicated that leadership behaviors and leadership styles represent the most commonly examined aspects of leadership, leaving future research to scrutinize the relationship regarding shared or collective leadership (see Wang et al. [2014] for an exception).

With respect to other constructs (e.g., conflict) findings differ depending on the specific dimension one is examining. Conflict, for instance, is often broken down into task and relationship conflicts. Task conflict refers to disagreement regarding ways to approach the job, whereas relationship conflict is more related to interpersonal tension. While task conflict has been shown to be less detrimental than relationship conflict (de Wit et al. 2012), both types of conflict showed a strong negative correlation to satisfaction (De Dreu and Weingart 2003). Turning conflict into something positive has been a constant struggle within the team-diversity literature (e.g., van Jaarsveldt and Joubert 2015). It has emerged as part of team development that influences team learning (Raes et al. 2015) but also makes team members more dissatisfied (Cox 2003). Consequently, the emergence of conflict is likely to be detrimental to team learning as it is often associated with process loss. Our results suggest that conflict has a moderate negative relationship with team learning and team satisfaction. However, given that team-learning findings only represented one study, we further break down the relationship between conflict and team satisfaction. Consistent with previous research, task and relationship conflicts showed moderate negative relationships with team satisfaction.

The most influential inputs, however, were attitudinal-emergent states when considering all outcomes, likely to be due to the affective nature of team satisfaction and viability constructs. Collective efficacy (CE) and trust, for instance, showed to play an important role to a number of team outcomes. CE is the idea that team members' beliefs of their team's capabilities converge and thus can be conceptualized at the team level (Kozlowski and Klein 2000). Positive beliefs about members' capability to perform effectively leads to better performance (Gully et al. 2002; Jung and Sosik 2003) in both lab and field settings (Katz-Navon and Erez 2005; Scott-Young and Samson 2009; Stajkovic et al. 2009). This was also true in our meta-analysis, in addition to the relationship of CE and team viability. Even just the belief that the team can manage conflict can increase its viability (Jehn et al. 2008). Considering the definition of viability to be the desire to remain a part of the same team in the future, this is highly related to whether the team members perceive their team as strong and with the potential for being successful. On a similar note, trust—or the willingness to be vulnerable to another party (see Mayer et al. 1995)—shows a strong positive relationship to team learning, explaining 42% of the variance in team learning. Trust has also been more often empirically linked to satisfaction within the team. Specifically, higher levels of trust translate to higher satisfaction (Costa 2003; Gladstein 1984). Within the team, trust can lead to satisfaction between dyads, or pairs of individuals who share interdependent goals (Smith and Barclay 1997). Accordingly, our meta-analytic results support these claims.

Last but not least, team cohesion, which has been investigated as an antecedent to team outcomes for over 60 years, shows strong evidence base both within the military and broader literature. Several meta-analyses emerged to support the cohesion–performance link (e.g., Beal et al. 2003; Evans and Dion 1991; Gully et al. 1995; Oliver 1999). Our meta-analytic results support prior findings, and team cohesion explains a significant 14% of the variance and is positively related to team performance, team learning, and team viability. While these findings for team viability are not based on a wealth of research, cohesion has shown to significantly increase team viability (Barrick et al. 1998). Cohesion is considered one of the main antecedents of team effectiveness (Carron and Brawley 2000); thus, these positive outcomes of cohesion come without surprise.

Furthermore, we encourage researchers to begin identifying behavioral markers for those constructs that explain the most variance in key outcomes. Specifically, while initial team models have been theorized, further work is needed to identify a complete design architecture, including behavioral markers and metrics that can be used to model team processes and performance in the GIFT. This design architecture must be rooted not only in principles of intelligent tutoring but also be based upon the science behind teamwork and team performance. Given the need for a comprehensive design architecture based on team science that will inform the development of the GIFT, we propose refining team models and the content of each to guide the most important constructs. A way to assess these constructs, especially attitudinal and cognitive, through behavioral markers can be groundbreaking toward a more accurate and timely aid in team outcomes throughout tutoring.

6. Dynamic Assessment

While there are many considerations in the assessment of team constructs and corresponding performance, within this section we highlight two areas that the work from the project can offer insight and provide information relevant to the dynamic assessment module within the GIFT framework. First is in the area of behavioral markers. Traditionally, the assessment of teams has relied heavily on self-report measures and, to a lesser extent, observer ratings. Within the last several years, even outside the ITS context, there has been an increasing focus on what can be done to move away from the heavy reliance on the self-report measure toward measures that might be less obtrusive—behavioral markers are one approach—and move toward unobtrusive (and in some cases system captured) measures. Therefore, we sought to describe an approach that could be used to systematically develop behavioral markers of team constructs along with the results of following such a procedure for several key team constructs. Our goal is to develop a set of

behavioral markers that can be used for an intelligent team-tutoring system. ITSs use artificial intelligence to provide directed feedback during training situations.

Traditionally, the focus of intelligent tutoring systems has been on the development of an individual's cognitive skills. Yet, this dismisses the potential to leverage these types of systems to create opportunities for social learning in team environments (Singley et al. 1999). Thus, the current objectives were focused on developing behavior markers specifically in the team context. In order to do this, we first used a refined team model that was produced in collaboration with the US Army Research Laboratory (ARL). Specifically, we synthesized the current state of the team literature and, using GIFT (Sottilare et al. 2011) as the basis, created a contemporary team-model architecture. This design architecture was then used to identify crucial team states that would be the foundation of the behavioral markers.

6.1 Methodology

In order to develop the behavioral markers, we had to first identify critical team constructs that were indicators of effective team performance. Incorporating the GIFT framework (Sottilare et al. 2011), we articulated several team constructs that we believe are predictors (see Table 11) of team outcomes (e.g., performance, learning, viability, and satisfaction). Once these constructs were identified, we searched the literature for measures that have been used to assess these constructs. Next, we compiled the items from these measures into separate Excel documents. These Excel documents were then given to several subject-matter experts (SMEs) with the goal of 1) developing behavioral markers that represented each individual item, 2) identifying and removing items that measured the same behavior, and 3) developing subdimensions of the construct as a whole if it was necessary.

Table 11 Key team constructs based on meta-analytic results

1	Behaviors	Cognitions
Trust	Communication	Team mental models
Collective efficacy	Coordination	Transactive memory systems
Psychological safety	Conflict	Situation awareness
Cohesion	Leadership	
Justice	Transition processes	
	Action processes	
	Interpersonal processes	
	Conflict management	
	Organizational citizenship behaviors	

Once the SMEs completed their ratings, the group met and came to a consensus on whether the generated markers accurately represented the item. During this meeting, SMEs determined whether the construct and its subdimensions were

equally represented by the markers. If the group believed the current markers were deficient in representing any aspect of the construct, new markers were then generated and agreed upon. This was the final step in generating the final list of behavioral markers by construct. A short version of these steps is displayed in Table 12.

Table 12 Overview of process used to identify behavioral markers

Step	Action
1.	Identify crucial team construct related to team outcomes
2.	Search the literature for measures of these constructs
3.	Compile items from for each team construct
4.	Independently review Items • Create markers for each item • Identify duplicates • Come up with subdimensions
5.	Meet as a group to come to an agreement on markers and subdimensions
6.	Identify areas of subdimensions that may have a deficient markers
7.	Generate new makers to fill in deficiency

6.2 Marker Development

Using the process described in Section 6.1, we have generated a list of potential behavioral markers for the following constructs: team psychological safety, trust, collective efficacy, cohesion, conflict, and team communication. While we did not develop markers for the team cognition constructs, we refer the reader to Salas et al. (2007b), who did develop a set of markers for shared mental models and situation assessment. Next, we present each set of potential markers for the above mentioned constructs as well as providing a brief definition of the construct, along with those sources that formed the basis of marker development for each construct. We begin first with those constructs that reflect emergent states within the team (i.e., psychological safety, trust, collective efficacy, and cohesion), followed by those constructs that are more process focused (i.e., conflict and team communication).

6.2.1 Team Psychological Safety

Psychological safety is defined as “a shared belief that the team is safe for interpersonal risk taking” (Edmondson 1999, p. 354). It reflects a sense of confidence among team members that members will not be embarrassed or punished/rejected for speaking up (Edmondson 1999), thereby promoting a work

climate that is safe for interpersonal risk taking. It goes beyond interpersonal trust to include a work climate characterized by mutual respect where people are comfortable expressing their differences (Walumbwa and Schaubroeck 2009). The markers developed were created based on items that appear in papers on psychological safety published by Choo et al. (2007), Edmondson (1999), Kostopoulos and Bozionelos (2011), Ortega et al. (2010), and Schaubroeck et al. (2011).

Potential markers of psychological safety are the following:

- The number of times new ideas are encouraged, brought up, discussed openly, not actively ignored, or not offensively criticized.
- The number of time different/opposing ideas are encouraged, brought up, discussed openly, not actively ignored, or not offensively criticized.
- The number of times diversity (e.g., expertise) among team members is recognized, positively reinforced, or accepted.
- The number of times team members engage in inappropriate criticism or feedback (i.e., bashing team members' competence behind their back or incorporating interpersonal remarks during task feedback).
- Ratio of problems/number of times the problems are talked about.

6.2.2 Trust Markers

While there is no universally agreed upon definition of trust, two elements that are common to nearly all definitions reflect the idea of positive expectations and the willingness to become vulnerable. Perhaps one of the most cited definitions argues that trust is “the willingness of another party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” (Mayer et al. 1995, p. 712).

In this vein, we next present a number of trust markers. More specifically, it is a mixture of markers some of which describe the trustee's behavior (the target) and the trustor (person trusting). The former is typically indicative of trustworthiness and the later trust. Items here generally fall into markers that reflect assessments of competence, benevolence, or integrity, some of which are surface level and some of which are a deeper level (only appearing after a time). The markers were created based on material in papers published by Chang et al. (2012), Cook and Wall (1980), De Jong and Elfring (2010), and Dirks (2000).

The following are potential markers of trust or distrust:

- The number of times team members are reminded to follow through on their tasks (reverse coded).
- The number of times team members accept backup behaviors from their fellow teammates.
- The number of times “other” team members actively seek the opinions of others regarding work tasks.
- The amount of task information withheld from fellow team members.
- The number of times team members delegate task responsibilities, assign additional tasks, or ask the advice of their fellow team members.
- Use of *honest language* (e.g., “I’m just being honest with you”) in discussions.
- The number of times team members double-check their fellow team members’ work or call their competence into question in a public forum (i.e., with their fellow team members).
- The number of times team members point out their fears requiring the quality of information, competence, or other potentially negative information among their fellow team members. (e.g., questioning team members).

The following are potential markers of trustworthiness:

- Providing negative feedback in an inappropriate way (e.g., involving interpersonal insults) or at an inappropriate time (e.g., in front of other team members).
- The number of times team members complete task responsibilities in a timely/assigned manner.
- The number of times team members engage in inappropriate interpersonal behaviors (e.g., cutting people off, telling them to “Shut up,” or name-calling) during any part of the task (reverse coded).
- The number of times either good verbal or physical backup behaviors are provided.
- The number of times team members purposefully provide bad information (reverse coded).

6.2.3 Collective Efficacy

Collective efficacy has been defined as “a shared belief in a collective’s capabilities to organize and execute the course of action” (Bandura 1997, p. 477). In essence, a team experiences a sense of collective efficacy when it perceives the team members have the strong abilities to fulfill their roles on the team as they relate to the task in question. Interpersonal factors can moderate the relationship between perceived abilities and task perception such that when interpersonal factors are diminished (e.g., personal conflict is high), teams may feel they have less collective efficacy not due to task-related capabilities but interpersonal factors that impact task execution. Scales on collective efficacy tend to be task specific and based on the task, incorporate explicitly task-driven factors, or may also include some of the interpersonal factors (given that both taskwork and teamwork are needed for a team to successfully perform). Therefore, the markers that appear in Table 12 reflect this duality.

Table 13 is based on material in papers on collective efficacy published by Lent et al. (2006), Jehn et al. (1997), Edmonds et al. (2009), Tasa et al. (2007), Cronin et al. (2011), Luhtanen and Crocker (1992), Wong et al. (2009), Jones (1986), Matheiu et al. (2009), Woehr et al. (2013), Shivers-Blackwell (2004), Sargent and Sue-Chan (2001), Guzzo et al. (1993), Bray (2004), Chen et al. (2005), Riggs and Knight (1994), Hsu et al. (2007), and Cheng and Yang (2011). Within this set of markers, we have the most confidence in the ability-focused markers.

Table 13 Potential markers of collective efficacy

Ability focused
Team members actively request backup when needed.
Team members acknowledge input from teammates in taskwork and incorporate input into team task.
Team members acknowledge fellow teammates’ inputs and suggestions.
Team communication portrayed positive messages, such as reinforcing team ability to complete task, avoid negative or anxiety of failure, and so on.
Team members request assistance from or direct questions to a team member rather than an external member or external instructor.
Team members express confidence by acknowledging each team member’s strengths, use phrases to indicate team ability such as “We can do it” and “Go, team”, and do not express doubts of team success.
Goals are framed as achievable, despite challenges.
Team members discuss and develop a plan/course of action in a timely fashion.
Team members display confidence in completing challenging goals, do not indicate doubt in success, and so on.
Team members think out of the box to develop solutions to problems even in the face of time constraints or challenging goals (i.e., task difficulty).
Reliance on one’s self to complete goals that may require other teammates (reverse coded).

Table 13 Potential markers of collective efficacy (continued)

Ability focused
Disengaging from group or team by not communicating, not requesting backup behavior, and not accepting input from team members (reverse coded).
Team members indicate they are not sure about successfully meeting team/task goals (reverse coded).
Team members do not request backup or support when needed (reverse coded).
Team members withhold information, or say they do not believe their input will be incorporate when sharing (reverse coded)
Team members do not incorporate suggestions from other team members in team planning (reverse coded).
Team members develop solutions to problems even in the face of time constraints or challenging goals (i.e., task difficulty).
Team members commit to task and do not procrastinate.
Members express preference in belonging to the team and in the team's ability.
Contextualized
Express confidence and willingness to resolve disagreements.
Team members cooperate with one another and make effort to avoid conflict.
Team members expressed confidence that they could efficiently resolve conflicts regarding opinions, preferences, differences, and so on, about how to complete the task and begin working toward this goal.
Team members expressed confidence that they could efficiently resolve conflicts regarding which tasks to prioritize and complete the task and begin working toward this goal.
Manage interpersonal and/or task conflict effectively while prioritizing team tasks (discussion of task and execution of team task-related actions).
Team members discuss and replan due to task and/or goal changes.
Team members prioritize team needs in planning team tasks and doing actions to execute team tasks without prompting.
Demonstrate clear concise communication with all team members.
Team members demonstrated open communication with all team members despite background.
Team members efficiently discuss alternative plans and solutions and resolve conflicts in planning tasks and assigning roles.
Team members avoid contact with teammates, are not involved in team communication/interaction.
Team members are receptive and open to feedback regarding performance.
Team members strive to contact teammates, make an effort to be involved in team communication/interaction.

6.2.4 Cohesion

While cohesion has been defined in many way, it has generally been argued to reflect an attraction or bond within the team. More specifically, cohesion has been defined as “the bonding together of members of a unit in such a way as to sustain their will and commitment to each other, their unit, and the mission” (Johns et al. 1984, p. 4). Carless and DePaola (2000) conceptualize cohesion as having three dimensions: social, task, and group pride; as such, we have examined the existing scales and papers that have reported to have cohesion measures in them and delineated a first round of markers that tap each of the three dimensions. In this vein, task cohesion is defined as the group’s shared commitment and ability to execute the group task or goal or the group’s capacity for teamwork (Siebold 1999; Craig and Kelly 1999). Social cohesion reflects the group’s bond that promotes development and maintenance of social relationships (Carless and DePaola 2000). Finally, group pride can be described as a shared sense of unity and honor derived from membership in the group and the group’s accomplishments.

The markers that follow were created based on items that appear in the following published papers on team cohesion: Carless and DePaola (2000), Carron et al. (1985), Hoegl and Gemuenden (2001), Chang and Bordia (2001), Henry et al. (1999), McCroskey and McCain (1974), Hobman and Bordia (2006), O’Reilly et al. (1989), Wong (2004), Podsakoff and Mackenzie (1994), Sargent and Sue-Chan (2001), Shin and Choi (2010), Shivers-Blackwell (2004), Solansky (2011), Watson et al. (1991), Zaccaro (1991), and Jehn (1995). Table 14 lists potential behavioral markers of team cohesion.

Table 14 Potential markers of team cohesion

Task cohesion
Members are actively working together and pitching in to reach team goals.
Ratio of push to pull of information within the team.
Member task engagement (e.g., immersed in the team task as evidenced through posture, verbal intonation, and behavior).
Members pick up the slack for one another.
Proactive offering of backup behavior when needed.
Team members are performing tasks that help the team reach its goal.
Motivational drive and persistence to reach targeted goal even when circumstances are difficult (stressors present).
Frequency of task-relevant communication.
Valuing member contributions; seeking input from all members.
Contributing to discussions about new course of action and/or problem solving.
Owning up to mistakes when they occur.
Negative comments about the speed of work of a team member.
Individual member workload taken into account when assigning tasks.
Team members own up to their actions and responsibilities within the team (verbal acknowledgement).
Frequency of productive meetings.
Willingness to drive toward a plan despite challenges (could also manifest itself in length of meetings, involvement when meetings go long).
Team members are willing to compromise.
• Negotiate taskload; • Adapt to meet goals; and • Considerate of individual differences within the team (e.g., culture, race, and preference)
Request for backup; phrases such as “I’ve got your back”, “You can count on me”, and “You do XXX, and I’ll do XXX”.
Phrases like “We make a great team”, acknowledging each other’s strengths and how they complement weaknesses in the group at large.
Phrases like “Great job, everyone”, “Go, team”, “You’re the best”, “Good work”; positive affirmations toward the team’s work.
Conversations being dominated by a single individual (reverse coded).
Input only sought from a select few; not including everyone in discussions (reverse coded).
Social loafing (reverse coded).
Use of negative tone and affect in describing the task and importance of the mission (reverse coded).
Frequency of task-irrelevant communication (reverse coded).
Team discusses different goals, different words are used to discuss the mission—teams do not have a shared mental model (reverse coded).
Individuals are working on tasks that do not contribute to team goals (reverse coded).
Team members consistently try to work independently on interdependent tasks (reverse coded).
Team members do not request backup when needed from fellow team members (reverse coded).

Table 14 Potential markers of team cohesion (continued)

Social cohesion
Team members rarely socialize outside the work context.
Number of social events organized by the team outside work.
Number of social events voluntarily attended.
Team members rarely share personal stories.
Introducing and/or mentioning each other's names (first or last, or rank as appropriate).
Willingness to speak up for a team member outside the team-task environment.
Members talk about common bonding experiences.
Communicating general interest in personal well-being.
Group members identify with the group.
Use of words like "we", "us", and "team" instead of "I" and "they".
OCBs targeted at the team; supportive behavior—backup, offering to switch days (needs to be contextualized).
Team members make personal attacks (reverse coded).
Negative, task-irrelevant communication (reverse coded).
Cursing or negative tone of voice targeted at fellow team members (reverse coded).
Heart rate or galvanic skin responses during discussions with team members (exploratory).
Members see themselves as having nothing in common with fellow group members (reverse coded).
Group pride
Messages of positive affirmation about being part of the team, usually made outside the immediate group.
Desire to stay in the group; come back and train with same people.
Coming up with a team name, call signs, patches, T-shirts, and logos.
Sticking up for the team, defending the team and its members.
Highlighting the team's accomplishments.

6.2.5 Conflict Markers

Team conflict has been defined as "the process resulting from the tension between team members because of real or perceived differences" (De Dreu and Weingart 2003, p. 741). Closely related is the notion of conflict management, which refers to the process through which team members engage in strategies to effectively manage conflict as it arises.

In examining published scales reported to measure conflict, we found a mixture of conflict and conflict-management markers. Therefore, we included both in Table 14. These markers were based on material in papers on conflict published by Rutkowski et al. (2007), Espinosa et al. (2012), Gupta et al. (2010), Barker et al. (1988), Huang (2010), Jehn (1995), Kostopoulos and Bozionelos (2011), Rispens (2012), Jehn and Mannix (2001), and Pearson et al. (2002). Tables 15 and 16 list potential behavioral markers of team conflict and conflict management. (In the tables, "R" indicates an item should be reverse coded.)

Table 15 Potential markers for team conflict

General conflict markers
Frequency of disagreements over performance episode(s).
Have substantial agreement on goals, strategies, and processes (R).
Nonverbal signals such as a furrowed brow, loud volume, or red face (R).
Team member verbally communicates frustration with team (e.g., phrases, loud sighs, or indication of unwillingness to continue working with someone or on a task) (R).
Team members raise their voices at each other (R).
Interpersonal conflict markers
Frequency of disagreements about opinions.
Ratio of task to relationship conflicts, taking into account frequency of disagreements.
Task conflict markers
Frequency of disagreements about which tasks have highest priority during performance episode(s).
Frequency of disagreements about who should do what over performance episode(s).
Frequency of disagreements about who should receive what resources over performance episode(s).

Table 16 Potential markers for team conflict management

Team member communicates the benefits of his/her opinion, strategy, or approach.
Team member or leader encourages a “We are in it together” attitude.
Team member(s) or leader seeks a solution that will be good for all of us.
Team members adhere to current plans/opinions (e.g., stick to protocol) until teammates provide reasonable rationale otherwise for changing.
Team members are discouraged from showing anger or losing their cool (e.g., punishment).
Team members are open and receptive to differing opinions, strategies, or approaches.
Team members are respectful of one another (no name-calling, no talking down to team members).
Team members disclose instances of conflict/tension to individuals outside the conflict (e.g., leaders, supervisors, other teammates).
Team members identify conflict.
Team members initiate addressing conflict when it arises.
Team members refrain from showing anger that may arise, use strategies to keep one’s self calm.
Team members take initiative to communicate differing opinions, strategies, or approaches.
Team members take initiative to notice and communicate with team members who have expressed differing opinions, strategies or approaches with one another.
Two or more team members discuss details of merging their respective opinions, strategies or approaches.
Team members avoid communicating differing opinions, strategies or approaches (R).
Team member automatically defers to a strategy, opinion, or approach despite being aware of a conflict (R).
Team member demands (e.g., verbal, threats, intimidation, or denial) that his/her opinion is correct (R).
Team member or leader overstates his/her needs and position to get his/her way (R).

6.2.6 Team Communication

Communication has been defined as “the process by which information is clearly and accurately exchanged between two or more team members in the prescribed manner and with proper terminology; the ability to clarify or acknowledge the receipt of information” (Cannon-Bowers et al. 1995 p. 345). Our team-communication markers were created based material in papers that contain scales, published by Bunderson and Sutcliffe (2002), De Dreu (2007), Lee et al. (2010), Hirst (2009), Huang and Cummings (2011), Jong et al. (2005), Tung and Chang (2011), Walther and Bunz (2005), Schippers et al. (2007), Cronin et al. (2011), Espinosa et al. (2012), Faraj and Sproull (2000), Fletcher and Major (2006), Gajendran and Joshi (2012), Gupta et al. (2009), and Haas (2006).

In examining published measurement instruments that pertain to communication, we found that when thinking about communication markers one must be cognizant that team communication is a complex endeavor; as such, in developing markers you can think about those that speak to the characteristics of the communication as well as the content of that communication (Table 17).

Table 17 Communication markers’ components

Characteristic	Content
Task/Nontask	Transition phase
Concise	Mission analysis
Timely	Goal specification
Clear	Strategy formulation
Closed loop	Action phase
Push vs. pull	Monitoring goal progress
Unique vs. shared	Systems monitoring
	Team monitoring/backup
	Coordination
	Interpersonal
	Conflict management
	Motivation/confidence building
	Affect management

Table 18 lists potential behavioral markers for communication.

Table 18 Potential markers of team communication

Subdimension	Marker
General information sharing	The number of times task-relevant information is shared.
	Team-member communication is concise.
	Team-member communication is timely.
	Team-member communication is clear.
	Frequency of requests for clarification (R).
	Frequency with which team members engage in closed-loop communication.
	Team members verbally acknowledging when a team member shares information or makes a suggestion.
	Frequency with which team members confirm receipt of information presented to them (e.g., acknowledgment) by other team members (“Okay”, “Got it”, “Roger, that”, etc.).
	Frequency of closed-loop sequences (task specific).
	The number of times unsolicited, task-relevant information is pushed to team members (anticipatory communication).
Contextualized information sharing	The number of times team members provide unsolicited, but important, task-specific information.
	The number of times team members provide unsolicited, but important, role-specific information.
	The number of times task-relevant information is requested (pulled).
	The number of times team members share unique, role-specific information with one another.
	The number of times expert knowledge is verbalized to the team.
	Team members tell others about the role they are supposed to be playing and how they will execute their responsibilities.
	The amount/number of times unique (new/insightful) information is shared among team members.
	The ratio of shared unique to shared common information.
	The number of times team members share inaccurate information (R).
Transition phase	Number of times team members clearly state their expectations regarding the main mission tasks.
	Number of times the team shares information regarding environmental conditions.
	Number of times team members discuss roles and responsibilities (establishment of, sharing information pertaining to, who has a role and its capabilities/responsibilities).
	Number of times team members discuss the resources available to the team.
	Number of times team discusses strategy and approaches to the task.
	Number of times goals are discussed.
Action phase	Number of times team members update one another on progress toward goals (e.g., what remains to be accomplished or current progress).

Table 18 Potential markers of team communication (continued)

Subdimension	Marker
Contextualized information sharing	
	Number of times team members share information regarding systems monitoring. • Number of times team members update one another concerning issues experienced when completing tasks (e.g., task challenges and performance barriers) • Number of times team members share information regarding the tracking of team resources • Number of times team members share information regarding the tracking of environmental conditions Number of times team members provide verbal feedback to one another. Number of times team members provide coaching to team members. Number of times team members provide verbal assistance to a fellow team member in terms of clarification statements to facilitate the accomplishment of an action associated with that member's role. Number of times team members verbally provide backup to another member. Number of times information is shared regarding the sequencing of team-member action. Number of times information is shared regarding the timing of interdependent action on the team.
Contextualized information sharing	
Workflow knowledge (general)	Frequency count of the number of times task knowledge was requested from each team member (more of a social-network measure).
Interteam communication (general)	One could ask the same items as within information sharing, but add the caveat that the information exchange is about exchanging information outside the team. • Boundary-spanning perspective • Multiteam system perspective

6.3 A Way Forward: The Testing of the Proposed Markers

The described effort represents a first step in a much longer process of marker development that would involve experimentation to ensure the usability of the items (as some may need further contextualization or differ in instrumentality dependent on the particular task characteristics) as well as experimentation aimed at establishing the validity (e.g., convergent and divergent validity) of the item set. This may also include factor analysis for those marker sets that may be argued to be multidimensional.

We have also described some potential next steps for beginning to test the efficacy and validity of the proposed markers. Moreover, we focused on developing markers

of emergent states as these often represent constructs in which existing scales represent attitudes or cognitive aspects of teams and, as such, may need more testing and refinement as compared with behavioral constructs whose measures reflect behaviors by definition. Future work should not only continue to develop behavioral markers for the remaining identified constructs that meta-analytic work indicated were important in facilitating team outcomes, but begin to test out the markers developed in this initial effort.

6.3.1 Gaining Evidence of Content Validity

One potential method by which additional evidence for content validity could be gained is as follows. The initial steps would consist of developing a study protocol and corresponding measures such that the ARL Institutional Review Board's (IRB's) approval could be obtained. With regard to the development of the survey that would be given to the identified SMEs, the construct of interest would be defined at the top of the survey. Then, we would include each of the markers developed for a given construct and for each one the SMEs would provide a rating of the degree they felt that particular item reflected the construct of interest. Most likely we would use a 5-point Likert scale for these ratings. At the conclusion of rating each individual marker we would list the markers in totality and ask a final question that would tap the degree to which the markers fully sampled the construct of interest; again, most likely using a 5-point Likert scale. However, this time we would also include an open response so, in the case they felt the construct was not being fully represented, they could highlight what area they felt was deficient.

6.3.2 Gaining Evidence toward Construct Validity

A second group of studies are those that can begin to provide evidence of construct validity for each of the aforementioned sets of behavioral markers. While there are a number of approaches that can be used to show evidence for the construct validity of a given measure, we recommend the use of a multitrait-multimethod approach. Using a Multitrait-Multimethod Matrix approach (Campbell and Fiske 1959) one can examine the degree of convergence and divergence of the developed measure with other measures and, in doing so, gain insight into construct validity. We will demonstrate construct validity for our metrics by comparing our set of behavioral markers for a particular construct against a number of validated metrics for that same construct. There are also a number of measures that can be given along with our metrics that would be expected not to correlate with our measures, providing discriminant validity.

The initial steps in this effort would be to map out each construct's nomological net, thereby providing an initial foundation of constructs whose measures may

provide convergent and divergent validity evidence. We would primarily use theory to do this, but we also have some additional insight based on the MASEM efforts conducted in Year 2. Once the nomological net is developed, the next step would be to identify established measures that would be expected to converge as well as those where divergence would be expected. Once the measures have been identified the next step would be to decide the setting in which to examine the relationships. Within this same set of studies one could also seek evidence of predictive validity. In essence, predictive validity examines the degree to which the developed measure predicts key outcomes of interest. In the cases of the developed sets of markers this may be specific team outcomes (as evidence in the meta-analysis) or other team constructs; for example, trust-predicting information sharing within the team.

6.4 Team Indexing and Aggregation

Team-level assessment seeks to speak to phenomena that are particular to teams that sometimes cannot exist on an individual level alone (e.g., emergent states). Despite the need to speak to team-level phenomena for effective team-development intervention implementation and evaluation, there are several steps that are necessary to achieve a team score. For instance, theoretical decisions must be made regarding the level at which the construct exists and how it may differ from individual-level phenomena, such as compositional, compilational, and emergent (Kozlowski and Klein 2000).

To guide and justify aggregation, constructs must first be identified as team-level variables. Particularly, the researcher or practitioner must ensure there is a clear conceptual basis for aggregating data (Woehr et al. 2015). As much of team data are most commonly collected at the individual level (Tesluk et al. 1997), constructs should be examined for the particular way in which individual-level constructs relate to the construct at a team level. It is, first, important to note this report assumes team-assessment needs have been identified and that constructs are expected to appear at the team level in some fashion. Second, the identified team-level construct should be examined for configural properties (i.e., how does the variable appear at the team level?).

The review by Woehr et al. (2015) of the team-assessment literature reveals that 61%–96% of studies using team-level assessment fail to include or address one or more necessary steps to adequately infer team-level constructs. This amassed confusion may be because publications for each topic of team assessment (e.g., theoretical decision-making guidance and statistical guidance) are dispersed in various publication outlets. For instance, LeBreton and Senter (2008) provides a comprehensive step-by-step guide for calculating and selecting appropriate team

agreement indices, whereas Kozlowski and Klein (2000) advise on many theoretical issues related to multilevel assessment. While each paper, book, and chapter addressing multilevel issues is integral to advancing robust assessment of teams, the field may benefit from an integrative review to provide high-level guidance for practitioners, students, and future research.

In response to this need, the current report attempts to provide integrative guidance for individuals seeking to leverage team-level assessment. In the effort to provide counsel, the report provides a practical “how to” on best practices to guide 1) conceptual development of team constructs, 2) refinement of measure and item referents, 3) selection of agreement indices, and 4) choice of an aggregation technique. Specifically, through this attempt to unify the guidance at large, this report aims to complement current best practices for multilevel assessment. To facilitate a more direct translation of the guidance to practice, we offer a scenario of teams in a practical setting that will be used for demonstrative purposes at each step in the multilevel assessment process.

6.4.1 Conceptually Developing Team Constructs

The configural properties of teams fall into one of two primary categories—either compilational (i.e., additive model) or compositional (i.e., dispersion model) (Kozlowski and Klein 2000). The configural property of composition is based on the premise that the individual-level (i.e., lower-level) variable is isomorphic with the team-level variable, such that the construct being assessed is essentially identical at both the individual and team levels. It is important to note that team constructs infrequently identify identically at multiple levels (Firebaugh 1978) and that, more frequently, team constructs are considered fuzzy compositional processes. In fact, fuzzy compositional processes occur when an aggregated higher-level variable (e.g., team-level variable that has been aggregated from individual response data) is related or similar to individual processes, yet contains higher-level contextual influences. In this way, the variables are similar at the individual level and team level; however, the team-level construct contains higher-order influences than can be captured via individual response data alone. Variables that are compilational in nature, however, occur when the team construct is expected to vary at a lower level. Specifically, when team variables are compilational, the higher level is theoretically distinct from the lower level (Chan 1998).

6.4.2 Refining Measure and Item Referents

While there are best practices to guide development of team measures, including surveys (Wageman et al. 2005) and behavioral markers (Rosen et al. 2011), emphasis is placed on ensuring measurement is at the team level rather than psychometrically robust. This is not to say validity and reliability are not important to team assessment; rather, refinement of key measurement strategies is necessary to capture constructs at a team level. Wording of particular items and markers can be modified in an instructional segment (e.g., “Respond to the following regarding your team”) or reflected in the item itself. Further, the item should tap into a particular construct of interest theorized to appear at the team level such that it promotes between group variability (Klein et al. 2001).

Further considerations for item refinement stem from considerations for the features of the team for which the measure is being refined. For instance, a team is traditionally defined as consisting of at least two individuals who have specific roles and functions to execute within the team so that individuals can work collectively and interdependently toward a shared goal (Salas et al. 1992). This definition postulates that a group of two people who work together toward completing a common goal, who contribute to a role and/or function, are a team. However, the literature on teams has noted the stark contrasts in teams of three or more individuals and dyads (i.e., teams of two individuals). Essentially, research has highlighted that larger teams are more complex and, as there is more potential for variation at the team level, have different measurement needs; such as, information can be shared differently within a triad than a dyad, resulting in differences of shared mental models at a team level (Mohammed and Dumville 2001).

6.4.3 Selecting Agreement Indices

As team concepts are thought to emerge from individual responses, researchers and practitioners should capture the variance and agreement among team members prior to aggregating data. At this step, the importance of the original conceptualization of the pattern of team configuration and emergence increases. Namely, the next step in the team-assessment process occurs once item referents have been modified to capture team-level data and have been administered to collect data at an individual level. More specifically, team-agreement indices are selected and calculated to determine the level of agreement or dispersion within individual data. This informs the dispersion and agreement among teams and can be used to justify aggregation to a team level. The two most common methods to test for appropriate statistical convergence for multilevel assessment are interrater agreement (IRA) indices and interrater reliability (IRR) indices. IRR calculates the consistency with which a

judge or rater (i.e., teammate) ranks an item or target consistently in comparison with other judges. In other words, this provides a score of relative consistency among teammates without accounting for an absolute or “true” score. IRA, on the other hand, estimates an absolute consistence for multiple judges for one or more targets when there is an absolute or “true” score for the target.

LeBreton and Senter (2008) provide comprehensive guidance for selecting and computing the appropriate team-agreement index. However, in order to make an informed decision on the selection of appropriate team-agreement indices, one must first refer to the conceptual definition of the team construct’s configuration as to how the team level relates to individual-level constructs. For instance, truly isomorphic team constructs do not require the calculation of a team-agreement index. The reason for this is that the individual level and team level of the construct match identically and individual-level responses are not expected to vary within the team or differ from the team-level variable. For fuzzy compositional and compilational variables, however, justification for aggregation is warranted. It is important to identify the number of judges and the number and type of targets (i.e., whether or not they have an accurate response). Pearson product-moment coefficients are the most popular form of IRR; yet, selection of an appropriate IRA is challenging. This is particularly true as each team-agreement index is highly interrelated (Roberson et al. 2007). However, determination of an appropriate rWG index (James et al. 1984) depends on the researcher’s or practitioner’s responses to two questions, alluded to earlier in this segment:

- 1) Are there multiple judges rating a single target or multiple targets?
- 2) Do(es) target(s) have an absolute or “true” score (i.e., is there a correct answer to the target)?

When there are multiple judges, $rWG(j)$ is most appropriate; however, when multiple targets with multiple true scores are present, it has been advocated that a newly created rWG index be used (LeBreton and Senter 2008)— r^*WG —as it can statistically determine the maximum disagreement that exists among judges (Lindell et al. 1999). Calculation of this and similar rWG indices can account for skewed distribution in the data (Cohen et al. 2001). Intraclass correlations (ICCs) calculated among individual raters estimate the absolute as well as interrater consistency, such that they are advantageous as they provide an estimate of both IRA and IRR. As each agreement index is subject to the pitfalls of its calculation, it is recommended that both a measure of ICC and rWG be leveraged in justifying aggregation. The in-depth method for calculation of these indices, as well as further conceptual guidance for index selection, is presented in LeBreton and Senter

(2008). Resulting scores should meet specified cut-off criteria for adequate agreement prior to aggregation of individual-response data.

6.4.4 Choosing an Aggregation Technique

Aggregation refers to the act of achieving higher-level constructs through statistical procedures applied to individual-level data. Many methods are available to create team-level indices from individual-level data, inundating practitioners and researchers with an abundance of information. However, to make an informed decision, one must refer to the team construct's configural properties to select an accurate aggregation method. Once agreement among team members has been established using the chosen agreement indices—assuming it is theoretically necessary to justify aggregation through agreement indices (such as in fuzzy composition models)—the next selection is the method of aggregating to the team level.

Aggregation can involve averaging or summing individual team-member responses to important indices of team-level phenomenon such as accuracy or similarity. However, in some instances team members vary in terms of their influence on the team-level outcomes. For example, one team member may have strong opinions on the strategy a team should employ for a shared-decision task, meaning certain team-level indices may be more strongly influenced by one person rather than by the team as a whole. To account for these invariances among team members, Kozlowski and Klein (2000) have advocated for matching team-level aggregation techniques to the theoretically appropriate methodologies. In referring back to the identified configural and emergent properties of the team constructs in question, this helps to better inform the selection of an appropriate aggregation method. Dispersion or composition models reflect the distribution of individual characteristics within the team (Molleman 2005). Thereby, the variance's minimum and maximum are typically associated with the aggregation methods used for constructs within this configural model (DeNisi 2000), whereas additive or compilation models rely on computing a sum or average to reflect a team score.

6.5 Discussion

While the current report focuses more exclusively on the needs of the practitioner, evidence begins to speak to the changing nature of teams, particularly as teams become more geographically dispersed and reliant on technology (Murase et al. 2012) and become more fluid in membership (Andreatta 2010). As such, Murase et al. (2012) call for a shift in the way teams are assessed and understood in the workplace such that levels of virtuality and dispersion in networks can be better understood. While the current report seeks to align future research efforts with the

best practices communicated within the literature on multilevel assessment at large, we hope this report can serve as a building block upon which changes in teams and team assessment needs can be further developed and better understood. While team assessment may be headed toward understanding networks and dealing with fewer stable teams, theoretical guidance on capturing individual variation in the nature of the response can continue to inform what is happening at the team level (Smith-Jentsch et al. 2009).

7. Feedback

Oftentimes, there are simply tasks, activities, or missions that cannot be accomplished by an individual alone. Similarly, the support of a team serves to increase overall performance and reliability on a given task. In either case, a team cannot be optimally effective if team members are not able to work together in a cohesive manner. Consequently, teamwork is gaining more attention across fields as a vital component for achieving optimal outcomes (Driskell et al. 2010; LePine et al. 2008). As the use of teams continues to rise (Marks et al. 2001; Tröster et al. 2014), it is increasingly important that we understand how to maximize their performance.

It is not uncommon for teams to be less than the sum of their parts (Kerr and Tindale 2004). An example of this situation is when a team of experts is not necessarily an *expert team*. To further expound this, a team comprising highly skilled individuals does not by any means oblige an effective or high-performing team. It has been found that teams have the potential to improve performance (Mathieu et al. 2000; Salas et al. 2008a; Salas and Fiore 2004); however, there are many factors to consider. Providing feedback is an effective way to influence these factors before, during, and after the team completes its tasks. With that in mind, a recent meta-analysis showed that accurate feedback can enhance performance by approximately 25% over those teams that did not engage in feedback (Tannenbaum and Cerasoli 2012). The effect of feedback seems to hold in business units and also influence important psychological outcomes such as job satisfaction, morale, and turnover intentions (Pritchard et al. 1988). However, the integration of how specific feedback elements should exist in order for feedback to be effective is still pending.

In order to counter such deficiency, the purpose of this section is to highlight best practices to inform the structuring of both individual and team performance and process feedback. With that in mind, we reviewed the literature to compile the best practices to deliver feedback in the most effective manner within the team intelligent-tutoring context. First, we outline the best practices in regard to trainee characteristics, feedback content, and implementation. These are later summarized

in a table to help synthesize the information of this report. Second, we bring the challenges specific to the team-tutoring environment to the forefront throughout these sections. Emerging technologies are often used as a mechanism to deliver information more efficiently, but it is important to recognize their pitfalls. Finally, we mention gaps in the literature and how they shape future research.

7.1 Team-Feedback Review

Feedback is a way to provide an enhanced understanding of how individuals and/or teams are doing in regards to desired performance. Feedback has been identified as a critical tool to facilitate learning for over two decades (Komaki et al. 1980). Team feedback can be defined as the communication of information concerning actions, events, processes, or behaviors relative to a task completed by a team (London 2003; London et al. 2005). One of the primary purposes of feedback is for team members to communicate in a manner that results in the improvement of individual development and interpersonal collaboration (London et al. 2005). A common method for providing feedback is known as “debriefing” (i.e., after-action reviews). Debriefing techniques are one mechanism by which feedback is delivered. More specifically, debriefing is a feedback tool that can facilitate the constructive communication of past events, the discussion of both positive and negative aspects of the team, and the development of future action plans (Tannenbaum and Cerasoli 2012; Smith-Jentsch et al. 2008).

There are different methods of debriefing—as a form of feedback delivery—of which one of these methods is Team Dimensional Training (Smith-Jentsch et al. 1998). This particular debriefing technique uses guided team self-correction to enhance team members’ knowledge of teamwork processes and mastery of task-related skills (Smith-Jentsch et al. 1998). Guided team self-correction is one of the most popular and effective feedback models (Salas et al. 2007a; Smith-Jentsch et al. 2008). Guided team self-correction can be defined as a feedback tool in which a facilitator keeps the discussion focused, fosters a psychologically safe environment, encourages engagement in the session, and guides team members on constructive feedback delivery and reception (Smith-Jentsch et al. 1996). Feedback also includes team members monitoring each other’s performance, providing constructive feedback regarding errors, and offering advice for improving performance (Blickensderfer et al. 1997). According to Hinsz et al. (1997), feedback has the potential to challenge previously existing attributions and allowing groups to use such feedback in better ways than they would if individual feedback was given instead.

We will adopt the term feedback in this report for being the broader umbrella term of other methods (e.g., debriefing). Feedback has been shown to be beneficial to enhance teamwork mental models, teamwork processes, and team-performance outcomes (Smith-Jentsch et al. 2008; Tannenbaum and Cerasoli 2012). For instance, feedback regarding the team's performance is essential to enrich any practice opportunity when training for teamwork (Dreifuerst 2009; Issenberg et al. 2005). Feedback is not only important during training, but it is also essential throughout team action periods to continuously develop and reinforce proper teamwork. By discussing what team members did correctly or not, team members can develop a better, shared understanding of their tasking and subsequently enhance their performance (Smith-Jentsch et al. 2008). The more they perform does not necessarily lead to better understanding of how to perform in an optimal manner, but the discussion of it and developmental feedback are key components when the goal is to improve team performance (Mathieu et. al. 2000). This is likely due to the possibility of adaptation when mistakes are brought to the forefront and members converge their understanding in regards to their performance (DeChurch and Mesmer-Magnus 2010).

7.2 Is Team Feedback Always the Right Solution?

The immense support from the literature with regard to individual-level feedback stands in stark comparison to the scarcity of team-level research in this area. The juxtaposition of this research results in a quandary identified by DeShon et al. (2004): Should feedback be provided to individual team members or the team as a whole? If one team member's actions or behaviors are the root cause of the negative performance of the team as a whole, then feedback should be provided on the individual-level to address errors. Conversely, if poor performance can be attributed to the team as a whole then feedback should follow suit and address the whole group.

Research often supports the importance of both levels of feedback when applied to interdisciplinary tasks that require both individual and team-level action (DeShon et al. 2004). Regarding individual feedback, Klaus and Glaser (1970) recommend that individual-performance feedback should be given only if the task allows that such feedback occurs regularly in the operational environment. Regarding the team feedback, one must consider that team performance is not just based on teamwork processes, but it also includes taskwork (Smith-Jentsch et al. 2008). These authors reiterate the potential of feedback to improve both teamwork and taskwork. More specifically, feedback should guide individuals toward teamwork where interdependency lies, but also ensure they have the proper proficiency to execute a task. As an illustration from the medical field, facilitators are required to be

competent not only regarding their clinical knowledge but also their communication to delineate the proper areas of improvement and to aid on resolution of issues (Fanning and Gaba 2007).

It is not surprising we expect an automated tool to face difficulties in identifying when to provide the appropriate feedback. More specifically, the tailoring of feedback to specific targets can be challenging to trace behavior of individuals during task execution that is highly interdependent. Consequently, when team feedback is the right solution, the team intelligence tutoring has to take careful steps to provide the team feedback (i.e., teamwork and taskwork). Otherwise, problems with potential ostracism of team members may emerge if the machine targets someone and disperses the individual feedback in front of the other teammates. Therefore, differentiating whether individual and/or team feedback should be given becomes a crucial decision to improve team performance.

7.3 Best Practices for Team Debrief

Even though feedback seems to improve performance, the conditions under which feedback can be most effective remain unanswered (Appelgren et al. 2014). In fact, the impact of feedback on performance on both the group and individual level can vary widely (Kluger and DeNisi 1996). There is not a “one size fits all” approach to feedback, but multiple methods have been supported throughout the literature to facilitate this process. We identified 13 best practices as part of team-member characteristics, feedback design, or work environment that researchers and practitioners should be aware of. Finally, special attention is given to the potential challenges that can emerge and hinder the implementation of such best practices when the context is team intelligent tutoring.

Baldwin and Ford’s (1988) training model was leveraged as a way to organize feedback best practices. More specifically, the model inputs were used as organizing mechanisms to delineate feedback best practices. The words “trainees” and “training” in the original model were replaced by “team members” and “feedback”, respectively, to fit the purpose of going beyond the training context. With that being said, we have categorized the best practices into team-member characteristics, feedback design, or work environment.

7.3.1 Team-Member Characteristics

There are a number of characteristics that can make feedback more effective. First and foremost, team members must be ready to provide and receive feedback. For instance, demonstrating flexibility, enthusiasm, and motivation while engaging in team feedback are key elements (Ahmed et al. 2012). It is crucial that the feedback

is honest and brings up strengths and weaknesses of performance episodes (Lederman 1992; Salas et al. 2008b) in order to guide the team's reflection regarding their performance. When concerned about characteristics of the facilitator of team feedback, one should target both technical competence and good interpersonal skills (Tannenbaum and Goldhaber-Fiebert 2012).

Even though we just pointed out individual characteristics that are important to consider for team feedback, the interaction of those characteristics when individuals belong to a team is just as important. In congruence with Kerr and Tindale (2004), a team can succumb to inefficiency despite being composed of experts. It is not uncommon that even functional experts require training prior to team feedback (Tannenbaum and Goldhaber-Fiebert 2012). As a whole, trainees must possess emotional resilience in addition to task competence to receive, provide, and implement team feedback. Some individuals may be more hesitant than others to discuss their flaws, while others may be too focused on what went wrong and neglect the positive aspects of team processes and performance. Accordingly, Smith-Jentsch and colleagues (2008) emphasize the importance of discussing both negative and positive events that comprised the team's mission, task, or activity. Considering such emotionally loaded scenarios, team members must be cognitively and emotionally ready in order to benefit most from team feedback.

Best Practice 1: Prepare team members for (even negative) feedback.

The participation of team members is one of the main variables in determining whether the team feedback will be successful or not. With that in mind, individuals should be instructed to actively engage throughout feedback sessions (Smith-Jentsch et al. 2008). Such participation can help in the development of a common understanding regarding their goals, areas of improvement, and what optimal teamwork is. It is important to reinforce the proper behaviors such as the communication of points of view, different perspectives, and emotions (Mort and Donahue 2004). However, team members should not dwell on specific emotional issues without the concern for understanding and improving those issues. With that in mind, programs have been developed to account for the release of emotions earlier in the feedback process (e.g., debriefing for meaningful learning) (Dreifuerst 2010, 2012). According to these programs, team members can be more objective when they reflect upon the scenarios experienced in the simulation. Furthermore, even the positioning of team members can influence whether or not team members are willing to participate in the feedback. Having all team members physically visible to one another seems to ease their communication and interaction (Smith-Jentsch et al. 1998; Salas et al. 2008b). Thus, engagement in diagnosing the team can be maximized in a number of ways.

Best Practice 2: Ensure team members are actively engaged in diagnosing the performance.

In order to encourage team members to contribute in a meaningful way, both relevant information sharing and self-learning should occur within the teams (Mort and Donahue 2004). In the presence of a facilitator, he/she should not just be providing examples of accurate or inaccurate behavioral manifestations. It is more beneficial if team members are required to reflect and come up with the examples themselves (Cicero et al. 2012; Tannenbaum and Cerasoli 2012), and the facilitator should only interject when needed. Allowing time for everyone to reflect, even if silence remains, should be acceptable and could often encourage participation (Arafeh et al. 2010; Overstreet 2010). The discussion of feedback among team members can enhance learning (Helstad and Lund 2012; Holmquist 2007), resulting in the encouragement of critical thinking and development of new insights (Bartram and Roe 2008). However, guidance from facilitators may be required when team members cannot identify examples on their own and/or they are providing a wrongful scenario that could hinder the overall understanding of ideal performance.

Best Practice 3: Facilitators should interject the discussion of feedback only when needed.

To properly deliver team feedback, the main target should be the team's states, activities, and outputs that are produced or emerge during their encounter. One may want to focus inward when thinking about good and bad examples, but it is more important to keep in mind the overarching goal of converging everyone's understanding regarding teamwork rather than concentrating on individual outputs. This is especially true when team members undergo high levels of stress, which is likely to make them focus on their individual task (Driskell and Salas 1991; Kleinman and Serfaty 1989). Knowing that team members will possibly have different perceptions about the same event (Ahmed et al. 2012), focusing on reaching a common understanding often becomes challenging, but it is an attainable goal. Accordingly, Salas and colleagues (2008b) emphasize the importance of collective orientation to mitigate the individual tendencies and prioritize the overall team.

The development of effective team feedback must account for the complex nature of the interactions present within teams. For example, the team feedback may or may not be directly related to an individual. In order to fully benefit from team feedback, team members need to focus on the "big picture" instead of focusing solely on individual behaviors. It has been shown that teams receiving both individual- and team-level feedback or just team-level feedback showed a higher

level of commitment to team goals than participants who only received individual-level feedback (DeShon et al. 2004). These findings illustrate the importance of tailoring feedback to the whole team.

Best Practice 4: Tailor the feedback to the team and not just to the individuals.

7.3.2 Feedback Design

The feedback should be designed around scientifically driven findings, including the content, target, and feedback source. Rightfully, feedback can be used to target an individual's attention to the right processes and performance episodes (Appelgren et al. 2014). Research suggests that in any given scenario, it is most effective to offer both process and performance feedback situations (Walsh et al. 2009). Even though performance feedback is more common and often the sole technique implemented when thinking of feedback, the focus on process is just as (if not more) important. Process feedback focuses on the behavioral manifestations that take place during a task rather than the results themselves. Without process feedback, learning can occur by chance instead of a more systematic approach to assure such desired outcome.

Recent research also suggests feedback should establish a process orientation as the primary focus (Salas et al. 2008a; Smith-Jentsch et al. 1998) to allow for increased adaptability (Gabelica et al. 2012; Weaver et al. 2010). There is growing support in training and simulation research for the benefits of focusing on process or mastery-level feedback (Hattie and Timperley 2007; Salas and Cannon-Bowers 2001; Salas et al. 2008a). It is going beyond whether or not team members achieved the desired outcome(s), but understanding what happened while they were coming together to execute the task and how they accomplished that. This process focus not only eliminates the interference of extraneous variables but also allows for the development of more generalizable skills that can transfer to different tasks (Kozlowski et al. 1995).

Best Practice 5: Target the processes, not just outcomes, during team feedback.

In addition to ignoring team processes, another common mistake that can significantly diminish the effectiveness of feedback is not being specific enough when pointing out examples. Tannenbaum and Cerasoli (2012) identified the importance of being more specific—versus general—in regard to performance as one of the essential criteria (Tannenbaum and Cerasoli 2012). These authors also point out the importance of structuring the debriefing in an organized manner that will facilitate participation. Examples that are given throughout the team feedback should be specific, open-ended, and behaviorally based to maximize their utility (Smith-Jentsch et al. 2008). This is true for both positive and negative examples of

team processes and performance (Weaver et al., 2010). The reason for requesting specificity includes the richness in information and the decrease in subjectivity (Cannon-Bowers et al. 2010).

Practice activities, for instance, enable the provision of direct feedback regarding individual and team performance, which allows team members to gain information about their behaviors and how to modify them (Murthy et al. 2008). Taking this into consideration, the specificity of examples can not only maximize the understanding of the behaviors but also facilitate the implementation of them by providing more guidance than any general example would otherwise. However, Leboe et al. (2005) point out examples of scenario-specific corrections during debriefing (e.g., feedback sessions) that were not generalizable to broader novel events. This highlights the importance of properly distinguishing between specific scenarios and possible “novel” events. Consequently, the applicability of the information learned through feedback depends on the connection with the on-the-job situations one may face. For this, researchers recommend the discussion of both negative and positive events that may emerge (Smith-Jentsch et al. 2008). Throughout the team discussions, team members must be ready to provide the team with specific actions or behaviors that can be analyzed constructively, resulting in the development of future action plans.

Best Practice 6: Be specific, not generic, when referring to (positive or negative) examples and their implementation.

In short, the objective of team debrief sessions is to constructively diagnose both positive and negative team performance, with the ultimate goal of improving future performance. Besides focusing on process and being specific, it is key to focus on the developmental aspect of feedback by targeting task and not the individuals. Developmental intent is an essential debriefing element defined by Tannenbaum and Cerasoli (2012, p. 233) as “a clear, primary intent for improvement or learning that is nonpunitive rather than judgmental or administrative”. An illustration of such task-focused developmental feedback would be identifying a deficiency in team performance by diagnosing underlying causes, developing lessons learned, and cultivating a strategy for improving performance (Beaubien and Baker 2004). This allows for members to address their own issues and develop plans for future actions in a constructive environment that provides structured topics for discussion among the team (Smith-Jentsch et al. 1998).

Not only should quality be a concern for the content of feedback but also regarding the source of feedback (Ilgen et al. 1979). We advocate an increase of source(s) both in terms of quantity and quality when possible. Ideally, having more than two sources beyond the individual level helps increase the credibility of feedback.

Researchers have promoted the use of a greater number of sources when providing feedback (Tannenbaum and Cerasoli 2012; Weaver et al., 2010). Accordingly, others have reiterated the benefits of self- and peer feedback (Kolbe et al. 2013; Salas et al. 2007b; Smith-Jentsch et al. 2008). A meta-analysis on “360° feedback” identified this intervention as a positive influence on behavior when implemented correctly (Nowack and Mashihhi 2012). However, authors warn the use of this feedback as potential reason for disengagement or even decreased performance if the process of implementation has flaws.

Furthermore, the feedback can be improved in regard to the delivery means (Hawley and Mares 2007). For instance, team feedback that incorporated both visual and auditory media improved team performance and shared cognition more so than either format individually (Oden 2008). Videos can serve as tools to recall specific incidents without the distortion of an individuals’ perception (Arafah et al. 2010; Grant et al. 2010) as well as reinforcing the appropriate learning objectives (Carron et al. 2011). Tannenbaum and Cerasoli (2012) emphasized the employment of multiple sources of feedback as well as assessing the quality of the feedback. They went on to point out that quality and quantity are intertwined as a feedback best practice. The more information sources you implement when delivering feedback, the more the credibility (e.g., quality) of the feedback is enhanced. Therefore, as supported by Tannenbaum and Cerasoli (2012), we suggest increasing not only the quantity of sources but also the quality of it (objective ratings, video, etc.) to avoid misinterpretation or oversight of important behaviors.

Best Practice 7: Invest in quality of feedback and its sources to increase effectiveness.

Feedback should be aligned with the desired processes and outcomes the team is expected to engage in and deliver. Drawing from Trait Activation Theory (Tett and Guterman 2000), certain exercises can make certain behaviors more prone to happen. With that in mind, when developing scenarios or situations in which feedback is implemented, consider whether desired behaviors are likely to occur. Accordingly, measures must be designed to capture performance in these planned teamwork opportunities. Measurement tools should then be scenario specific (Brannick et al. 1997), in contrast to global assessments that are prone to overlooking important teamwork behaviors in this context (Marks et al. 2001). Brannick and colleagues (1997) used event-based performance analysis through the Targeted Acceptable Responses to Generated Events or Tasks (also known as TARGETS) methodology. This is a performance-measurement tool that can be applied across multiple settings with reliable accuracy. Scenario-specific measurement tools also control for individual bias while providing objective performance data.

However, feedback is often confounded with different types of evaluations (e.g., grades). To differentiate from these when providing feedback, one should not be assessing or comparing the performance of teammates. The conversation should be about the exchanging information to facilitate the development of teamwork skills, not to praise or demoralize any team member. Discussion of mistakes or knowledge gaps are intrinsic aspects of a successful debrief, as they are prospects for development of future actions (Salas et al. 2008b). Feedback pertaining to errors must be stated in a constructive manner (i.e., solution oriented and nonthreatening) in order to be useful. For training purposes, arguably, the most essential criterion for effective debriefing is the relevance of the debrief discussion to the objectives for training. Alignment of training objectives and feedback content first requires that the simulation scenarios and measurement tools—which inform feedback—are linked to training objectives (Brannick et al. 1997; Chronister and Brown 2017). Thus, objective analysis of learner’s behavior should be measured in order to appropriately diagnose team processes and performance (Rosen et al. 2008). This highlights the importance of psychometrically sound assessment tools to aid team feedback.

Best Practice 8: Develop and validate measures to use for feedback assessment and delivery.

7.3.3 Work Environment

In order for debriefing to be an effective tool, a positive work environment must be established (Ahmed et al. 2012; Smith-Jentsch et al. 2008). This positive environment includes high levels of psychological safety and an organizational climate that encourages learning and development. Specifically, Edmondson (1999) highlights the importance of psychological safety (i.e., shared feeling that it is safe to take risks) to team learning and as a precursor of seeking feedback, sharing information, asking for help, and talking about errors—all of which are key components of effective feedback. Furthermore, if the organization encourages proper execution of tasks and implementation of resources, team members will be more inclined to speak up if tasks are not being conducted properly (Leroy et al. 2012). The promotion of psychological safety and proper organizational climate are incremental to facilitating discussion of mistakes, issues, and conflicts that arise both during and after the debriefing session (Edmondson 1999). It is extremely important team members feel at ease to engage in debriefing and to express their ideas without fear of receiving negative consequence (Dismukes et al. 1997; Edmondson 2003; Tannenbaum and Cerasoli 2012). Accordingly, individuals must trust that the organization will not punish for any faults or errors admitted during a debrief.

When an outside member (e.g., automated tool) is the responsible source to provide the team feedback, comparable issues that virtual teams suffer may emerge, such as the lack of trust (Bowers et al. 2008; Rosen et al. 2006). Organizations must then clearly communicate the purpose of debriefing to team members, that it is for developmental purposes and it does not incur punitive actions (Salas et al. 2008b; Tannenbaum and Cerasoli 2012). Properly facilitated debriefing sessions begin by clearly outlining the goals and objectives of the simulation or performance episode (Buykx et al. 2011). This proper facilitation ensures trust is built properly. In turn, teams and team members receive the highest amount of benefit from the debrief. Taking this together, the welcoming work environment should promote psychological safety, create an organizational climate conducive of voice, and build trust among team members.

Best Practice 9: Create a welcoming environment to allow for relevant information exchange during team feedback.

It is important to set the tone early on to maximize the benefits of team feedback. Specifically, motivational feedback can help members to increase effort and spur buy-in toward their goals. Positive motivation can be in the form of rewards, recognition, or other ways that serve to increase feelings of collective efficacy. For instance, simple preflight briefings—as opposed to no briefings—were sufficient to shape better crew interaction (Ginnett 2110). Promoting “early wins” can help team members build collective efficacy (Tasa et al. 2007). Team members are generally not as receptive to strategy adjustments that early on, and are not as prepared to absorb knowledge either, particularly if it is a novel task. Thus, the initial steps taken by the facilitator to ease team members’ belief and contributions to the team feedback can help validate the purpose of this activity (in addition to timing that will be discussed next).

Furthermore, the learners’ perception of safety is influenced by how emotions are managed during the delivery and reception of feedback (Savoldelli et al. 2005). Taking this into consideration, simulated events can trigger a plethora of emotions (e.g., stress and anger) in team members that each individual must handle after the event has taken place. Feedback sessions require open, honest communication that addresses areas that are not always amicable, such as mistakes, interpersonal conflict, and miscommunication. This is the area of a feedback session in which challenges surface. Honest communication, a safe environment, and constructive discussion are all necessary aspects of a debrief (Cant and Cooper 2011; Edmondson 1999, 2003; Fanning and Gaba 2007; Gum et al. 2011), yet these aspects can contradict each other (e.g., feelings of vulnerability can arise if the team member must address negative aspects of a task). Thus, it is paramount that facilitators maintain a heightened sense of awareness with regard to team members’

emotions and behaviors during the feedback session (Savoldelli et al. 2005). In addition, facilitators should also be trained to handle difficult discussions as they arise and keep the group focused on constructive goals and communication (Savoldelli et al. 2005). Proper facilitator training can ensure that discussions containing sensitive topics can be addressed effectively.

Best Practice 10: Set up a plan to deal with goals and emotions early on the process of team feedback.

The timing of the feedback is still in its nascent stages. Even though timing can play a critical role in feedback reception and interpretation, Dismukes and colleagues (2000) noted the subjective observation of timing within feedback is often understudied. Primarily based on the coaching research of Hackman and Wageman (2005), researchers have established what to focus on at different points throughout the entire life cycle of the team. They emphasize the team should be ready to receive feedback and able to implement it; hence, the importance of timing. This includes a focus on motivating in the beginning, strategizing in the middle, and education at the end of the team life cycle. When feedback delivery occurs at an inappropriate time, such as when the team's attention should be diverted to other portions of its tasking, feedback can be detrimental rather than helpful. With that in mind, we will expand on the timing of feedback in regard to focus on motivating, strategizing, and educating.

Regarding motivation, individuals experience anxiety about their performance as well as a loss in retention of critical incidents when a significant lapse in time occurs following the performance episode and preceding the feedback session (Arafeh et al. 2010). Additionally, the examples may not be recalled with such accuracy if too much time goes by in between performance episode and team debriefing. To effectively critique using a video, pause the performance episode at the specific times when critical behaviors are (or should have been) performed—with the caveat the feedback should remain clear, concise, and motivating. This allows participants to receive feedback and evaluation “in the moment” (Cant and Cooper 2011). Feedback should be given when performance is still recent and fresh in team members' memories.

Regarding strategizing and educating, it is important to recall the different team-process phases a team goes through. Considering the transition and action phases described by Marks and colleagues (2001) as temporal phases that are differentially important. Transition phases are those “periods of time when teams focus primarily on evaluation and/or planning activities to guide their accomplishment of a team goal or objective” (Marks et al. 2001, p. 359). Conversely, action phases are those “periods of time when teams are engaged in acts that contribute directly to goal

accomplishment (i.e., taskwork)”. Their team-process framework is based on the idea teams perform in “temporal cycles of goal-directed activity, called ‘episodes’” (Marks et al. 2001, p. 359). Using this notion, feedback can draw from this framework as a guide to discuss performance episodes at the appropriate time.

Individuals have demonstrated a preference for feedback right after their performance, which has also been found to enhance learning (e.g., Cant and Cooper 2011; Cicero et al. 2012). Health-care professionals around the world—surgeons, anesthesiologists, and nurses from the United Kingdom, United States, and Australia—have adopted this preference for immediate terminal feedback (Ahmed et al. 2012). This resembles the transition phase that is more prone of strategizing. Marks and colleagues (2001) highlight mission analysis, goal specification, and strategy formulation and planning as key processes. On the other hand, there is strong evidence that concurrent feedback does serve to immediately improve performance (Park et al. 2000; Wulf and Shea 2004). With the use of technology, videos can serve as an assessment and concurrent-feedback methodology for showing specific behaviors from the action phases (Cant and Cooper 2011). These can then have a more educational focus to develop a shared mental model of what desired performance is for the team.

Taking this together, research has begun to show there may be value in delayed feedback with regard to promoting retention (Astwood 2009), as well as concurrent positive feedback (e.g., rewards) with regard to increasing motivation (Gabelica et al. 2012). Whether the feedback is concurrent or terminal (e.g., timing), the technology (e.g., videos) and focus (motivational, educational, etc.) should be taken into consideration to fit the team’s needs. These mixed results regarding feedback timing point to an area where more research is needed.

Best Practice 11: Keep the timing of feedback in mind to maximize the benefits.

Teams learn to debrief themselves both during training and on the job. The discussions should address both positive and negative critical incidents that lead to improved plans for future actions (Salas et al. 2008b). Team members should use feedback to reflect about performance and come up with ways to improve it (Peltier et al. 2005). During discussions, it is important to talk about the direct cause of issues and weaknesses during the performance episode, rather than summarizing occurrences (Shapiro et al. 2008). Basic summarization of past events or neglecting to discuss both positive and negative aspects of performance can hinder future teamwork improvement (Smith-Jentsch et al. 2008). In turn, team members should come out of feedback sessions with an action plan that helps the improvement of performance.

Feedback should be clear, concise, and constructive, focusing on the task and not specific individuals (Shute 2008). If feedback promotes the wrong behaviors, these will be reinforced and likely to be repeated by the team. Consequently, organizations should support the use of feedback and continuously encourage the use of trained skills. This can be done through cues to remind trainees of what they learned in the team-training program, opportunities to use the trained skills, incentives for using trained skills, and feedback on the trained skills (Ford et al. 1992; Quiñones 1995; Rouiller and Goldstein 1993; Tracey et al. 1995). Hence, the work environment is a key element in determining whether the lessons learned during team feedback will be implemented on the job or not.

Best Practice 12: Foster a work environment that allows for implementation of feedback (i.e., development of an action plan).

7.4 Implementation in Team-Tutoring Environment: Critical Question

The aforementioned best practices are by no means an exhaustive list of what to consider when implementing team feedback, nor do they take into account the many complex issues regarding team training as a whole. Specifically, the team ITS provides a new avenue to deliver feedback. However, this mechanism also presents a number of challenges to the implementation of effective feedback. These challenges consist of trust development, establishment of credibility, and accurate interpretation of feedback. It is for this reason that attention should be drawn to the fundamental differences between human and virtual interaction.

Rather than finding ways to incorporate new technologies into existing team-feedback models, the focus should be on developing feedback models with specific technologies and uses in mind. Parasuraman and Riley (1997) support that individual differences and the context of tasking should be accounted for when using automated systems to replace human functions. The issues an individual encounters when automated technology is implemented can be magnified in a team setting (Paris et al. 2000). With that in mind, we identify questions that should be brought to the forefront of those implementing feedback via a team intelligent tutor.

7.5 How to Establish Credibility in the Automated Feedback Tool?

A common challenge that automated tools possess is the development and repair of trust. Nohria and Eccles (1992) suggested face-to-face interaction cannot be replaced with an electronic substitute when the purpose of communication is ambiguous, abstract, or difficult. Conversely, these authors also noted empirical

support for the assertion that electronic communication could be a successful conduit for feedback if it is addressing unambiguous, routine tasks. Though specific findings have supported the difficulty for trust-building in a virtual environment (Nohria and Eccles 1992), more research is required to adequately assess the implications for trust development in a virtual setting. However, there is initial evidence that virtual tools can increase their credibility through adding social context to the feedback (e.g., goals and knowledge). Individuals are aware when there is a virtual world and may be more likely to mask intentions or social contexts. This awareness can hinder an individual's propensity to trust a virtual medium if proper context is not provided (Burauskas and Aldama 2008). Therefore, it is important to customize the feedback tool to fit team's objectives, competence, and desired outcomes in order to increase the credibility of such tool.

7.6 Does the Type of Task Influence How Team Feedback Is Implemented?

Task complexity is particularly important in regards to the timing of feedback. For more complex tasks, concurrent feedback poses the risk of interrupting the natural flow of the team members and inhibiting learning (Astwood 2009; Gabelica et al. 2012). In highly complex tasks with a high level of interdependency, concurrent feedback can be at odds with developing an effective process-level focus. However, given the positive impact concurrent feedback has been shown to have on performance (Park et al. 2000; Wulf and Shea 2004), we must be careful not to eliminate it unless entirely warranted.

Along these lines, team intelligent tutoring has to not only adapt to the type of task members are executing, it also has to tailor the feedback system to fit in regard to content, specificity, and timing. Though empirical research has failed to investigate how timing of feedback delivery and reception impacts performance outcomes, it is worth consideration due to the extensive existing literature incorporating the time of feedback administration into the methodology of experimental designs (Dismukes et al. 2000; Smith-Jentsch et al. 2008). As aforementioned, the complexity of the task can require different types of feedback, and a deep understanding of these parameters and differences must be inputted prior to implementation of feedback in order for it to be successful.

7.7 Is One Source of Feedback Delivery Ever Enough?

Recognizing the challenges and limited resources, the feedback may only be provided by the facilitator. In these situations, we suggest eliciting input from team members whenever possible. While less formal than most team dimensional

training models, it will still encourage team members to analyze their own performance rather than becoming overly reliant on external feedback. This increases process-level learning while also avoiding potential guidance hypothesis issues (Wulf and Shea 2004). Research supports the assertion that a decrease in member satisfaction with regard to debriefing can occur when individuals feel they cannot communicate ideas or actions in opposition of the majority or the organization (Scott et al. 2013). Consequently, drawing from team members during feedback discussions seems to be a key element of successful feedback.

Within the automated tool, a similar approach can be done when the machine may be the only source of feedback available, but it can be programmed to draw from teammates' opinions and examples. This can give team members the feeling that they have a "voice" in the process (Edmondson 2003). A practical tool that has been used as a guided, team-led feedback mechanism is *DebriefNow*, developed by the Group for Organizational Effectiveness (gOE 2015). This system uses algorithms and compares the discrepancy of team members' responses to structured questions in regard to their discrepancy to highlight troublesome areas of team experiences. The promising future of this tool illustrates how the lack of a facilitator can be potentially countered by the development of the proper feedback system.

7.8 Is More Always Better?

As aforementioned, the reliance on concurrent feedback—in comparison to only terminal feedback—is definitely growing. However, feedback can negatively impact performance when the learner develops a dependency or complete reliance on instructor evaluation, such that they cannot perform adequately in the absence of feedback. This takes place when feedback from the instructor is provided with excessive frequency. Central to this debate is the "guidance hypothesis"; Khan and Franks (2004, p. 58) define it as "providing frequent feedback [that] would cause performers to be reliant on this information and prevent the processing of other sources of feedback". Hence, while the immediate performance benefits of concurrent feedback are undeniable, we must avoid issues with team's long-term retention.

Properly facilitated debriefing (i.e., feedback) sessions can lead to a more in-depth evaluation of performance and encouraged higher levels of engagement in team members (Dismukes et al. 2000). Therefore, we continue to advocate the relevance and benefits of team feedback. However, we point out the caveat that there can be too much feedback (Goodman 1998). Consequently, one should pay attention regarding the timing of feedback as constant and explicit feedback may lead to counterproductive outcomes.

8. Workshop Summary

In March 2016, the University of Central Florida, Rice University, and ARL's Human Research and Engineering Directorate jointly hosted a workshop, "Building an Intelligent Tutoring System for Teams: What Matters". The purpose of this multiday workshop was to bring together academic and government researchers, practitioners, and government sponsors to explore knowledge gained with regard to the unique requirements and challenges in developing ITSs for teams as well as highlight some potential ways forward. In doing so, we sought to foster collaborative discussions among researchers, instructional designers, and ITS design engineers. With this interdisciplinary approach it is our goal to foster 1) critical thinking about the challenging endeavor of creating team ITSs, 2) cross-fertilization of knowledge across disciplines, and 3) actionable objectives for future research.

Twenty-five individuals attended the workshop and participants were multidisciplinary, including but not limited to human-factors psychologists, industrial/organizational psychologists, mechanical and software engineers, cognitive psychologists, and education-management specialists. Participants also came from a range of organizations. The following organizations were represented at the workshop:

- University of Central Florida
- Rice University
- ARL
- Wyle Science
- Technology and Engineering Group
- Iowa State University
- Aptima
- Georgia Tech
- Embry Riddle University
- Dignitas
- Educational Testing Service
- University of Wisconsin–Madison
- IMMEX Publications

- US Naval Air Warfare Center–Training Systems Division
- Chi Systems

The structure of the workshop was such that participants heard a series of presentations that pertained to unpacking the domain and pedagogical modules for teams, populating the dynamic assessment module for teams, and providing feedback to teams in the context of ITSs. These primary presentations will pave the way for break-out groups designed to foster discussion around critical areas and gaps in knowledge. The breakout sessions were organized into four themes that tied back into the presentations that were given earlier in the day: 1) competency module, 2) pedagogical module content/design, 3) dynamic assessment, and 4) delivering feedback in a team ITS. Next, in Sections 8.1–8.4 we briefly describe the output of each of these sessions. The first breakout group was tasked with discussion of a series of questions and challenges that relate to information within the competency module.

8.1 Breakout Session 1: Team Competency

The primary questions that were addressed by this breakout group dealt with the competency module and revolved around three areas: 1) identifying the top five challenges in populating the competency module, 2) identifying research needs, and 3) identifying those communities that need to be talking in order to move research and practice forward.

The group identified the following as the top five challenges: 1) task interdependency, 2) development of an expert model of teamwork, 3) decisions regarding the degree to which to promote self-regulation was providing feedback within the ITS, 4) ensuring the authoring of the ITS is tied to the assessment, and 5) team roles.

With respect to the type of research that is needed to move this area forward, the group identified five primary areas. First, is the need for longitudinal data. Most of what is known about teams is based on teams that are performing for short periods of time and are often ad hoc in nature. Therefore, little is truly known in terms of how specific emergent states and processes develop and change over time; most of the work in this area that incorporates any notions of temporality is theoretical with a few exceptions. Second, related to the need for longitudinal research, is the need for evidence-based design. This relates to ensuring that the content within the domain module is built based on systematically derived scientific evidence. While there has been much progress made in the area of teams over the last 50 years, there still remain many unanswered questions, including but not limited to the impact of

temporal characteristics and the interrelationship among particular sets of team processes (e.g., which processes account for the predominant amount of variance when pitted against one another). As a field rarely is there an examination of the relative importance of the identified behavioral processes with respect to one another in predicting different team outcomes—the same can be said of the cognitive and affective states shown to be important in teams. Nor do experiments typically look across behavior, cognitive, and affective factors to see which specific combinations predict the most variance in team outcomes; instead, there is a tendency to focus at a more microlevel on the relative contribution of a smaller set of predictors (whether it be affective, behavioral, or cognitive). The third area of research that was identified was team-task analysis. This research need relates to the actual need to conduct a team-task analysis to identify within a particular domain what are the KSAOs that are transportable versus those that are team or task specific. While theory can help (see Salas et al. 2009; Cannon-Bowers et al. 1995), it does not provide the complete picture. Oftentimes team-training systems are created without a systematic team-task analysis, thereby leaving the possibility of training targeting the needed KSAOs. The fourth area that was identified dealt with research to determine which KSAOs are most amenable to integrate within an ITS for teams. In other words, where are the low-hanging fruit, the ones that might be most easily integrated and taught within the context of an automated system?

Finally, in terms of the communities that need to be involved in discussions in order to tackle the challenges, the group identified psychologists, computer scientists, engineers, linguists, mathematicians, and anthropologists. The breakout group specified this list is by no means all-inclusive but represents those which may be most prominent in these discussions.

The group also discussed which team characteristics should be accounted for prior to populating the domain module for use by a specific team. The team characteristics that were identified included the following: degree and type of interdependence, cultural orientation (e.g., power distance and individualism/collectivism), leadership structure (e.g., collective or hierarchical), and team type (e.g., action, management, knowledge, product, and development).

There was also discussion of what would be needed within the domain module so that someone who is not a team's expert could come in and pick the most efficacious content for their specific situation. Some of the discussion revolved around the need for a list of generalizable and prioritized skills, an effective delivery method, and precise assessment.

8.2 Breakout Session 2: Team Pedagogy

The second breakout group was tasked with discussing issues related to the content within the pedagogical model (and associated challenges). The group spent the predominant amount of its time discussing the top six challenges in populating and designing the content within each of the pedagogical models. The group first discussed how team training presents many of the same challenges as individual training. However, the team-training environment is additionally hard to control as it quickly gets complex and chaotic. Second, we often do not know how to best measure what needs to be measured. Within a team-training environment it becomes more difficult to detect, assess, and mitigate due to the complexity and number of moving interdependent parts. We often do not know how to best measure what needs to be measured within such complex environments (although a variety of methods are available). Nor are we particularly effective at simulating the complexity of “real life” social interactions in such a way as to truly scaffold and facilitate training effectiveness. Related to social interactions is the fact we currently need live people together for the best experience, which is costly in terms of resources (time and money).

A third challenge pertains to the extent to which you can train part versus whole aspects of teamwork. In this vein, the group talked about one way to accomplish this is through the experimental control of the simulation by some of the participants (e.g., confederates). Another way is through quasi-experimental control whereby you can account for what the participants have received and respond accordingly.

A fourth challenge that was discussed is how to account for constraints on scenario design that can push people toward learning. The group discussed how this approach makes the whole thing tractable so that people can train on a particular skill “x”. They discussed considerations with respect to how to open the scenario up enough, but also being able to extract what we need to measure—the challenge of walking that balance.

A fifth challenge discussed is how to define a good policy for feedback and feed forward within the system; specifically, with respect to the timing, content, and who gets it. Related to feedback is the question of how to detect the behaviors that need to improve. In this vein, there are decisions regarding taskwork and teamwork. In addition, another challenge is how to constrain the scenario or event so that the chance to observe exists. Teams often engage in implicit coordination whereby the explicit actions to observe are limited. These challenges led the group to identify a need for models that integrate structural dynamics of team and task dynamics—network phenomena that need integration. Additionally, there was a call for the

exploration of psychological and neurological measures to help understand team state and team behaviors over time.

A sixth, and final, challenge discussed by the group is the identification of the analytic and experimental technologies that are needed to be built into an ITS to develop the answers obtained. Models and assessments need to model the complexity of the task environment. In this vein, perhaps a measurement tool would be nonlinear and dynamical versus more traditional linear methods—one example might be the push toward network-analysis tools. Additionally, there is a need for the ITS to manipulate scenarios in different ways to determine the bounds on team behavior. What are the variables (moderator or mediators) that change team behavior? Once identified, there is a need for systems that can vary those bounds. On a related note is the need to understand the impact of team size and type of team member. Lastly, the group mentioned the importance of being able to assess the team state well enough to assess utility and drive feedback.

8.3 Breakout Group 3: Team Assessment

The third breakout group was tasked with discussing questions related to dynamic assessment. The first topic discussed was challenges to team assessment. The group identified several challenges, including 1) a need for longer-term assessment; 2) an understanding of the nature of interdependence; 3) an understanding of the various collaboration types; 4) building up a knowledge set for ITSs; 5) how to ensure measures are not so context specific that they cannot be used for transportable skills; 6) the need for systematically shared measures; 7) incorporation of multimodality assessment (verbal and nonverbal); and 8) the noisy nature of the environment.

Some of these challenges and needs the group talked about in more specificity. For example, one of the challenges mentioned for team assessment is the need to for longer-term assessment so that we have a better understanding of how teams develop over time and the importance of various team constructs over time and how these constructs may/may not change over time. However, it is difficult to collect team data and collecting team data over long periods becomes even more difficult. In this vein the group mentioned the sample-size requirements needed for a stable analysis and how this often becomes difficult when examining teams over time. There is often an inherent trade-off in length of time studied and ease with which sample-size requirements are met, although some of the newer statistical techniques can assist to some extent. Another challenge that was discussed was the need for better understanding of the nature of interdependence. More specifically, in terms of assessment, how does the degree and type of interdependence impact the

members who should be assessed on the team (e.g., dyadic pairs or entire team), and when is the most efficacious time to assess? There was also discussion of the need to understand the various types of collaboration and what this means for team assessment; specifically, the need to better understand the interaction between team composition and collaboration style on performance, as well as the notion of “fake collaboration” that has been seen in some of the educational domains. The other challenge that was discussed at great length was the noisy nature of the environment and the impact it has on team assessment. Some other topics: the fact that within teams things are not always as they seem on the surface (e.g., implicit vs. explicit coordination); determining the periods of performance that make sense to provide feedback (e.g., what are your thresholds of when to intervene); the use of accountable talk as a potential way forward; and that interpretation of assessment is often context dependent. This later point led to discussions on how to deliver feedback in context; understanding what algorithms make sense to use and where they do not; and the notion that sometimes the system can only be used for a portion of the assessment—that the system may feedback the data but there may be a need for human interpretation given the many moving parts that may affect data interpretation and sense-making. Finally, with respect to the challenge of building up the knowledge set for the ITS, the group discussed the need to determine which patterns are the same across tasks, contexts, and domains. Systematic research is needed in this area. However, very closely related was the need for systematically shared measures. Currently in team research there are many measures and each researcher uses his/her “pet” measure, which may make it hard to compare across studies. Therefore, there has been a recent push for systematically shared measures. This is a culture shift, though, and will take time and mandates to facilitate.

The group also talked about how to move beyond self-report questionnaires; more specifically, which emerging technologies may facilitate movement toward less-obtrusive measures. Technologies that were mentioned included sociometric badges, commercial off-the-shelf and SepherTech bioharness/physiological measures, Kinect automated video analysis, linguistic analysis, dedicated instrumented spaces and embedded measurement within performance tasks, facial-recognition software, social-network analysis, and vocal intensity. However, in moving toward some of these more unobtrusive technologies there are inherent challenges. For example, with regard to sociometric badges there is a level of expertise needed to interpret the data that is different from the expertise most team researchers currently have. With respect to linguistic analysis, challenges are faced in terms of the volume of data needed to train the system and the resources to transcribe the data given current state-of-the art vocal recognition (transcription is needed to train the system). Additionally, clear communication data can be difficult

to obtain in noisy environments where there is often cross-talk among team members.

The group also discussed how unobtrusive measures can sometimes cause unexpected IRB concerns—that via unobtrusiveness the researcher might inadvertently capture something unintended. Moreover, that the degree to which measures are perceived “unobtrusive” may vary based on how long participants are expected to wear the devices and the environment in which they are operating. For example, at first eye-trackers may seem obtrusive but, generally, after wearing them for a short while participants often forget they are on and they become essentially “unobtrusive”. Other factors that may impact the degree to which measures are seen as obtrusive pertain to how meaningful participants feel the information is that is being collected. For example, members of the group mentioned that while surveys are often seen as obtrusive, participants’ feedback indicates that when surveys are short and are seen as meaningful, participants actually perceive them to be less intrusive and request to take them more frequently.

Another topic of discussion dealt with determining the timing of assessment; more specifically, how to determine when to inject measures. The group ended up breaking this into two large buckets: one regarding the acceptability of the timing of the measure and the second based on information gleaned from the scientific literature. With regard to the science the following were mentioned as potential sources of guidance: 1) work on team development, 2) natural points of reflection within the team’s life span and/or task, 3) work dictating how often you need to measure to have reliable results (e.g., immediately before a critical event or after), 4) domain modeling to provide insight into the critical points and subdimensions, 5) the purpose of assessment (i.e., training and assessment), 6) turnover rate, and 7) work emerging from the community of NASA researchers suggesting it may take up to 45 d for emergent team states to fully develop. From a more practical standpoint, where concern revolves around the acceptability of the measurement timing by participants, the following were mentioned as guiding forces: piloting, life span of team, organizational culture, user versus organizational control, and degree of user control. With respect to the timing of measurement, it was discussed that the length of time that the team has been together or will be together on the task of interest may interact to impact the frequency with which they are amenable to be assessed and provide quality responses. Organizational culture also has been known to play a role in the degree to which participants accept not only being assessed, but in terms of the frequency with which they are willing to accept (i.e., provide good responses to). Guidance can also be seen in terms of the user-control literature. Specifically, the group talked about ways to increase user control as compared to it being perceived that everything was under the organization’s control

with little user input into how often assessment occurred. For example, some suggestions included having a window of time that participants can respond rather than an exact time and allowing participants to choose from multiple formats in which to respond (e.g., video based; self-report, Likert-style paper and pencil; or journaling). In essence, this would be having multiple equivalent formats for the assessment of any given construct, thereby allowing the user to pick the format they prefer.

8.4 Breakout Group 4: Team Feedback

The last breakout group was tasked with discussing questions revolving around the delivery of feedback within an ITS.

The group began its discussions with a number of assumptions in place. More specifically, the group members identified seven assumptions they were operating with throughout the discussion. First, with respect to the composition of the team, it was to be live versus synthetic. Second, the initial focus would be with respect to cognitive tasks. Third, team roles would be interdependent and teams would be intact as compared to ad hoc. Fourth, during the task any feedback that is given cannot contribute significantly to cognitive load; however, the feedback needs to have sufficient complexity that it requires an ITS plus feedback. Fifth, there are limited learning objectives at any given time to reduce the cognitive load and assessments are most likely performance based. Sixth, assessments should be individual and team based and preference would be for them to be more state like (e.g., mastery). Finally, the tutor is providing all of the feedback.

Once these assumptions were identified, the group began to discuss what the members perceived as the top five challenges to delivering feedback in an ITS. One of the challenges concerns the ability to segregate individual and team feedback so that both are able to be given, but done so separately as needed. A second challenge deals with issues pertaining to the management of feedback: timing, type of feedback, delivery of feedback, and priority determination. Additionally, there are questions with respect to the timing of predictive versus reactive feedback. In terms of delivery, questions exist with respect to the most appropriate modality and the delivery of architectural feedback. Then, finally, with regard to prioritization of feedback there are challenges with respect to the deconfliction and contextualization. A third group of challenges identified by the group relate to targeting and achieving effects; more specifically, determining exactly what effect is being sought and the identification of the best way to express the effect. Related are issues pertaining to how to adapt feedback to roles and responsibilities. A fourth group of challenges relates to the ability to integrate and adapt feedback to deep-

assessment data. This adaptation should be based on things such as the role, individual, team, history, and environment and pedagogical strategy need. The final challenge that the group discussed pertains to the presentation of feedback. Issues discussed included the determination of channels, strategies, mode conflict, and sensitivity to potential saturation.

There was also discussion related to the ITS's architecture. Within this architecture there is a need for content and data threaded through learning/educational objectives. Also within the architecture is the need for feedback to have deep access to assessment data and to delivery environment (not just a pipeline to a "feedback display"). Finally, the architecture needs to support dynamic, pedagogically based scenario control and modification.

The group also discussed the translation of best practices. In this vein, it was recommended that individuals should read Goldberg and Cannon-Bowers (2015). The need to generate a taxonomy of problem types/characteristics and feedback options was also discussed. Once this taxonomy is created there is a need to generate generic algorithms for each. There is a recognition that one size does not fit all and there are many ways to get to a possible solution (e.g., "more than one way to skin a cat").

9. Conclusions

Overall, the presentations given throughout the workshop served to highlight the state of the field's knowledge with regard to 1) the particular team knowledges, skills, and attitudes that may be relevant to train within an ITS whose objective was team development; 2) types of pedagogical techniques that have proven useful in the past for training teams and discussions of the challenges in terms of integrating these into an ITS; 3) methodologies that are currently used for team assessment, the degree to which these may lend themselves to dynamic assessment, and emerging technologies that may facilitate closer to "real time" assessment; and 4) guidance and challenges pertaining to the delivery of team feedback within an ITS as well as highlighting ongoing work in this area. While many themes emerged, as can be seen through the summaries of the breakout groups, key among these were the complexity in moving from an individual-focused ITS to one that is focused on teams. While there are complexities that come into play across the board for each of the aspects discussed in the breakout groups, perhaps the largest area of continual discussion was the notion of team assessment and the delivery of the resulting feedback within the context of an ITS—adding the notion of wanting near real-time assessment, which suggests unobtrusive and automated measures of team process becomes even more complex. Many of the complexities discussed highlight the

need for additional research; however, there is also a subset of challenges that were purely at a practical level—for example, how to design the system such that it is adaptable to the individual needs of its users without requiring that these users be experts in team training. Even within these more practical questions there is often a need for additional research. For example, a diagnostic tool could be built on the front end of the system whereby it would use the science of teams to guide the user through a series of questions that would then point them to particular team KSAs that are lacking. Additional meta-analytic and qualitative analysis could also be conducted to examine whether particular pedagogical methods commonly used for teams are more efficacious depending on the nature of the team construct to be targeted (i.e., behavior, cognition, attitude, or combination). The productive discussions that occurred during the workshop not only highlighted the exciting and complex nature of the eventual goal of designing GIFT for teams, but provided a way forward along with areas of potential collaboration across disciplines to maximize the synergy in moving the vision forward.

10. References

- Ahmed M, Sevdalis N, Paige J, Paragi-Gururaja R, Nestel D, Arora S. Identifying best practice guidelines for debriefing in surgery: a tri-continental study. *American J Surg.* 2012;203:523–529.
- Aleven V, McLaren B, Roll I, Koedinger K. Toward meta-cognitive tutoring: a model of help seeking with a cognitive tutor. *Int J Art Intel Ed.* 2006;16(2):101–128.
- Andreatta PB. A typology for health care teams. *Health Care Man Rev.* 2010;35(4):345–354.
- Appelgren A, Penny W, Bengtsson SL. Impact of feedback on three phases of performance monitoring. *Exp Psych.* 2014;61(3):224–233.
- Arafeh JM, Hansen SS, Nichols A. Debriefing in simulated-based learning: facilitating a reflective discussion. *J Peri Neo Nurs.* 2010;24(4):302–309.
- Astwood RS Jr. Effective timing of feedback during scenario based team training within a simulated environment [unpublished dissertation]. [Orlando (FL)]: University of Central Florida; 2009.
- Baker RS, D’Mello SK, Rodrigo MMT, Graesser AC. Better to be frustrated than bored: the incidence, persistence, and impact of learners’ cognitive–affective states during interactions with three different computer-based learning environments. *Int J Human-Comp Stud.* 2010;68(4):223–241.
- Baldwin TT, Ford JK. Transfer of training: a review and directions for future research. *Pers Psych.* 1988;41(1):63–105.
- Bandura A. Editorial. *American J Health Prom.* 1997;12(1):8–10.
- Barker J, Tjosvold D, Andrews IR. Conflict approaches of effective and ineffective project managers: a field study in a matrix organization. *J Man Stud.* 1988;25(2):167–178.
- Barrick MR, Stewart GL, Neubert MJ, Mount MK. Relating member ability and personality to work-team processes and team effectiveness. *J App Psych.* 1998;83(3):377–391.
- Bartram D, Roe RA. Individual and organisational factors in competence acquisition. In: Nijhof WJ, editor. *The learning potential of the workplace.* Rotterdam, Netherlands: Sense Publishers; 2008. p. 71–96.

- Beal DJ, Cohen RR, Burke MJ, McLendon CL. Cohesion and performance in groups: a meta-analytic clarification of construct relations. *J App Psych*. 2003;88(6):989–1004.
- Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? *Qual Safe Health Care*. 2004;13(1):i51–i56.
- Blickensderfer E, Cannon-Bowers JA, Salas E. Theoretical bases for team self-correction: fostering shared mental models. *Adv Interdis Stud Work Teams*. 1997;4:249–279.
- Bowers C, Smith PA, Cannon-Bowers J, Nicholson D. Using virtual worlds to assist distributed teams. In: Zemliansky P, St. Amant K, editors. *Handbook of research on virtual workplaces and the new nature of business practices*. Hershey (PA): Information Science Reference; 2008. p. 408–423.
- Brannick MT, Salas E, Prince C, editors. *Team performance assessment and measurement: theory, methods, and applications*. Mahwah (NJ): Lawrence Erlbaum Associates; 1997.
- Bray SR. Collective efficacy, group goals, and group performance of a muscular endurance task. *Small Group Res*. 2004;35(2):230–238.
- Bunderson JS, Sutcliffe KM. Comparing alternative conceptualizations of functional diversity in management teams: process and performance effects. *Academy Man J*. 2002;45(5):875–893.
- Burauskas G, Aldama JI. Trust in virtual communities [master's thesis]. [Lund, Sweden]: Lund University; 2008.
- Burke CS, Stagl KC, Klein C, Goodwin GF, Salas E, Halpin SM. What type of leadership behaviors are functional in teams? A meta-analysis. *Leader Quart*. 2006;17:288–307.
- Buykx P, Kinsman L, Cooper S, McConnell-Henry T, Cant R, Endacott R, Scholes J. Educating nurses to identify patient deterioration—a theory-based model for best practice simulation education. *Nurse Ed Today*. 2011;31(7):687–693.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psych Bull*. 1959;56(2):81–105.
- Cannon-Bowers JA, Bowers C, Procci K. Optimizing learning in surgical simulations: guidelines from the science of learning and human performance. *Surg Clin N America*. 2010;90:583–603.

- Cannon-Bowers JA, Bowers C. Team development and functioning. In: Zedeck S, editor. *APA handbook of industrial and organizational psychology, vol 1: building and developing the organization*. Washington (DC): American Psychological Association; 2011. p. 597–650.
- Cannon-Bowers JA, Tannenbaum SI, Salas E, Volpe CE. *Defining competencies and establishing team training requirements*. Hoboken (NJ): Wiley and Sons; 1995. *Team effectiveness and decision making in organizations*; p. 333–380.
- Cant RP, Cooper SJ. The benefits of debriefing as formative feedback in nursing education. *Australian J Adv Nurs*. 2011;29(1):37–47.
- Carless SA, De Paola C. The measurement of cohesion in work teams. *Small Group Res*. 2000;31(1):71–88.
- Carron AV, Brawley LR. Cohesion conceptual and measurement issues. *Small Group Res*. 2000;31(1):89–106.
- Carron PN, Trueb L, Yersin B. High-fidelity simulation in the nonmedical domain: practices and potential transferable competencies for the medical field. *Adv Med Ed Prac*. 2011;2:149–155.
- Carron AV, Widmeyer WN, Brawley LR. The development of an instrument to assess cohesion in sport teams: the group environment questionnaire. *J Sport Psych*. 1985;7(3):244–266.
- Chan D. Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *J App Psych*. 1998;83(2):234–246.
- Chang A, Bordia P. A multidimensional approach to the group cohesion-group performance relationship. *Small Group Res*. 2001;32(4):379–405.
- Chang JW, Sy T, Choi JN. Team emotional intelligence and performance: interactive dynamics between leaders and members. *Small Group Res*. 2012;43(1):75–104.
- Chen G, Thomas B, Wallace JC. A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *J App Psych*. 2005;90(5):827–841.
- Cheng HH, Yang HL. Student team projects in information systems development: measuring collective creative efficacy. *Australasian J Ed Tech*. 2011;27(6):881–895.

- Choo AS, Linderman KW, Schroeder RG. Method and psychological effects on learning behaviors and knowledge creation in quality improvement projects. *Man Sci.* 2007;53(3):437–450.
- Chronister C, Brown D. Comparison of simulation debriefing methods. *Clin Sim Nurs.* 2017;8(7):e281–e288.
- Cicero MX, Auerbach MA, Zigmont J, Riera A, Ching K, Baum CR. Simulation training with structured debriefing improves residents' pediatric disaster triage performance. *Prehos Dis Med.* 2012;27(03):239–244.
- Cohen A, Doveh E, Eick U. Statistical properties of the rWG (J) index of agreement. *Psych Meth.* 2001;6(3):297–310.
- Cook J, Wall T. New work attitude measures of trust, organizational commitment and personal need non-fulfilment. *J Occup Psych.* 1980;53(1):39–52.
- Costa AC. Work team trust and effectiveness. *Pers Rev.* 2003;32(5):605–622.
- Cox KB. The effects of intrapersonal, intragroup, and intergroup conflict on team performance effectiveness and work satisfaction. *Nurs Admin Quart.* 2003;27(2):153–163.
- Craig TY, Kelly JR. Group cohesiveness and creative performance. *Group Dyn Theory Res Prac.* 1999;3(4):243–256.
- Cronin MA, Bezrukova K, Weingart LR, Tinsley CH. Subgroups within a team: the role of cognitive and affective integration. *J Org Behav.* 2011;32(6):831–849.
- DeChurch LA, Mesmer-Magnus JR. Measuring shared team mental models: a meta-analysis. *Group Dyn.* 2010;14:1–14
- De Dreu CKW. Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective. *J App Psych.* 2007;92(3):628–638.
- De Dreu CK, Weingart LR. Task versus relationship conflict, team performance, and team member satisfaction: a meta-analysis. *J App Psych.* 2003;88(4):741–749.
- De Jong BA, Elfring T. How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Acad Man J.* 2010;53(3):535–549.
- DeNisi AS. Performance appraisal and performance management: a multilevel analysis. In: Klein KJ, Kozlowski SWJ, editors. *Multilevel theory, research,*

- and methods in organizations: foundations, extensions, and new directions. San Francisco (CA): Jossey-Bass; 2000. p. 121.
- DeShon RP, Kozlowski SWJ, Schmidt AM, Milner KR, Wiechmann D. A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *J App Psych.* 2004;89(6):1035–1056.
- de Wit FR, Greer LL, Jehn KA. The paradox of intragroup conflict: a meta-analysis. *J App Psych.* 2012;97(2):360–390.
- Dillenbourg P. What do you mean by ‘collaborative learning’? Oxford (UK): Elsevier; 1999. Collaborative-learning: cognitive and computational approaches; p. 1–15.
- Dirks KT. Trust in leadership and team performance: evidence from NCAA basketball. *J App Psych.* 2000;85(6):1004–1012.
- Dismukes RK, Jobe KK, McDonnell Lori K. LOFT debriefings : an analysis of instructor techniques and crew participation. 1997. NASA technical memorandum.
- Dismukes RK, McDonnell LK, Jobe KK. Facilitating LOFT debriefings: instructor techniques and crew participation. *Inter J Av Psych.* 2000;10(1):35–57.
- D’Mello SK, Graesser A, King B. Toward spoken human–computer tutorial dialogues. *Human–Comp Int.* 2010;25(4):289–323.
- Dreifuerst KT. The essentials of debriefing in simulation learning: a concept analysis. *Nurs Ed Pers.* 2009;30(2):109–114.
- Dreifuerst KT. Debriefing for meaningful learning: fostering development of clinical reasoning through simulation [unpublished dissertation]. [Bloomington (ID)]: Indiana University; 2010.
- Dreifuerst KT. Using debriefing for meaningful learning to foster development of clinical reasoning in simulation. *J Nurs Ed.* 2012;51(6):326–333.
- Driskell JE, Salas E. Group decision making under stress. *J App Psych.* 1991;76(3):473–478.
- Driskell JE, Salas E, Hughes S. Collective orientation and team performance: development of an individual differences measure. *Hum Fact.* 2010;52(2):316–328.
- Edmonds WA, Tenenbaum G, Kamata A, Johnson MB. The role of collective efficacy in adventure racing teams. *Small Group Res.* 2009;40(2):163–180.

- Edmondson AC. Psychological safety and learning behavior in work teams. *Admin Sci Quart.* 1999;44(2):350–383.
- Edmondson AC. Speaking up in the operating room: how team leaders promote learning in interdisciplinary action teams. *J Man Stud.* 2003;40(6):1419–1452.
- Edmondson AC, Bohmer RM, Pisano GP. Disrupted routines: team learning and new technology implementation in hospitals. *Admin Sci Quart.* 2001;46(4):685–716.
- Elson-Cook M. Student modeling in intelligent tutoring systems. *Art Intel Rev.* 1993;7:227–240.
- Espinosa JA, Cummings JN, Pickering C. Time separation, coordination, and performance in technical teams. *IEEE Trans Eng Man.* 2012;59(1):91–103.
- Evans CR, Dion KL. Group cohesion and performance: a meta-analysis. *Small Group Res.* 1991;22(2):175–186.
- Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Sim Health.* 2007;2(2):115–125.
- Faraj S, Sproull L. Coordinating expertise in software development teams. *Man Sci.* 2000;46(12):1554–1568.
- Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *American Soc Rev.* 1978;43(4):557–572.
- Fletcher TD, Major DA. The effects of communication modality on performance and self-ratings of teamwork components. *J Comp Med Comm.* 2006;11(2):557–576.
- Fletcher JD, Sottilare RA. Shared mental models of cognition for intelligent tutoring of teams. In: Sottilare R, Graesser A, Hu X, Holden H, editors. *Design recommendations for intelligent tutoring systems—volume 1: learner modeling.* Orlando (FL): Army Research Laboratory (US); 2013. p. 237–252.
- Ford JK, Quiñones MA, Sego DJ, Sorra JS. Factors affecting the opportunity to perform trained tasks on the job. *Pers Psych.* 1992;45(3):511–527.
- Gabelica C, Bossche PVD, Segers M, Gijssels W. Feedback, a powerful lever in teams: a review. *Ed Res Rev.* 2012;7(2):123–144.
- Gajendran RS, Joshi A. Innovation in globally distributed teams: the role of LMX, communication frequency, and member influence on team decisions. *J App Psych.* 2012;97(6):1252–1261.

- Ginnett RC. Crews as groups: their formation and their leadership. *Crew Res Man.* 2110;79–110.
- Gladstein DL. Groups in context: a model of task group effectiveness. *Admin Sci Quart.* 1984;29(4):499–517.
- [gOE] The Group for Organizational Effectiveness. Products. 2015 [accessed 14 January 2019]. <http://www.groupoe.com/products.html>.
- Goldberg B, Cannon-Bowers J. Feedback source modality effects on training outcomes in a serious game: pedagogical agents make a difference. *Comp Human Behav.* 2015;52:1–11.
- Goldberg BS, Sottolare RA, Brawner KW, Holden HK. Predicting learner engagement during well-defined and ill-defined computer-based intercultural interactions. In: D’Mello S, Graesser A, Schuller B, Martin JC, editors. *International Conference on Affective Computing and Intelligent Interaction. ACII 2011. Lecture notes in computer science*, vol. 6974. Berlin, Germany: Springer. 2011; p. 538–547.
- Gomes E, Weber Y, Brown C, Tarba SY. *Mergers, acquisitions, and strategic alliances*. Basingstoke (UK): Palgrave Macmillan; 2011.
- Goodman JS. The interactive effects of task and external feedback on practice performance and learning. *Org Behav Hum Dec Proc.* 1998;76(3):223–252.
- Graesser AC. Conversations with AutoTutor help students learn. *Inter J Art Intel Ed.* 2016;26(1):124–132.
- Graesser AC, Hu X, Sottolare R. Intelligent tutoring systems. In: Fischer F, Hmelo-Silver CE, Goldman SR, Reimann P, editors. *International handbook of the learning sciences*. New York (NY): Routledge; 2018. p. 246–255.
- Grant JS, Moss J, Epps C, Watts P. Using video-facilitated feedback to improve student performance following high-fidelity simulation. *Clin Sim Nurs.* 2010;6(5):177–184.
- Gully SM, Devine DJ, Whitney DJ. A meta-analysis of cohesion and performance effects of level of analysis and task interdependence. *Small Group Res.* 1995;26(4):497–520.
- Gully SM, Incalcaterra KA, Joshi A, Beaubien JM. A meta-analysis of team-efficacy, potency, and performance: interdependence and level of analysis as moderators of observed relationships. *J App Psych.* 2002;87(5):819–832.

- Gum L, Greenhill J, Dix K. Sim TRACT a reflective conceptual framework for simulation debriefing. *J Transfor Ed*. 2011;9(1):21–41.
- Gupta A, Mattarelli E, Seshasai S, Broschak J. Use of collaborative technologies and knowledge sharing in co-located and distributed teams: towards the 24-h knowledge factory. *J Strat Info Systems*. 2009;18(3):147–161.
- Gupta VK, Huang R, Niranjana S. A longitudinal examination of the relationship between team leadership and performance. *J Leader Org Studies*. 2010;17(4):335–350.
- Guzzo RA, Yost PR, Campbell RJ, Shea GP. Potency in groups: articulating a construct. *British J Soc Psych*. 1993;32(1):87–106.
- Haas MR. Acquiring and applying knowledge in transnational teams: the roles of cosmopolitans and locals. *Org Sci*. 2006;17(3):367–384.
- Hackman JR, Wageman R. A theory of team coaching. *Acad Man Rev*. 2005;30(2):269–287.
- Hattie J, Timperley H. The power of feedback. *Rev Ed Res*. 2007;77(1):81–112.
- Hawley JK, Mares AL. Developing effective adaptive missile crews and command and control teams for air and missile defense systems. Fort Bliss (TX): Army Research Laboratory (US); 2007 Mar. Report No.: ARL-SR-149.
- Helstad K, Lund A. Teachers' talk on students' writing: negotiating students' texts in interdisciplinary teacher teams. *Teach Teacher Ed*. 2012;28(4):599–608.
- Henry KB, Arrow H, Carini B. A tripartite model of group identification: theory and measurement. *Small Group Res*. 1999;30(5):558–581.
- Hinsz VB, Tindale RS, Vollrath DA. The emerging conceptualization of groups as information processors. *Psych Bull*. 1997;121(1):43–64.
- Hirst G. Effects of membership change on open discussion and team performance: the moderating role of team tenure. *European J Work Org Psych*. 2009;18(2):231–249.
- Hobman EV, Bordia P. The role of team identification in the dissimilarity-conflict relationship. *Group Proc Intergroup Rel*. 2006;9(4):483–507.
- Hoegl M, Gemuenden HG. Teamwork quality and the success of innovative projects: a theoretical concept and empirical evidence. *Org Sci*. 2001;12(4):435–449.

- Holmquist M. Managing project transformation in a complex context. *Creat Inno Man.* 2007;16(1):46–52.
- Hsu MH, Chen IYL, Chiu CM, Ju TL. Exploring the antecedents of team performance in collaborative learning of computer software. *Comp Ed.* 2007;48(4):700–718.
- Huang JC. Unbundling task conflict and relationship conflict: the moderating role of team goal orientation and conflict management. *Inter J Conf Man.* 2010;21(3):334–355.
- Huang S, Cummings JN. When critical knowledge is most critical centralization in knowledge intensive teams. *Small Group Res.* 2011;42(6):669–699.
- Ilgen DR, Fisher CD, Taylor MS. Consequences of individual feedback on behavior in organizations. *J App Psych.* 1979;64(4):349–371.
- Issenberg BS, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach.* 2005;27(1):10–28.
- James LR, Demaree RG, Wolf G. Estimating within-group interrater reliability with and without response bias. *J App Psych.* 1984;69(1):85–98.
- Jehn KA. A multimethod examination of the benefits and detriments of intragroup conflict. *Admin Sci Quart.* 1995;40(2):256–282.
- Jehn KA, Chadwick C, Thatcher SM. To agree or not to agree: the effects of value congruence, individual demographic dissimilarity, and conflict on workgroup outcomes. *Inter J Conf Man.* 1997;8(4):287–305.
- Jehn KA, Greer L, Levine S, Szulanski G. The effects of conflict types, dimensions, and emergent states on group outcomes. *Group Dec Neg.* 2008;17(6):465–495.
- Jehn KA, Mannix EA. The dynamic nature of conflict: a longitudinal study of intragroup conflict and group performance. *Acad Man J.* 2001;44(2):238–251.
- Johns JH, Bickel MD, Blades AC, Creel JB, Gatling WS, Hinkle JM, Kindred JD, Stocks SE. Cohesion in the US military. Washington (DC): National Defense University; 1984.
- Johnson WL, Lester JC. Face-to-face interaction with pedagogical agents, twenty years later. *Inter J Art Intel Ed.* 2016;26(1):25–36.
- Jones GR. Socialization tactics, self-efficacy, and newcomers' adjustments to organizations. *Acad Man J.* 1986;29(2):262–279.

- Jong AD, Ruyter KD, Lemmink J. Service climate in self-managing teams: mapping the linkage of team member perceptions and service performance outcomes in a business-to-business setting. *J Man Stud.* 2005;42(8):1593–1620.
- Jung D, Sosik J. Group potency and collective efficacy: examining their predictive validity, level of analysis, and effects of performance feedback on future group performance. *Group Org Man.* 2003;28:366–391.
- Katz-Navon T, Erez M. When collective and self-efficacy affect team performance, the role of task interdependence. *Small Group Res.* 2005;36:437–465.
- Kerr NL, Tindale RS. Group performance and decision making. *Ann Rev Psych.* 2004;55:623–655.
- Khan MA, Franks IM. The utilization of visual feedback in the acquisition of motor skills. In: Williams AM, Hodges NJ, editors. *Skill acquisition in sport: research, theory and practice.* London, England: Routledge; 2004. p. 45–62.
- Klaus DJ, Glaser R. Reinforcement determinants of team proficiency. *Org Behav Hum Perf.* 1970;5(1):33–67.
- Klein KJ, Conn AB, Smith DB, Sorra JS. Is everyone in agreement? An exploration of within-group agreement in employee perceptions of the work environment. *J App Psych.* 2001;86(1):3–16.
- Kleinman DL, Serfaty D. Team performance assessment in distributed decisionmaking. *Proceedings of the Symposium on Interactive Networked Simulation for Training;* 1989; Orlando, FL. p. 22–27.
- Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psych Bull.* 1996;119(2):254–284.
- Kolbe M, Weiss M, Grote G, Knauth A, Dambach M, Spahn DR, Grande B. TeamGAINS: a tool for structured debriefings for simulation-based team trainings. *BMJ Qual Safe.* 2013;22:541–553.
- Komaki J, Heinzmann AT, Lawson L. Effect of training and feedback: component analysis of a behavioral safety program. *J App Psych.* 1980;65(3):261–270.
- Kostopoulos KC, Bozionelos N. Team exploratory and exploitative learning: psychological safety, task conflict, and team performance. *Group Org Man.* 2011;36(3):385–415.

- Kozlowski SWJ, Klein KJ. A multi-level approach to theory and research in organizations: contextual, temporal, and emergent processes. In: Klein KJ, Kozlowski SWJ, editors. *Multilevel theory, research, and methods in organizations*. San Francisco (CA): Jossey-Bass; 2000. p. 3–90.
- Kozlowski SWJ, Gully SM, Smith EA, Nason ER, Brown KG. Sequenced mastery training and advance organizers: effects on learning, self-efficacy, performance, and generalization. Paper presented at: 10th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL; 1995.
- Leboe JP, Whittlesea BW, Milliken B. Selective and nonselective transfer: positive and negative priming in a multiple-task environment. *J Exp Psych: Learn Mem Cog*. 2005;31(5):1001–1029.
- LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. *Org Res Methods*. 2008;11(4):815–852.
- Lederman LC. Debriefing: toward a systematic assessment of theory and practice. *Sim Gam*. 1992;23:145–160.
- Lee P, Gillespie N, Mann L, Wearing A. Leadership and trust: their effect on knowledge sharing and team performance. *Man Learn*. 2010;41(4):473–491.
- Lent RW, Schmidt J, Schmidt L. Collective efficacy beliefs in student work teams: relation to self-efficacy, cohesion, and performance. *J Voc Behav*. 2006;68(1):73–84.
- LePine JA, Piccolo RF, Jackson CL, Mathieu JE, Saul JR. A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Pers Psych*. 2008;61(2):273–307.
- Leroy H, Dierynck B, Anseel F, Simons T, Halbesleben JR, McCaughey D, Sels L. Behavioral integrity for safety, priority of safety, psychological safety, and patient safety: a team-level study. *J App Psych*. 2012;97(6):1273–1281.
- Lindell MK, Brandt CJ, Whitney DJ. A revised index of interrater agreement for multi-item ratings of a single target. *App Psych Measure*. 1999;23(2):127–135.
- Litman D. Speech and language processing for adaptive training. In: Durlach PJ, Lesgold AM, editors. *Adaptive technologies for training and education*. Cambridge, MA: Cambridge University Press; 2012. p. 247–260.
- London M. *Job feedback: giving, seeking, and using feedback for performance improvement*. 2nd ed. Mahwah (NJ): Lawrence Erlbaum Associates; 2003.

- London M, Polzer JT, Omoregie H. Interpersonal congruence, transactive memory, and feedback processes: an integrative model of group learning. *Hum Res Dev Rev.* 2005;4(2):114–135.
- Luhtanen R, Crocker J. A collective self-esteem scale: self-evaluation of one's social identity. *Pers Soc Psych Bull.* 1992;18(3):302–318.
- Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. *Acad Man Rev.* 2001;26(3):356–376.
- Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. *J App Psych.* 2000;85(2):273–283.
- Mathieu JE, Rapp TL, Maynard MT, Mangos PM. Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. *Hum Perf.* 2009;23(1):22–40.
- Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *Acad Man Rev.* 1995;20(3):709–734.
- McCroskey JC, McCain TA. The measurement of interpersonal attraction. *Speech Mono.* 1974;41:261–266.
- Mohammed S, Dumville BC. Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries. *J Org Beh: Inter J Ind Occ Org Psych Behav.* 2001;22(2):89–106.
- Molleman E. Diversity in demographic characteristics, abilities and personality traits: do faultlines affect team functioning? *Group Dec Neg.* 2005;14(3):173–193.
- Mort TC, Donahue SP. Debriefing the basics. In: Dunn WF, editor. *Simulators in critical care and beyond.* Des Plaines (IL): Society of Critical Care Medicine; 2004. p. 76–83.
- Murase T, Doty D, Wax AMY, DeChurch LA, Contractor NS. Teams are changing: time to “think networks”. *Ind Org Psych.* 2012;5(1):41–44.
- Murthy NN, Challagalla GN, Vincent LH, Shervani TA. The impact of simulation training on call center agent performance: a field-based investigation. *Man Sci.* 2008;54(2):384–399.
- Nkambou R, Mizoguchi R, Bourdeau J. *Advances in intelligent tutoring systems.* Heidelberg (Germany): Springer; 2010.

- Nohria N, Eccles RG. Networks and organizations: structure, form and action. Boston (MA): Harvard Business School Press; 1992.
- Nowack KM, Mashihi S. Evidence-based answers to 15 questions about leveraging 360-degree feedback. *Cons Psych J Prac Res*. 2012;64(3):157–182.
- Nye BD, Graesser AC, Hu X. AutoTutor and family: a review of 17 years of natural language tutoring. *Inter J Art Intel Ed*. 2014;24(4):427–469.
- Oden K. Distributed team training: effective team feedback [dissertation]. [Orlando (FL)]: University of Central Florida; 2008.
- Oliver RL. Whence consumer loyalty? *J Market*. 1999;63:33–44.
- O'Reilly CA 3rd, Caldwell DF, Barnett WP. Work group demography, social integration, and turnover. *Admin Sci Quart*. 1989;34(1):21–37.
- Ortega A, Sánchez-Manzanares M, Gil F, Rico R. Team learning and effectiveness in virtual project teams: the role of beliefs about interpersonal context. *Spanish J Psych*. 2010;13(1):267–276.
- Overstreet M. Ee-chats: the seven components of nursing debriefing. *J Cont Ed Nurs*. 2010;41(12):538–539.
- Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Hum Fact J Hum Fact Ergo Soc*. 1997;39(2):230–253.
- Paris CR, Salas E, Cannon-Bowers JA. Teamwork in multi-person systems: a review and analysis. *Ergonomics*. 2000;43(8):1052–1075.
- Park JH, Shea CH, Wright DL. Reduced-frequency concurrent and terminal feedback: a test of the guidance hypothesis. *J Motor Behav*. 2000;32(3):287–296.
- Pearson AW, Ensley MD, Amason AC. An assessment and refinement of Jehn's intragroup conflict scale. *Inter J Conf Man*. 2002;13(2):110–126.
- Peltier JW, Hay A, Drago W. The reflective learning continuum: reflecting on reflection. *J Market Ed*. 2005;27(3):250–263.
- Podsakoff PM, MacKenzie SB. An examination of the psychometric properties and nomological validity of some revised and reduced substitutes for leadership scales. *J App Psych*. 1994;79(5):702–713.
- Prince M. Does active learning work? A review of the research. *J Eng Ed*. 2004;93(3):223–231.

- Pritchard RD, Jones SD, Roth PL, Stuebing KK, Ekeberg SE. Effects of group feedback, goal setting, and incentives on organizational productivity. *J App Psych.* 1988;73(2):337–358.
- Psotka J, Mutter SA. Intelligent tutoring systems: lessons learned. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
- Quiñones MA. Pretraining context effects: training assignment as feedback. *J App Psych.* 1995;80(2):226–238.
- Raes A, Vanderhoven E, Schellens T. Increasing anonymity in peer assessment by using classroom response technology within face-to-face higher education. *Stud Higher Ed.* 2015;40(1):178–193.
- Rispens S The influence of conflict issue importance on the co-occurrence of task and relationship conflict in teams. *App Psych.* 2012;61(3):349–367.
- Riggs ML, Knight PA. The impact of perceived group success-failure on motivational beliefs and attitudes: a causal model. *J App Psych.* 1994;79(5):755–766.
- Roberson QM, Sturman MC, Simons TL. Does the measure of dispersion matter in multilevel research? A comparison of the relative performance of dispersion indexes. *Org Res Meth.* 2007;10(4):564–588.
- Rosen B, Furst S, Blackburn R. Training for virtual teams: an investigation of current practices and future needs. *Hum Res Man.* 2006;45(2):229–247.
- Rosen MA, Bedwell WL, Wildman JL, Fritzsche BA, Salas E, Burke CS. Managing adaptive performance in teams: guiding principles and behavioral markers for measurement. *Hum Resource Man Rev.* 2011;21(2):107–122.
- Rosen MA, Salas E, Silvestri S, Wu TS, Lazzara EH. A measurement tool for simulation-based training in emergency medicine: the simulation module for assessment of resident targeted event responses (SMARTER) approach. *Sim Health.* 2008;3(3):170–179.
- Rouiller JZ, Goldstein IL. The relationship between organizational transfer climate and positive transfer of training. *Hum Reso Dev Quart.* 1993;4(4):377–390.
- Rus V, Graesser A. The question generation task and evaluation challenge. Memphis (TN): Institute for Intelligent Systems, University of Memphis. 2009 Feb. Workshop report.

- Rutkowski AF, Saunders C, Vogel D, Van Genuchten M. “Is it already 4 am in your time zone?” Focus immersion and temporal dissociation in virtual teams. *Small Group Res.* 2007;38(1):98–129.
- Salas E. *Team training essentials: a research-based guide*. London (UK): Routledge; 2015
- Salas E, Cannon-Bowers JA. The science of training: a decade of progress. *Ann Rev Psych.* 2001;52(1):471–499.
- Salas E, Cooke NJ, Rosen MA. On teams, teamwork, and team performance: discoveries and developments. *Hum Fact J Hum Fact Ergo Soc.* 2008a;50(3):540–547.
- Salas E, Dickinson TL, Converse SA, Tannenbaum SI. Toward an understanding of team performance and training. In: Swezey RW, Salas E, editors. *Teams: their training and performance*. Norwood (NJ): Ablex Publishing; 1992. p. 3–29.
- Salas EE, Fiore SM. *Team cognition: understanding the factors that drive process and performance*. Washington (DC): American Psychological Association; 2004.
- Salas E, Klein C, King H, Salisbury M, Augenstein JS, Birnbach DJ, Upshaw C. Debriefing medical teams: 12 evidence-based best practices and tips. *Joint Comm J Qual Pat Safe.* 2008b;34(9):518–527.
- Salas E, Nichols DR, Driskell JE. Testing three team training strategies in intact teams: a meta-analysis. *Small Group Res.* 2007a;38(4):471–488.
- Salas E, Rosen MA, Burke CS, Nicholson D, Howse WR. Markers for enhancing team cognition in complex environments: the power of team performance diagnosis. *Av Space Env Med.* 2007b;78(1):B77–B85.
- Salas E, Rosen MA, Burke CS, Goodwin GF. *The wisdom of collectives in organizations: an update of the teamwork competencies*. London, UK: Routledge; 2009. *Team effectiveness in complex organizations: cross-disciplinary perspectives and approaches*; p. 39–79.
- Sargent LD, Sue-Chan C. Does diversity affect group efficacy? The intervening role of cohesion and task interdependence. *Small Group Res.* 2001;32(4):426–450.
- Savoldelli GL, Naik VN, Hamstra SJ, Morgan PJ. Barriers to use of simulation-based education. *Canadian J Anes.* 2005;52(9):944–950.

- Schaubroeck J, Lam SS, Peng AC. Cognition-based and affect-based trust as mediators of leader behavior influences on team performance. *J App Psych.* 2011;96(4):863–871.
- Schippers MC, Den Hartog DN, Koopman PL. Reflexivity in teams: a measure and correlates. *App Psych.* 2007;56(2):189–211.
- Scott C, Allen JA, Bonilla DL, Baran BE, Murphy D. Ambiguity and freedom of dissent in post-incident discussion. *J Bus Com.* 2013;50(4):383–402.
- Scott-Young C, Samson D. Team management for fast projects: an empirical study of process industries. *Int J Op Prod Man.* 2009;29(6):612–635.
- Shapiro MJ, Gardner R, Godwin SA, Jay GD, Lindquist DG, Salisbury ML, Salas E. Defining team performance for simulation-based training: methodology, metrics, and opportunities for emergency medicine. *Acad Em Med.* 2008;15(11):1088–1097.
- Shin Y, Choi JN. What makes a group of good citizens? The role of perceived group-level fit and critical psychological states in organizational teams. *J Occ Org Psych.* 2010;83(2):531–552.
- Shivers-Blackwell SL. Using role theory to examine determinants of transformational and transactional leader behavior. *J Leader Org Stud.* 2004;10(3):41–50.
- Shute VJ. Focus on formative feedback. *Rev Ed Res.* 2008;78(1):153–189.
- Siebold GL. The evolution of the measurement of cohesion. *Mil Psych.* 1999;11(1):5–26.
- Singley MK, Fairweather PG, Swerling S. Team tutoring systems: reifying roles in problem solving. In: Hoadley CM and Roschelle J, editors. *CSCL 1999. Proceedings of the Computer Support for Collaborative Learning Conference*; 1999 Dec 12–15; Palo Alto, CA. Mahwah (NJ): Lawrence Erlbaum Associates; 1999; p. 538–548.
- Sleeman D, Brown JS, editors. *Intelligent tutoring systems*. Orlando (FL): Academic Press, Inc.; 1982.
- Smith JB, Barclay DW. The effects of organizational differences and trust on the effectiveness of selling partner relationships. *J Market.* 1997;61(1):3–21.
- Smith-Jentsch KA, Cannon-Bowers JA, Tannenbaum S, Salas E. Guided team self-correction: impacts on team mental models, processes, and effectiveness. *Small Group Res.* 2008;39(3):303–327.

- Smith-Jentsch KA, Kraiger K, Cannon-Bowers JA, Salas E. Do familiar teammates request and accept more backup? Transactive memory in air traffic control. *Hum Fact*. 2009;51(2):181–192.
- Smith-Jentsch KA, Salas E, Baker D. Training team performance-related assertiveness. *Pers Psych*. 1996;49:909–936.
- Smith-Jentsch KA, Zeisig RL, Acton B, McPherson JA. Team dimensional training: a strategy for guided team self-correction. In: Cannon-Bowers JA, Salas E, editors. *Making decisions under stress: implications for individual and team training*. Washington (DC): American Psychological Association; 1998. p. 271–297.
- Solansky ST. Team identification: a determining factor of performance. *J Man Psych*. 2011;26(3):247–258.
- Sottolare R. Making a case for machine perception of trainee affect to aid learning and performance in embedded virtual simulations. Paper presented at: NATO Research Workshop (HFM-RWS-169) on Human Dimensions in Embedded Virtual Simulations; 2009 Oct; Orlando, FL.
- Sottolare R. Toward the development of an intelligent tutoring system for distributed team training through passive sensing. Paper presented at: Intelligent Tutoring Systems (ITS) 2010 Conference; 2010 June; Pittsburgh, PA.
- Sottolare R. Considerations in the development of an ontology for a generalized intelligent framework for tutoring. Paper presented at: International Defense and Homeland Security Simulation Workshop, I3M Conference; 2012 Sep; Vienna, Austria.
- Sottolare RA, Brawner KW, Goldberg BS, Holden HK. The generalized intelligent framework for tutoring (GIFT). Orlando (FL): Army Research Laboratory (US); 2012a. Concept paper toward GIFT software documentation.
- Sottolare R, Brawner K, Sinatra A, Johnston J. An updated concept for a generalized intelligent framework for tutoring (GIFT). Orlando (FL): Army Research Laboratory (US). 2017 May.
- Sottolare R, Brawner K. Exploring standardization opportunities by examining interaction between common adaptive instructional system components. *Proceedings of the First Adaptive Instructional Systems (AIS) Standards Workshop*; 2018 Mar; Orlando, FL.

- Sottilare RA, Burke CS, Salas E, Sinatra AM, Johnston JH, Gilbert SB. Designing adaptive instruction for teams: a meta-analysis. *Int J Art Intel Ed*. 2018;28(2):225–264.
- Sottilare R, Goldberg B. Designing adaptive computer-based tutoring systems to accelerate learning and facilitate retention. *Cog Tech J*. 2012;17(1):19–34.
- Sottilare R, Goldberg B, Brawner K, Holden H. A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). *Proceedings of the Interservice/Industry Training Simulation and Education Conference*; 2012b Dec; Orlando, FL.
- Sottilare RA, Graesser AC, Hu X, Sinatra AM. Introduction to team tutoring and GIFT. In: Sottilare RA, Graesser AC, Hu X, Sinatra AM, editors. *Design recommendations for intelligent tutoring systems—volume 6: team tutoring*. Orlando (FL): Army Research Laboratory (US); 2018. p. 1–15.
- Sottilare R, Holden HK, Brawner KW, Goldberg B. Challenges and emerging concepts in the development of adaptive computer-based tutoring systems for team training. Paper presented at: *Interservice/Industry Training Simulation and Education Conference*; 2011 Nov; Orlando, FL.
- Sottilare R, Ragusa C, Hoffman M, Goldberg B. Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. Paper presented at: *Interservice/Industry Training Simulation and Education Conference*; 2013 Dec; Orlando, FL.
- Stajkovic A, Lee S, Nyberg A. Collective efficacy, group potency, and group performance: meta-analyses of their relationships, and test of mediation model. *J App Psych*. 2009;94:814–828.
- Tannenbaum SI, Cerasoli CP. Do team and individual debriefs enhance performance? A meta-analysis. *Hum Fact*. 2012;55(1):231–245.
- Tannenbaum SI, Goldhaber-Fiebert SN. Medical team debriefs: simple, powerful, underutilized. In: Salas E, Frush K, editors. *Improving patient safety through teamwork and team training*. New York (NY): Oxford University Press; 2012. p. 249–256.
- Tasa K, Taggar S, Seijts GH. The development of collective efficacy in teams: a multilevel and longitudinal perspective. *J App Psych*. 2007;92(1):17–27.
- Tesluk P, Mathieu JE, Zaccaro SJ, Marks M. *Task and aggregation issues in the analysis and assessment of team performance*. Mahwah (NJ): Lawrence

- Erlbaum Associates; 1997. Team performance assessment and measurement: theory, methods, and applications; p. 197–224.
- Tett RP, Guterman HA. Situation trait relevance, trait expression, and cross-situational consistency: testing a principle of trait activation. *J Res Pers.* 2000;34(4):397–423.
- Tracey JB, Tannenbaum SI, Kavanagh MJ. Applying trained skills on the job: the importance of the work environment. *J App Psych.* 1995;80:239–252.
- Tröster C, Mehra A, van Knippenberg D. Structuring for team success: the interactive effects of network structure and cultural diversity on team potency and performance. *Org Behav Hum Dec Proc.* 2014;124(2):245–255.
- Tung HL, Chang YH. Effects of empowering leadership on performance in management team: mediating effects of knowledge sharing and team cohesion. *J Chinese Hum Res Man.* 2011;2(1):43–60.
- van Berlo MPW. Team training vs. team building and cooperative learning: defining the field of research. Soesterberg (Netherlands): Human Factors Research Institute; 1997 Sep. TNO Report No.: TM-97-B109.
- van Jaarsveldt DE, Joubert A. Navigating diversity with nursing students through difficult dialogues: a qualitative study. *Int J Africa Nurs Sci.* 2015;2:34–41.
- VanLehn K. The behavior of tutoring systems. *Int J Art Intel Ed.* 2006;16(3):227–265.
- Wageman R, Hackman JR, Lehman E. Team diagnostic survey: development of an instrument. *J App Behav Sci.* 2005;41(4):373–398.
- Walumbwa FO, Schaubroeck J. Leader personality traits and employee voice behavior: mediating roles of ethical leadership and work group psychological safety. *J App Psych.* 2009;94(5):1275–1286.
- Walsh CM, Ling SC, Wang CS, Carnahan H. Concurrent versus terminal feedback: it may be better to wait. *Acad Med.* 2009;84(10):S54–S57.
- Walther JB, Bunz U. The rules of virtual groups: trust, liking, and performance in computer-mediated communication. *J Comm.* 2005;55(4):828–846.
- Wang D, Waldman DA, Zhang Z. A meta-analysis of shared leadership and team effectiveness. *J App Psych.* 2014;99(2):181–198.
- Watson WE, Michaelsen LK, Sharp W. Member competence, group interaction, and group decision making: a longitudinal study. *J App Psych.* 1991;76(6):803–809.

- Weaver SJ, Rosen MA, Salas E, Baum KD, King HB. Integrating the science of team training: guidelines for continuing education. *J Cont Ed Health Prof.* 2010;30(4):208–220.
- Woehr DJ, Arciniega LM, Poling TL. Exploring the effects of value diversity on team effectiveness. *J Bus Psych.* 2013;28(1):107–121.
- Woehr DJ, Loignon AC, Schmidt PB, Loughry ML, Ohland MW. Justifying aggregation with consensus-based constructs: a review and examination of cutoff values for common aggregation indices. *Org Res Meth.* 2015;18(4):704–737.
- Wolfram HJ, Mohr G. Transformational leadership, team goal fulfillment, and follower work satisfaction: the moderating effects of deep-level similarity in leadership dyads. *J Lead Org Stud.* 2009;15(3):260–274.
- Wong A, Tjosvold D, Liu C. Innovation by teams in Shanghai, China: cooperative goals for group confidence and persistence. *British J Man.* 2009;20(2):238–251.
- Wong SS. Distal and local group learning: performance trade-offs and tensions. *Org Sci.* 2004;15(6):645–656.
- Woolf BP. Building intelligent interactive tutors. Burlington (MA): Morgan Kaufmann Publishers; 2009.
- Wulf G, Shea CH. Understanding the role of augmented feedback. In: Williams AM, Hodges NJ, editors. *Skill acquisition in sport: research, theory and practice.* London (UK): Routledge; 2004. p. 121–144.
- Zaccaro SJ. Nonequivalent associations between forms of cohesiveness and group-related outcomes: evidence for multidimensionality. *J Soc Psych.* 1991;131(3):387–399.

Appendix. Phase 2: Detailed Article Coding

This appendix appears in its original form, without editorial change.

Phase 2: Detailed Article Coding

Summary of Phase 2

During this phase, trained graduate research assistants will be coding the actual meta-analytic data. While the undergraduate coders are pulling articles, **ALL** senior coders on the project will code the first 50 articles for Phase 2 coding independently. After all senior coders have gone through the first 50 articles independently, they will meet and come to a consensus on those 50 articles. The codebook will then be revised of any necessary details.

At this point, the undergraduate research assistants should be finished with Phase 1. Senior coders will then be split up into pairs to complete the remaining coding. Each senior coder will keep record of their own independent coding, as well as their consensus ratings. Both of these excel documents should be stored in a shared location. Once consensus coding is completed, senior coders will transfer this information to a google document.

- The code 998 should be used if the coder is unsure on what category should be used
- The code 999 should be used if the variable is unknown/not reported.

Article Number

- This is the unique article number given to each PDF

Study ID

- This is the unique ID given to each study within the meta-analysis
- Study IDs are an addendum to the article number.
 - If article number 412 included two studies, the first study would be given the unique ID of 412.1 and the second study would be labeled 412.2.
 - If an article only has one study, you still need to add a .1 after the article number. For instance, if article number 560 only had one usable study, the Study ID would be 560.1

N

- N is sample size. Keep in mind that we are only looking at team level variables. Also, it may be possible for the sample size to change based upon the relationship being investigated.

Sample Type

- Which of the following best describes the type of sample?
 - 1 = Employed adults (military)

- This is any sample that includes a military sample (regardless of the country or branch)
- 2 = Employed adults (non-military)
 - Any sample of working adults other than military
- 3 = College students
 - Any sample of college students (full time, part time, graduate, undergraduate)
- 4 = Community sample of adults
 - A sample of adults that does not provide specifics regarding the occupation(s) of those included. For example, this could be a survey emailed out to a list-serve or collected via students in a class handing out surveys as a part of their credit. These participants may or may not be employed.
- 5 = Mixed
 - A sample that involves more than one of the above mentioned groups.

Qualitative Description of Sample

- How many teams (N)? Who were they (e.g., students from a large southwest university, accountants from an MNC)?

Study Design

- Study design is meant to capture two things: 1) whether there was a manipulation and 2) was there time between the measurements of variables.
 - 1 = True Experimental Design
 - Experimental designs occur when the some variable of interest was manipulated, there was random sampling, and random assignment
 - 2 = Quasi-Experimental Design
 - Experimental designs occur when the some variable of interest was manipulated, there was random sampling, but not random assignment
 - 3 = Correlational Design
 - Correlation designs occur when variables of interest were not manipulated and they were measured relatively simultaneously

Time Period

- 1 = Cross-Sectional Design
 - Cross-sectional designs occur when variables were measured relatively simultaneously

- 2 = Longitudinal Design
 - Longitudinal designs occur when there is a period of substantial time between the measurements of variables.

Team Size

- This is a numerical value representing how big the average team size was.
 - You should first look at if they report the mean team size (e.g., correlation matrix, in text)
 - If not, they may report how many teams they had as well as the total number of participants. If this is the case, divide the total number of participants by the number of teams (e.g., “Team size ranged from 4-6 as we gathered data from 52 teams with a 275 participants”; You would report 5.29 (275/52) as the team size)
 - If not, they may describe the targeted team size in the methods section (e.g., “we had 52 3 person teams”; You would report 3 as the team size)
 - If not, you will take the average of the range they provide (e.g., Team size ranged from 3-5; You will report 4 as the team size).

Task Type

- This is a description of what type of task the team engaged in or normally engages in.
 - 1 = Project
 - 2 = Production
 - 3 = Service
 - 4 = Action/Performing
 - 5 = Management
 - 6 = Parallel
 - 7 = Mixed

Interdependence

- Interdependence is the degree to which team members work together to achieve their goals and the degree to which their rewards and outcome are at the team level.
 - 1 = Low – When task performance was largely a function of individual effort; much of the team’s work was performed individually; members generally did not rely on one another to accomplish their work; and feedback, rewards, and goals occurred mainly at the individual level.
 - 2 = Moderate – When members relied on one another for some information and resources but were able to complete a significant portion of the task individually and the goals, outcome, and feedback were mixed.

- 3 = High – When there were mutual or reciprocal dependencies among all members; members' performance was dependent on information or resources provided by other members; and team goals, outcomes, and feedback were emphasized over those of individual members

Work Complexity

- Work complexity will be based on the primary studies' description of the team's task. Unlike other codes, this variable will be on a likert-like scale, ranging from 1-5. It will be based upon three dimension: Knowledge Use, Creative Thinking, and Information Sharing.
 - 1 = Least Complex Work – Low level of knowledge used in the task, does not require much creative thinking, and little information sharing required among group members.
 - Example: It is possible for members complete their work independently, no expertise is needed to complete the task, and there is presumed to be a solid routine to complete the tasks.
 - 2
 - 3
 - 4
 - 5 = Most complex work – High level of knowledge use, high levels of creative thinking, and a large amount of information needed to be exchanged amount team members.
 - Example: The team is highly interdependent in reaching their goal as they need to rely on others for information, evaluation of diverse solutions, and their expertise.

Distribution

- This concerns the degree to which team members were geographically separated from one another.
 - 1 = Distributed
 - A group of individuals who work across time, space, and organizational boundaries with links strengthened by webs of communication technology.
 - 2 = Partially Distributed
 - A team in which some collaborators work together in the same space (members are co-located) and some work remotely using computer-mediated communications.
 - 3 = Collocated
 - A group of individuals who work in close proximity to one another.

Task Familiarity

- Task familiarity is meant to capture the degree to which the team (as a whole) is familiar with the task they are performing (in months).

Team Stability

- This is a “trichotomous” variable indicating whether the teams under investigation were ad-hoc, Pre-existing intact, or newly formed intact
 - 1 = Ad-Hoc
 - These types of teams are teams that have not worked together in the past and will disband once the project they are working on is completed.
 - 2 = Intact-Old
 - These types of teams have worked together in the past and plan on continuing to work together well into the future.
 - 3 = Intact-New
 - These types of teams have **not** worked together in the past (newly formed), but plan on continuing to work together well into the future.

Team Membership

- How many hours has the teams worked together for the purposes of this study?
 - 1 day = 8 hours
 - 1 week = 5 days = 40 hours
 - 1 month = 22 days = 176 hours
 - 1 year = 260 days = 2080 hours

V1/V2

- These will be the relevant variables used in the meta-analytic calculations. It doesn't matter if this variable is considered by the study as an independent variable, a dependent variable or a control variable.
- Select from one of the following variables (Note: the author may call the variable something other than one of the variables listed, so review the measure to see how the variable is being operationalized prior to putting into one of the categories below).
- V2 should always be learning unless there are two learning variables.
 - 100 = Composition
 - 101 = Age
 - 102 = Gender
 - Lower values equal male, High values equal female
 - 103 = National Culture Diversity

- Lower values represent cultural homogeneity, higher values equal greater heterogeneity
- 104 = Professional Diversity
 - Lower values represent more homogeneity, higher values equal greater heterogeneity
- 105 = Age Diversity
 - SD or directly reporting the diversity of age
- 106 = Gender Diversity
 - SD or directly reporting the diversity of Gender
- 107 = Cognitive Ability
 - Any measure of intellectual prowess
- 108 = Cultural Experience
 - Average length of diverse cultural exposure
- 109 = Job Experience (Job Tenure)
 - Average length on job
- 110 = Team Experience (Team Tenure)
 - Average length of time on team
- 111 = Learning Goal Orientation
 - A desire to develop the self by acquiring new skills, mastering new situations, and improving one's competence
- 112 = Prove/Performance Goal Orientation
 - The desire to prove one's competence and to gain favorable judgments about it
- 113 = Avoidance Goal Orientation
 - The desire to avoid the disproving of one's competence and to avoid negative judgments about it.
- 114 = Emotional Stability
 - Reverse code if article says Neuroticism
- 115 = Extraversion
- 116 = Openness to Experience
- 117 = Agreeableness
- 118 = Conscientiousness
- 119 = Pre-Existing Knowledge
- 120 = Overall Demographic Diversity
 - If the authors use an 'average' diversity of multiple indicators (e.g., age, gender, background)
- 121 = Team Size
- 122 = Team Familiarity
 - How well do you know your team members
- 123 = Collective Intelligence

- 124 = Education
- 125 = Job Tenure Diversity
- 126 = Organizational Tenure Diversity
- 127 = Emotional Intelligence
- 128 = Education Diversity
- 200 = Attitudes
 - 201 = Trust
 - Shared belief that all team members will contribute as required by role and protect the team.
 - 202 = Collective Efficacy
 - Belief that the team has the ability to achieve its goals and be successful.
 - 203 = Psychological Safety
 - Shared feeling of safety within team allowing for interpersonal risk-taking.
 - 204 = Cohesion
 - A shared commitment towards the team and its tasks.
 - 205 = Justice
 - Perceived fairness (e.g. procedural justice, distributive justice, interactional justice)
 - 206 = Team Engagement
 - The extent to which team members are vigorous, dedicated, and absorbed in performing the required collaborative tasks (taskwork and teamwork).
 - 207 – Organizational Commitment
- 300 = Behaviors
 - 301 = Learning Behaviors/Processes
 - Behavioral processes that facilitate learning. These processes have included several constructs like exploratory questions, handling conflicts, listening carefully to one another, complementing information from other team members, experimentation, codification and many others. Presently, there is no agreed upon taxonomy of team learning behaviors, so, for now, we will trust the author's judgment.
 - 302 = Communication
 - A reciprocal process of team members' sending and receiving information that forms and re-forms a team's attitudes, behaviors, and cognitions. Communication variables may include (but are not limited to), the sharing of information, use

of information protocols, the quality of communication, the quantity of information.

- 303 = Conflict
 - The perceived incompatibilities in the interests, beliefs, or views held by one or more team members.
- 304 = Leadership Behaviors
 - The enactment of leadership behaviors to establish goals and set direction that leads to the successful accomplishment of these goals.
- 305 = Transition Processes
- 306 = Action Processes
- 307 = Interpersonal Processes
- 308 = Organizational Citizenship Behaviors
- 309 = Inter Team Collaboration
 - Interacting acting with and associating with members of other teams
- 310 – Proactivity
 - Challenging the status quo and taking initiative in improving current circumstances or creating new ones rather than passively adapting to present conditions (Crant, 2000).
- 311 = Coordination
 - The enactment of behavioral and cognitive mechanisms necessary to perform a task and transform team resources into outcomes.
- 312 = External Learning
 - Gathering information from outside the team
- 313 = Reflexivity
 - Discussing whether or not the team is working effectively
- 400 = Cognition
 - 401 = Mental Model Similarity
 - The degree to which team members have a common understanding of the task and team relevant knowledge.
 - 402 = Mental Model Accuracy
 - The degree to which team members have a similar mental model compared to that of an ‘expert’ model
 - 403 = Transactive Memory Systems
 - A shared system for encoding, storing, and retrieving information where each person is responsible only part of the total information.
 - 404 = Situation Awareness

- The perception of environmental elements with respect to time and/or space, the comprehension of their meaning, and the projection of their status after some variable has changed, such as time, or some other variable, such as a predetermined event.
- 500 = Contextual Variables
 - 501 = Influence in Transition
 - Openness to influence the transition phase (planning phase)
 - 502 = Interruptions
 - The degree to which the team is interrupted while performing/how dynamic (in terms of changes) the team's environment is.
 - 503 = Team Stability
 - An index of how consistent team membership was during the performance episode
 - 504 = Proximity
 - An index of the physical closeness of team members
 - 505 = Task Complexity
 - How difficult the team thought the task was
 - 506 = Environmental Stability
 - The general degree of changes (fluctuations) in the environment.
 - 507 = Supportive Context
 - The team's opportunity to reach outside the team for resources (e.g., information) AND/OR a supportive climate for teamwork
 - 508 = Enabling Structure
 - 509 = Interdependence
- 800 = Outcomes
 - 801 = Learning
 - This reflects a definitive change in the teams knowledge.
 - 802 = Performance
 - The evaluation of the outcomes of team processes relative to some set of criteria. A judgment of how well the results of teamwork meet some set of relatively objective or subjective standards.
 - 803 = Satisfaction
 - How much team members enjoyed being a part of their team.
 - 804 = Creativity
 - This refers to reflections of how creative the team processes were, but not the 'objective creativity' value assigned to the team's outcome

- This refers to how creative the final product or performance of the team was.
- 806 = Viability
 - The desire to remain a part of the same team for future performance episodes.
- 807 = Adaptation – Changes in team behaviors that are the result of learning.
- 808 – Turnover

V1_Name/V2_Name

- What did the study call the variable?
- An important note here: Every unique variable from the study should have a unique name. For example, if there are two measures of communication in a single study, they both must have unique names (e.g., communication, information sharing).
- Another important note. Once a variable is called something (E.g., information sharing is the variable name given for the 2nd measure of communication), it **MUST** remain the same in every aspect (e.g., spelling, capitalization) throughout that particular study.

V1_Description/V2_Description

- Provide the name of the measure as listed by the author, a description of the measure and citation of the source for the measure if appropriate, as well as a sample item. This can be copied and pasted straight from the article.

V1_Items/V2_Items

- How many items were in the measure?

V1_R/V2_R

- Provide the specific **Cronbach's alpha** for the reliability of the measure. If not listed or other reliability metric is provided, leave blank.

V1_Source/V2_Source

- What is the source of the variable? In other words, where is the data coming from? Select the most appropriate option.
 - 1 = Self
 - 2 = Team Member(s) (team members rating one another)
 - 3 = Observer
 - 4 = Supervisor
 - 5 = Subordinate(s)
 - 6 = Automated/Objective
 - 7 = Team consensus (team reaches agreement)

- 8 = Instructor (Course/Class)
- 9 = Mix of sources
- 10 = Customer/Client
- 11 = Stakeholder

ES_Type

- Which of the following types of effect sizes is the original effect size as listed in the article?
 - 1 = R
 - 2 = F
 - 3 = T
 - 4 = Z
 - 5 = d

Org_ES

- Provide the original effect size from the article.

Con_ES

- Report the converted (or original, if appropriate) R for this Variable 1-Variable 2 relationship.

Flip

- Should the effect size be flipped or not?
- If there is a variable that represents the opposite of a construct (e.g., *team performance errors* would be coded as team performance, but the effect size would represent the opposite direction of team performance), then the sign should be flipped.
- You may run into situations where both variables should be flipped (e.g., lack of learning's relationship with team performance errors). In this case, you should not flip the effect size.
 - 0 = Don't Flip
 - 1 = Will be Flipped

List of Symbols, Abbreviations, and Acronyms

AIS	adaptive instructional system
ARL	US Army Research Laboratory
CE	collective efficacy
CFI	comparative fit index
GIFT	Generalized Intelligent Framework for Tutoring
ICC	intraclass correlation
IRA	interrater agreement
IRB	Institutional Review Board
IRR	interrater reliability
ITS	intelligent tutoring system
K	number of effect studies analyzed
KSA	knowledge, skill, ability
KSAOs	knowledge, skills, abilities, and other characteristics
LEM	learning-effect model
LTLM	long-term learner model
MASEM	meta-analytic structural equation modeling
N	number of individuals evaluated
NASA	National Aeronautics and Space Administration
OCB	organizational citizenship behavior
Rho	corrected effect size
RMSEA	root mean square of approximation
Ro	uncorrected effect size
SME	subject-matter expert

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 CCDC ARL
(PDF) FCDD RLD CL
TECH LIB

1 CCDC ARL HRED
(PDF) FCDD RLH B
T DAVIS

1 CCDC ARL HRED
(PDF) FCDD RLH FA
V J RICE

1 CCDC ARL HRED
(PDF) FCSS DAS LHD
K GUNN

1 CCDC ARL HRED
(PDF) ARMC FIELD ELEMENT
FCDD RLW PD
C BURNS

1 CCDC ARL HRED
(PDF) AWC FIELD ELEMENT
FCDD DAS LHA
D DURBIN

1 CCDC ARL HRED
(PDF) FCDD DAS LHD
J REINHART

1 CCDC ARL HRED
(PDF) FCDD RLH BD
M BARNES

1 CCDC ARL HRED
(PDF) FCDD RLH
J CHEN

1 CCDC ARL HRED
(PDF) FCDD DAS LHA
A MARES

6 CCDC SC STTC
(PDF) SIMULATION & TRAINING
TECHNOLOGY CENTER
FCDD SCT DD
I MARTINEZ

FCDD SCT TT
A RODRIGUEZ
FCDD SCT CE
J HART
J JOHNSTON
A SINATRA
FCDD SCT AM
C METEVIER
FCDD SCT MS
B PETTIT

ABERDEEN PROVING GROUND

9 CCDC ARL
(PDF) FCDD RLH
J C LANE
P FRANASZCZUK
K MCDOWELL
FCDD DAS LHN
P SAVAGE-KNEPSHIELD
FCDD DAS LHN
C PAULILLO
FCDD DAS LHC
L GARRETT
FCDD RLH
M LAFIANDRA
FCDD RLH FD
A SCHARINE
FCDD RLH BD
D HEADLEY