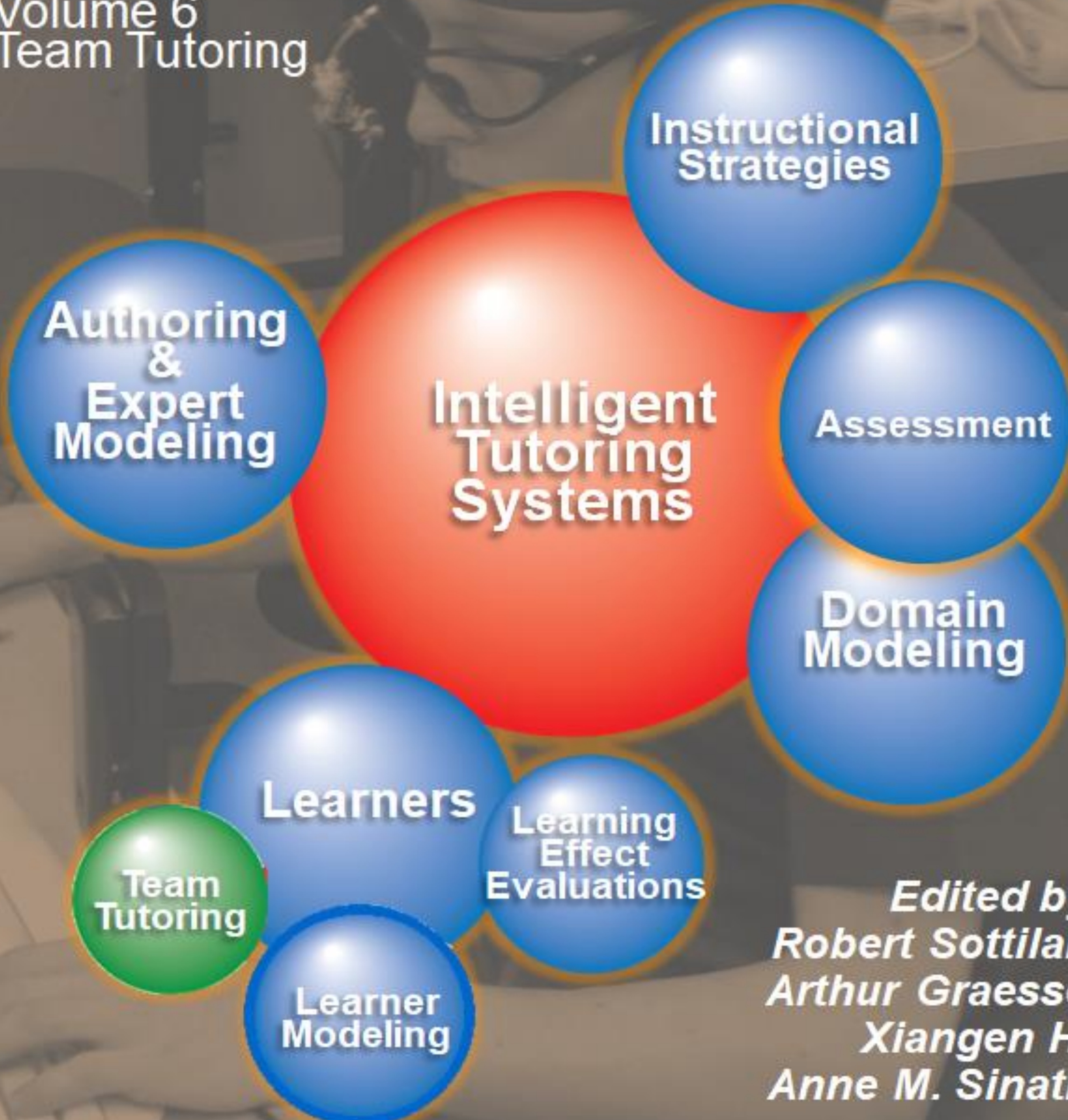
Design Recommendations for Intelligent Tutoring Systems

Volume 6
Team Tutoring



Edited by:
Robert Sottilare
Arthur Graesser
Xiangen Hu
Anne M. Sinatra

A Book in the Adaptive Tutoring Series

Design Recommendations for Intelligent Tutoring Systems

Volume 6
Team Tutoring

Edited by:
Robert A. Sottilare
Arthur C. Graesser
Xiangen Hu
Anne M. Sinatra

A Book in the Adaptive Tutoring Series

Printed in the United States of America
First Printing, August 2018

US Army Research Laboratory
Human Research & Engineering Directorate
Orlando, Florida

International Standard Book Number: 978-0-9977257-4-2

We wish to acknowledge the editing and formatting contributions of Carol Johnson, ARL.

Special thanks to Jody Cockroft, University of Memphis, for her efforts in coordinating the workshop that led to this volume.

Dedicated to current and future scientists and developers of adaptive learning technologies

CONTENTS

INTRODUCTION TO TEAM TUTORING & GIFT

Robert A. Sottilare[1], Arthur C. Graesser[2], Xiangen Hu[2], and Anne M. Sinatra[1] , Eds.

U.S. Army Research Laboratory - Human Research and Engineering Directorate[1]
University of Memphis Institute for Intelligent Systems[2]

This book on team tutoring is the sixth in a planned series of books that examine key topics (e.g., learner modeling, instructional strategies, authoring, domain modeling, assessment, impact on learning, team tutoring, machine learning for self-improving systems, potential standards, and learning effect evaluation methods) in intelligent tutoring system (ITS) design. This book focuses on team tutoring. The discussion chapters in this book examine topics through the lens of the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Brawner, Goldberg & Holden, 2012; Sottilare, Brawner, Sinatra, & Johnston, 2017). GIFT is a modular, service-oriented architecture created to reduce the cost and skill required to author ITSs, distribute ITSs, manage instruction within ITSs, and evaluate the effect of ITS technologies on learning, performance, retention, transfer of skills, and other instructional outcomes.

Along with this volume, the first five books in this series, Learner Modeling (ISBN 978-0-9893923-0-3), Instructional Management (ISBN 978-0-9893923-2-7), Authoring Tools (ISBN 978-0-9893923-6-5), Domain Modeling (978-0-9893923-9-6) and Assessment Methods (ISBN 978-0-9977257-2-8) are freely available at www.GIFTtutoring.org and on Google Play.

This introductory chapter provides a description of basic tutoring functions, provides a glimpse of assessment best practices, and examines the motivation for standards in the design, authoring, instruction, and evaluation of ITS tools and methods. We subsequently introduce GIFT design principles for individuals and teams. We believe this book can be used as a design tool for team tutoring. Before we discuss aspects of tutoring and ITSs, it is important to clarify what we and other stakeholders mean by teams, teamwork, team taskwork, and collaborative learning.

Teams, Teamwork, Team Taskwork, and Collaborative Learning

According to BusinessDictionary.com (2018), a team is "a group of people with a full set of complementary skills required to complete a task, job, or project. Team members (1) operate with a high degree of interdependence, (2) share authority and responsibility for self-management, (3) are accountable for the collective performance, and (4) work toward a common goal and shared rewards(s). A team becomes more than just a collection of people when a strong sense of mutual commitment creates synergy, thus generating performance greater than the sum of the performance of its individual members." While it is a highly complex task to design ITSs to teach individual learners, it is exponentially more difficult to design ITSs to instruct teams.

According Salas, a widely recognized researcher on teams, teamwork is the "coordination, cooperation, and communication among individuals to achieve a shared goal" (Salas, 2015, p.5), where teamwork behaviors are largely domain-independent. Teamwork includes all the social skills needed to function as a team and 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 Berio, 1997). Team taskwork is a subset of team training that is focused on developing proficiency in task domains required for a specific aspect of any individual team member's job (Salas, 2015). Team taskwork is a domain-dependent learning activity and may be confused with the concept of teambuilding or teamwork (Van Berio, 1997).

Collaborative learning (also known as cooperative learning) is "a situation in which two or more people learn or attempt to learn something together" (Dillenbourg, 1999, p. 1) where the process of collaborative learning reinforces active participation (Van Berio, 1997). Collaborative learning generally focuses on a learning goal and is primarily domain-dependent and, in the modern world, frequently includes computer-supported collaborative learning (CSCL) activities. Collaborative problem solving has the additional constraint that a problem needs to be solved that can be objectively assessed with respect to the quality of the solution (Graesser et al., 2017; OECD).

Teamwork and collaboration have been increasingly recognized as being important 21st century skills in the modern world that is increasing in complexity to the point where a single individual cannot complete many tasks alone. Instead, a group of people with different roles, skills and perspectives need to pool their resources in an interdependent manner (OECD, 2017). Interestingly, these skills have not been systematically integrated with the curriculum between kindergarten and college (Fiore et al., 2017; OECD, 2017) so there is a need to improve teamwork and collaboration in both schools and the workforce (Fiore, Graesser, & Greiff, 2018). This is where ITS can come to the rescue. An ITS can improve teamwork and collaboration in addition to the subject matters under focus. Next, we examine the major components of ITSs.

Components and Functions of Intelligent Tutoring Systems

It is generally accepted that an ITS has four major components (Elson-Cook, 1993; Nkambou, Mizoguchi & Bourdeau, 2010; Graesser, Hu, & Sottilare, 2018; Psotka & Mutter, 2008; Sleeman & Brown, 1982; VanLehn, 2006; Woolf, 2009): the domain model, the student model, the tutoring model, and the user-interface model. GIFT similarly adopts this four-part distinction, but with slightly different corresponding labels (domain module, learner module, pedagogical module, and tutor-user interface) and the addition of the sensor module, which can be viewed as an expansion of the user interface. In this volume, we also introduce the team model to support assessment of team learning objectives and groups performing collaborative tasks.

(1) 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.

(2) 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. For example, "knowledge tracing" tracks the learner's progress from problem to problem and builds a profile of strengths and weaknesses relative to the domain model (Anderson, Corbett, Koedinger & Pelletier, 1995). An ITS may also consider psychological states outside of the domain model that need to be considered as parameters to guide tutoring.

(3) 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. In mixed-initiative systems, the learners may also take actions, ask questions, or request help (Aleven, McClaren, Roll & Koedinger, 2006; Rus & Graesser, 2009), but the ITS always needs to be ready to decide "what to do next" at any point and this is determined by a tutoring model that captures the researchers' pedagogical theories.

(4) 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). In addition to the conventional human-computer interface features, some recent systems have incorporated natural language interaction (Graesser, 2016; Johnson & Lester, 2016; Nye, Graesser, & Hu, 2014), speech recognition (D'Mello, Graesser & King, 2010; Litman, 2013), and the sensing of learner emotions (Baker, D'Mello, Rodrigo & Graesser, 2010; D'Mello & Graesser, 2010; Goldberg, Sottilare, Brawner, Holden, 2011).

(5) The **team model**, an optional component of an ITS, is the group equivalent of an individual learner model. The team model must be able to track progress toward team task learning objectives for training or collaborative learning goals for learner development or problem-solving, but also has

the added complexity of monitoring teamwork states that moderate or influence team and individual learning and performance (Sottilare et al, 2017). Team models are more complex than individual learner models since it is not as simple as adding up individual learner performances to find the team's level of performance.

The designers of a tutor model 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 and that are delivered to the learner as instructional tactics (optimal actions by the tutor). Next, tactics are chosen based on the previously selected strategies and instructional context (the conditions of the training at the time of the instructional decision). This is part of the learning effect model (Sottilare, 2012; Fletcher & Sottilare, 2013; Sottilare, 2013; Sottilare, Ragusa, Hoffman & Goldberg, 2013), which has been updated and described in the next section.

Principles of Learning and Instructional Techniques, Strategies, and Tactics

Instructional techniques, strategies, and tactics play a central role in the design of GIFT. Instructional techniques represent instructional best practices and principles from the literature, many of which have yet to be implemented within GIFT at the writing of this volume. Examples of instructional techniques include, but are not limited to, error-sensitive feedback, mastery learning, adaptive spacing and repetition, and fading worked examples. Others are represented in the next section of this introduction. It is anticipated that techniques within GIFT will be implemented as software-based agents where the agent will monitor learner progress and instructional context to determine if best practices (agent policies) have been adhered to or violated. Over time, the agent will learn to enforce agent policies in a manner that optimizes learning and performance.

Some of the best instructional techniques have yet to be implemented in GIFT, but many instructional strategies and tactics have been implemented. Instructional strategies (plans for action by the tutor) are selected based on changes to the learner's state (e.g., cognitive, affective, and physical). If a sufficient change in any learner's state occurs, this triggers GIFT to select a generic strategy such as providing feedback. The instructional context along with the instructional strategy then triggers the specific selection of an instructional tactic (an action to be taken by the tutor). If the strategy is to "provide feedback," then the tactic might be to "provide feedback on the error committed during the presentation of instructional concept B in the chat window during the next turn." Tactics detail what is to be done, why, when, and how in the specific domain. An adaptive, intelligent learning environment needs to select the right instructional strategies at the right time, based on its model of the learner in specific conditions and the learning process in general. Such selections should be taken to maximize deep learning and motivation while minimizing training time and costs.

Motivations for Intelligent Tutoring System Standards

Some subject matters are difficult to learn without the scaffolding of a tutor, but unfortunately human tutors are not available 24-7 for point-of-need training. ITSs have been designed to fill this gap. Moreover, ITSs have been shown to be as effective as expert human tutors (VanLehn, 2011) in one-to-one tutoring for well-defined domains (e.g., mathematics or physics) and significantly better than traditional classroom training environments and students reading texts for an equivalent amount of time (Graesser, Rus, & Hu, 2017). ITSs have demonstrated significant promise, but 50 years of research have been unsuccessful in making ITSs ubiquitous in military training or the tool of choice in our educational system. This begs the question: "Why?"

Part of the answer lies in the fact that the availability and use of ITSs have been constrained by their high development costs, their limited reuse, a lack of standards, and their inadequate adaptability to the needs of learners. Educational and training technologies like ITSs are primarily investigated and developed in a few key environments: industry, academia, and government including military domains. Each of these environments has its own challenges and design constraints. The application of ITSs to military domains is further hampered by the complex and often ill-defined environments in which the US military operates today. ITSs are often built as domain-specific, unique, one-of-a-kind, largely domain-dependent solutions focused on a single pedagogical strategy (e.g., model tracing or constraint-based approaches) when complex learning domains may require novel or hybrid approaches. Therefore, a modular ITS framework and standards are needed to enhance reuse, support authoring, optimize instructional strategies, and lower the cost and skillset needed for users to adopt ITS solutions for training and education. It was out of this need that the idea for GIFT arose.

GIFT has three primary functions: authoring, instructional management, and evaluation. First, it is a framework for authoring new ITS components, methods, strategies, and whole tutoring systems. Second, GIFT is an instructional manager that integrates selected instructional theory, principles, and strategies for use in ITSs. Finally, GIFT is an experimental testbed used to evaluate the effectiveness and impact of ITS components, tools, and methods. GIFT is based on a combination of learner-centric, instructor-centric, and interaction-centric approaches with the goal of improving linkages in the updated adaptive tutoring learning effect model (see Figure 1; Sottilare, Burke, Salas, Sinatra, Johnston, & Gilbert, 2017).

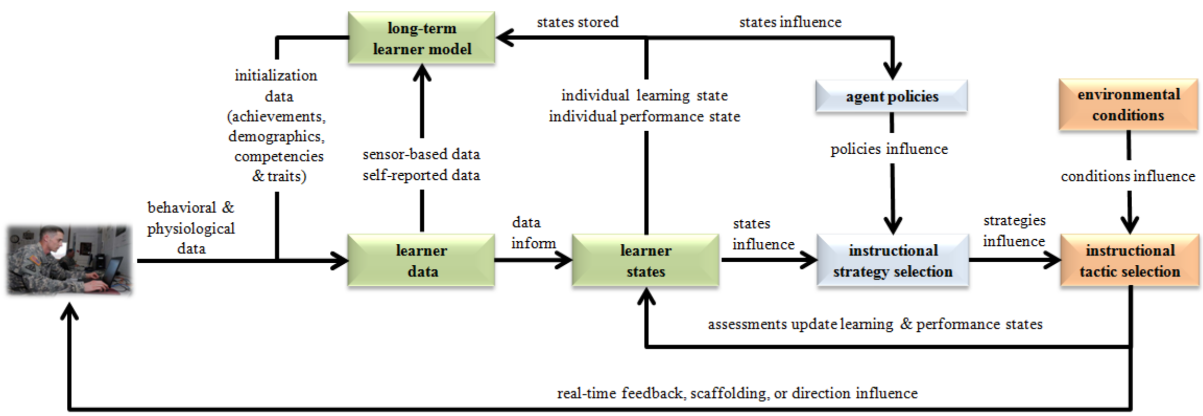


Figure 1. Updated learning effect model for individual learners (Sottilare et al., 2017)

A deeper understanding of the learner's behaviors, traits, and preferences that are collected through performance, physiological and behavioral sensors, and surveys will allow for more accurate evaluation of the learner's cognitive and affective states (e.g., engagement level, confusion, frustration). This will result in a better and more persistent model of the learner. To enhance the adaptability of the ITS, methods are needed to accurately classify learner states (e.g., cognitive, affective, psychomotor, social) and select optimal instructional strategies given the learner's current and predicted states. A similar chain of interactions between a team and an ITS can be constructed for teams working together collaboratively for a common purpose (see Figure 2, Sottilare, Burke, Salas, Sinatra, Johnston, & Gilbert, 2017). This representation is referred to as a learning effect model (LEM).

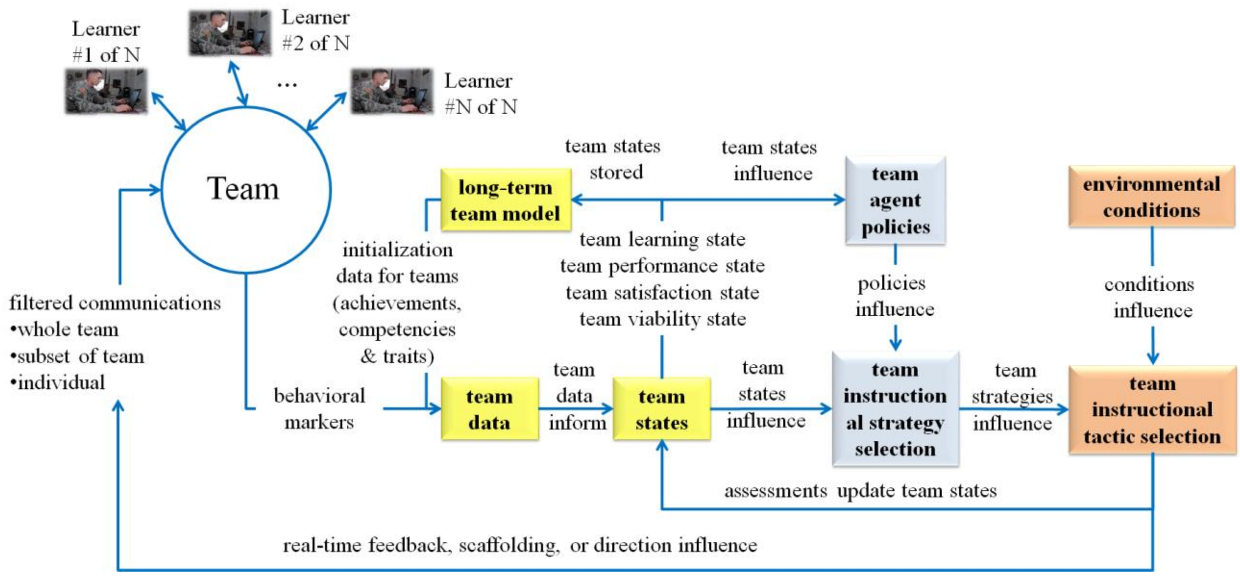


Figure 2. Updated learning effect model for teams (LEM; Sottilare et al., 2017)

A more comprehensive individual learner model or team model will allow an ITS to adapt more appropriately to address the learner's/team's needs by changing the instructional strategy (e.g., content, flow, or feedback). An instructional strategy that is better aligned to each learner's/team's needs is more likely to influence their learning gains in a positive way. It is with the goal of optimized learning gains in mind that the design principles for GIFT were formulated. That being said, it would also be worthwhile to consider other team goals, such as the quality of solutions in collaborative problem solving, the quality of decisions in group decision making, and the quality of products in collaborative design.

This version of the learning effect model has been updated to gain understanding of the effect of optimal instructional tactics and instructional contexts (both part of the domain model) on specific desired outcomes, including knowledge and skill acquisition, performance, retention, and transfer of skills from training or tutoring environments to operational contexts (e.g., from practice to application). The feedback loops in Figure 1 and 2 have been added to identify tactics as either a change in instructional context or interaction with the learner. This allows the ITS to adapt to the need of the individual learner or the team. Consequently, the ITS improves over time by reinforcing learning mechanisms.

GIFT Design Principles

The GIFT methodology for developing a modular, computer-based tutoring framework for training and education considered major design goals, anticipated uses, and applications. The design process also considered enhancing one-to-one (individual) and one-to-many (collective or team) tutoring experiences beyond the state of practice for ITSs today. A significant focus of the GIFT design was on maximizing the number of domain-independent elements resulting in domain-dependent element occurring only in the domain module only. This is a design tradeoff to foster reuse through interoperability and allows ITS decisions and actions to be made across any/all domains of instruction.

One design principle adopted in GIFT is that each module should be capable of gathering information from other modules according to the design specification. Designing to this principle resulted in standard message sets and message transmission rules (i.e., request-driven, event-driven, or periodic transmissions). For instance, the pedagogical module is capable of receiving information from the learner module to develop courses of action for future instructional content to be displayed, manage flow and challenge level, and select appropriate feedback. Changes to the learner's state trigger messages to the pedagogical module, which then recommends general courses of action (e.g., ask a question or prompt the learner for more information) to the domain module, which provides a domain-specific intervention (e.g., what is the next step?).

Another design principle adopted within GIFT is the separation of content from the executable code (Patil & Abraham, 2010). Data and data structures are placed within models and libraries, while software processes are programmed into interoperable modules. Efficiency and effectiveness goals (e.g., accelerated learning and enhanced retention) were considered to address the time available for military training and the renewed emphasis on 24-7 self-regulated learning. An outgrowth of this emphasis on efficiency and effectiveness led Dr. Sottilare to seek external collaboration and guidance. In 2012, ARL with the University of Memphis developed expert workshops of senior tutoring system scientists from academia and government to influence the GIFT design goals moving forward. Expert workshops have been held each year since 2012 resulting in volumes in the ***Design Recommendations for Intelligent Tutoring Systems*** series the following year. The *learner modeling* expert workshop was completed in September 2012 and Volume 1 followed in July 2013. An expert workshop on *instructional management* was completed in July 2013 and Volume 2 followed in June 2014. The *authoring tools* expert workshop was completed in June of 2014 and Volume 3 was published in June 2015. The *domain modeling* expert workshop was held in June 2015 and Volume 4 was published in July 2016. The *assessment* expert workshop was held in May 2016 and Volume 5 was published in June 2017. The *team tutoring* expert workshop was held in May 2017 and is published in this Volume 6. We recently conducted a workshop on *machine learning techniques for adaptive instruction* (self improving systems). Future expert workshops are planned for *adaptive instructional system (AIS) standards*, and *learning effect evaluation methods*.

Design Goals and Anticipated Uses

GIFT may be used for a number of purposes, with the primary ones enumerated below:

1. An architectural framework with modular, interchangeable elements and defined relationships to support stand-alone tutoring or guided training if integrated with a training system

2. A set of specifications to guide ITS development

3. A set of exemplars or use cases for GIFT to support authoring, reuse, and ease-of-use

4. A technical platform or testbed for guiding the evaluation, development, and refinement of concrete systems

The use cases have been distilled down into three primary functional areas: authoring, instructional management, and evaluation function. Discussed below are the purposes, associated design goals, and anticipated uses for each of the GIFT functions.

GIFT Authoring Function

The purpose of the GIFT authoring function is to provide technology (tools and methods) to make it affordable and easier to build ITSs and ITS components. Toward this end, a set of authoring interfaces with backend XML configuration tools continues to be developed to allow for data-driven changes to the design and implementation of GIFT-generated ITSs. The design goals for the GIFT authoring function have been adapted from Murray (1999, 2003) and Sottilare and Gilbert (2011). The GIFT authoring design goals are as follow:

- Decrease the effort (time, cost, and/or other resources) for authoring and analyzing ITSs by automating authoring processes, developing authoring tools and methods, and developing standards to promote reuse.
- Decrease the skill threshold by tailoring tools for specific disciplines (e.g., instructional designers, training developers, and trainers) to author, analyze, and employ ITS technologies.
- Provide tools to aid designers, authors, trainers, and researchers in organizing their knowledge.
- Support (structure, recommend, or enforce) good design principles in pedagogy through user interfaces and other interactions.
- Enable rapid prototyping of ITSs to allow for rapid design/evaluation cycles of prototype capabilities.
- Employ standards to support rapid integration of external training/tutoring environments (e.g., simulators, serious games, slide presentations, transmedia narratives, and other interactive multimedia).
- Develop/exploit common tools and user interfaces to adapt ITS design through data-driven means.
- Promote reuse through domain-independent modules and data structures.
- Leverage open-source solutions to reduce ITS development and sustainment costs.
- Develop interfaces and gateways to widely-used commercial and academic tools (e.g., games, sensors, toolkits, virtual humans).

As a user-centric architecture, anticipated uses for GIFT authoring tools are driven largely by the anticipated users, which include learners, domain experts, instructional system designers, training and tutoring system developers, trainers and teachers, and researchers. In addition to user models and graphical user interfaces (GUIs), GIFT authoring tools include domain-specific knowledge configuration tools, instructional strategy development tools, and a compiler to generate executable ITSs from GIFT components in a variety of formats (e.g., PC, Android, and iPad).

Within GIFT, domain-specific knowledge configuration tools permit authoring of new knowledge elements or reusing existing (stored) knowledge elements. Domain knowledge elements include learning objectives, media, task descriptions, task conditions, standards and measures of success, common misconceptions, feedback library, and a question library, which are informed by instructional system design principles that, in turn, inform concept maps for lessons and whole courses. The task descriptions, task conditions, standards and measures of success, and common misconceptions may be informed by an expert or ideal learner model derived through a task analysis of the behaviors of a highly skilled user. ARL is investigating techniques to automate this expert model development process to reduce the time and cost of developing ITSs. In addition to feedback and questions, supplementary tools are anticipated to author explanations, summaries, examples, analogies, hints, and prompts in support of GIFT's instructional management function.

GIFT Instructional Management Function

The purpose of the GIFT instructional management function is to integrate pedagogical best practices in GIFT-generated ITSs. The modularity of GIFT will also allow GIFT users to extract pedagogical models for use in tutoring/training systems that are not GIFT-generated. GIFT users may also integrate pedagogical models, instructional strategies, or instructional tactics from other tutoring systems into GIFT. The design goals for the GIFT instructional management function are the following:

- Support ITS instruction for individuals and small teams in local and geographically distributed training environments (e.g., mobile training), and in both well-defined and ill-defined learning domains.

- Provide for comprehensive learner models that incorporate learner states, traits, demographics, and historical data (e.g., performance) to inform ITS decisions to adapt training/tutoring.

- Support low-cost, unobtrusive (passive) methods to sense learner behaviors and physiological measures and use these data along with instructional context to inform models to classify (in near real time) the learner's states (e.g., cognitive and affective).

- Support both macro-adaptive strategies (adaptation based on pre-training learner traits) and micro-adaptive instructional strategies and tactics (adaptation based learner states and state changes during training).

- Support the consideration of individual differences where they have empirically been documented to be significant influencers of learning outcomes (e.g., knowledge or skill acquisition, retention, and performance).

- Support adaptation (e.g., pace, flow, and challenge level) of the instruction based the domain and task classification (e.g., cognitive learning, affective learning, psychomotor learning, and social learning).

- Model appropriate instructional strategies and tactics of expert human tutors to develop a comprehensive pedagogical model.

To support the development of optimized instructional strategies and tactics, GIFT is heavily grounded in learning theory, tutoring theory, and motivational theory. Learning theory applied in GIFT includes conditions of learning and theory of instruction (Gagne, 1985), component display theory (Merrill, Reiser, Ranney & Trafton, 1992), cognitive learning (Anderson & Krathwohl, 2001), affective learning (Krathwohl, Bloom & Masia, 1964; Goleman, 1995), psychomotor learning (Simpson, 1972), and social learning (Sottilare, Holden, Brawner, & Goldberg, 2011; Soller, 2001). Aligning with our goal to model expert human tutors, GIFT considers the intelligent, nurturant, Socratic, progressive, indirect, reflective, and encouraging (INSPIRE) model of tutoring success (Lepper, Drake, & O'Donnell-Johnson, 1997) and the tutoring process defined by Person, Kreuz, Zwaan, and Graesser (1995) and Graesser (2016) in the development of GIFT instructional strategies and tactics. Recently, GIFT's instructional management and authoring capacity was expanded to include the Interactive, Constructive, Active, and Passive (ICAP) framework (Chi & Wylie, 2014) in order to link cognitive engagement to active learning outcomes.

Human tutoring strategies have been documented by observing tutors with varying levels of expertise. For example, Lepper's INSPIRE model is an acronym that highlights the seven critical characteristics of successful tutors. Graesser and Person's (1994) 5-step tutoring frame is a common pattern of the tutor-learner interchange in which the tutor asks a question, the learner answers the question, the tutor gives short feedback on the answer, then the tutor and learner collaboratively improve the quality of (or embellish) the answer, and finally, the tutor evaluates whether the learner understands the answer. Cade, Copeland, Person, and D'Mello (2008) identified a number of tutoring modes used by expert tutors, which hopefully could be integrated with ITS.

A key next step in the evolution of GIFT will be to transform it from an individual tutoring architecture to a team tutoring architecture (Gilbert et al., 2018; Sottilare et al., 2018). This will expand the variety of domains which can be instructed by GIFT-based tutors, but will also increase the complexity of the authoring task. To overcome this increase in complexity, we anticipate that GIFT will evolve a number of automated authoring tools to reduce the authoring load and skill required to develop ITSs.

As a learner-centric architecture, anticipated uses for GIFT instructional management capabilities include both automated instruction and blended instruction, where human tutors, teachers, and trainers use GIFT to support their curriculum objectives. If its design goals are realized, it is anticipated that GIFT will be widely used beyond military training contexts as GIFT users expand the number and type of learning domains and resulting ITS generated using GIFT.

GIFT Evaluation Function

The GIFT Evaluation Function emphasizes the evaluation of effect on learning, performance, retention and transfer. The purpose of the GIFT evaluation function is to allow ITS researchers to experimentally assess and evaluate ITS technologies (ITS components, tools, and methods). The design goals for the GIFT evaluation function are depicted in Figure 3 and elaborated below:

- Support the conduct of formative assessments to improve learning.
- Support summative evaluations to gauge the effect of technologies on learning.
- Support assessment of ITS processes to understand how learning is progressing throughout the tutoring process.
- Support evaluation of resulting learning versus stated learning objectives.
- Provide diagnostics to identify areas for improvement within ITS processes.
- Support the ability to comparatively evaluate ITS technologies against traditional tutoring or classroom teaching methods.
- Develop a testbed methodology to support assessments and evaluations (Figure).

Figure 3 illustrates an analysis testbed methodology being implemented in GIFT. This methodology was derived from Hanks, Pollack, and Cohen (1993). It supports manipulation of the learner model, instructional strategies, and domain-specific knowledge within GIFT, and may be used to evaluate variables in the adaptive tutoring learning effect model (Sottilare, 2012; Sottilare, Ragusa, Hoffman, & Goldberg, 2013). It might also support manipulations of the team modeling. In developing their testbed methodology, Hanks et al. reviewed four testbed implementations (Tileworld, the Michigan Intelligent Coordination Experiment [MICE], the Phoenix testbed, and Truckworld) for evaluating the performance of artificially intelligent agents. Although agents have changed substantially in complexity during the past 20–25 years, the methods to evaluate their performance have remained markedly similar.

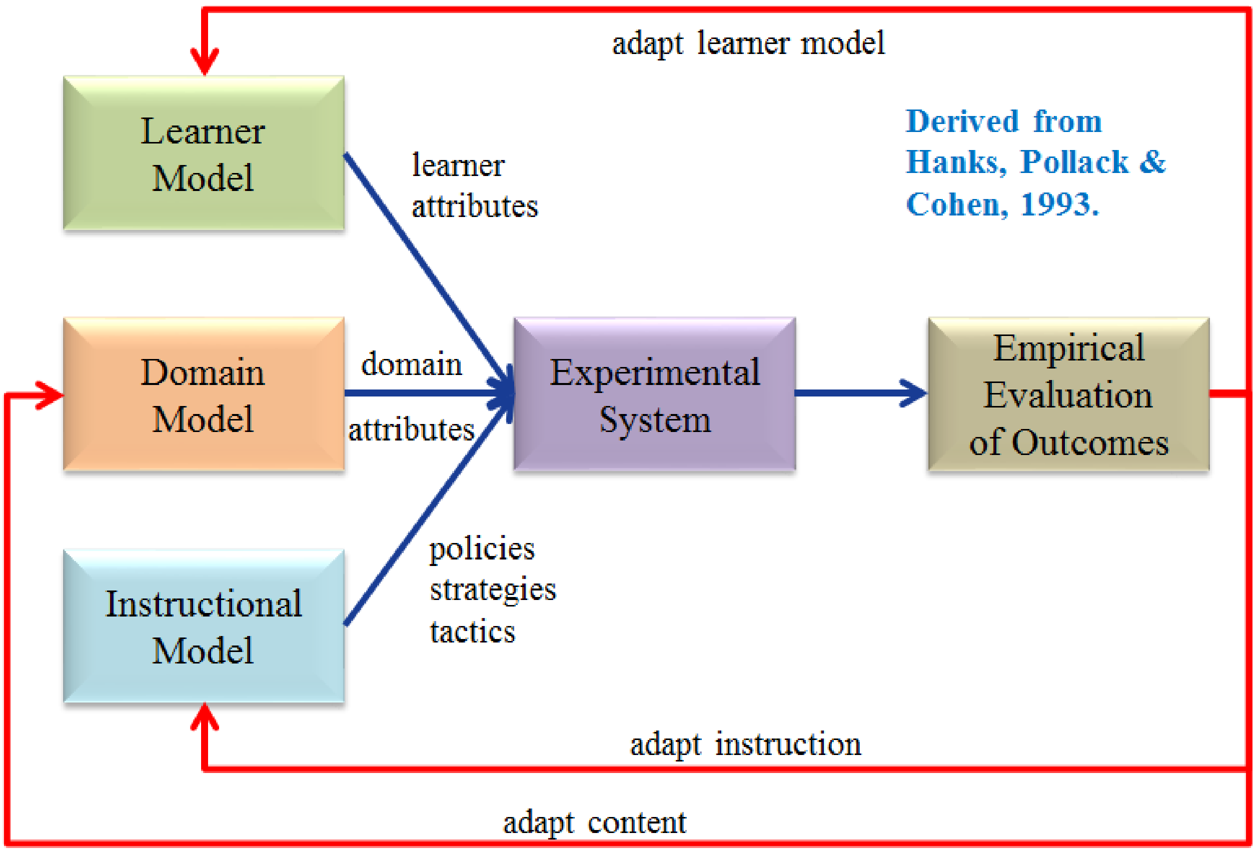


Figure 3. GIFT evaluation testbed methodology

The ARL adaptive training team designed the GIFT analysis testbed based upon Cohen's assertion (Hanks et al., 1993) that testbeds have three critical roles related to the three phases of research. During the exploratory phase, agent behaviors need to be observed and classified in broad categories. This can be performed in an experimental environment. During the confirmatory phase, the testbed is needed to allow more strict characterizations of agent behavior to test specific hypotheses and compare methodologies. Finally, in order to generalize results, measurement and replication of conditions must be possible. Similarly, the GIFT evaluation methodology in Figure enables the comparison/contrast of ITS elements and assessment of their effect on learning outcomes (e.g., knowledge acquisition, skill acquisition, and retention).

With a firm understanding of the design of ITSs in general and GIFT in particular, we can now shift our attention to the use of this book as a design tools for team ITSs.

How to Use This Book

This book is organized into four sections:

I. Team Modeling

II. Team Assessment Methods

III. Socio-Cultural Applications

IV. System Design of Team Tutors

Section I, *Team Modeling*, explores various aspects of ITSs for teams. Using GIFT as a basis, this section discusses the modeling aspects of teams, including the mission of the team, the interaction of learners within the team, and the roles and responsibilities of team members.

Section II, *Team Assessment Methods*, highlights the importance of defining measures and methods to accurately assess progress toward learning goals.

Section III, *Socio-Cultural Applications*, discusses aspects of team tutoring that are inherently social in nature. Whether teams are communicating with each other through typed messages, verbally, or using hand signals, communication is vital to team task performance.

Section IV, *System Design of Team Tutors*, focuses on the system design aspects for the tutoring of team taskwork using ITSs. Team taskwork is composed of domain-dependent measures of success whereas teamwork evaluates attitudes, behaviors, and cognition of team members independent of the task domain. When we discuss system design, we are specifically addressing how ITSs are designed to assess and interact with a group of learners working to complete a task. The type of task, the configuration of the team, and their roles and responsibilities vary to present interesting design challenges to ITS authors.

Chapter authors in each section were carefully selected for participation in this project based on their expertise in the field as ITS scientists, developers, and practitioners. *Design Recommendations for Intelligent Tutoring Systems: Volume 6 – Team Taskwork* is intended to be a design resource as well as a community research resource. We believe that Volume 6 can serve as an educational guide for developing ITS scientists and as a roadmap for ITS research opportunities. The authors of the chapters contained herein are experts in their area and the references provided (their own and those of others) compose a rich web of working professionals in the ITS field.

References

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101-128.

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167-207.

Anderson, L. W. & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives: Complete edition*. New York: Longman.

Baker, R.S., D'Mello, S.K., Rodrigo, M.T. & Graesser, A.C. (2010). 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. *International Journal of Human-Computer Studies, 68*, 223-241.

BusinessDictionary.com (2018). Team. Retrieved from: http://www.businessdictionary.com/definition/team.html

Cade, W., Copeland, J. Person, N., & D'Mello, S. K. (2008). Dialogue modes in expert tutoring. In B. Woolf, E. Aimeur, R. Nkambou & S. Lajoie (Eds.), Proceedings of the *Ninth International Conference on Intelligent Tutoring Systems* (pp. 470-479). Berlin, Heidelberg: Springer-Verlag.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

D'Mello, S. & Graesser, A.C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20*, 147-187.

D'Mello, S. K., Graesser, A. C. & King, B. (2010). Toward spoken human-computer tutorial dialogues. *Human Computer Interaction, 25*, 289-323.
Dillenbourg, P. (1999). What do you mean by collaborative learning? Collaborative-learning: Cognitive and computational approaches, 1, 1-15.
Elson-Cook, M. (1993). Student modeling in intelligent tutoring systems. *Artificial Intelligence Review, 7*, 227-240.
Fiore, S.M., Graesser, A.C., & Greiff, S. (2018). Collaborative problem solving education for the 21st century workforce. *Nature Human Behavior*, *June,* 1-3.
Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., … von Davier, A. A. (2017). *Collaborative problem solving: Considerations for the National Assessment of Educational Progress*. Alexandria, VA: National Center for Education Statistics.
Fletcher, J.D. & Sottilare, R. (2013). Shared Mental Models and Intelligent Tutoring for Teams. In R. Sottilare, A. Graesser, X. Hu, and H. Holden (Eds.) Design *Recommendations for Intelligent Tutoring Systems: Volume I - Learner Modeling*. Army Research Laboratory, Orlando, Florida. ISBN 978-0-9893923-0-3.
Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart & Winston.
Gilbert, S., Slavina, A., Sinatra, A. M., Bonner, D., Johnston, J., Holub, J., MacAllister, A., Dorneich, M., Winer, E. (2018). Creating a team tutor using GIFT. *International Journal for Artificial Intelligence in Education, 28,* 286-313.
Goldberg, B.S., Sottilare, R.A., Brawner, K.W. & Holden, H.K. (2011). Predicting Learner Engagement during Well-Defined and Ill-Defined Computer-Based Intercultural Interactions. In S. D'Mello, A. Graesser, B. Schuller & J.-C. Martin (Eds.), *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011) (Part 1: LNCS 6974)* (pp. 538-547). Berlin Heidelberg: Springer.
Goleman, D. (1995). *Emotional intelligence*. Bantam Books, New York (1995).
Graesser, A.C. (2016). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education, 26,* 124-132.
Graesser, A.C., Foltz, P.W., Rosen, Y., Shaffer, D.W., Forsyth, C., and Germany, M. (2018). Challenges of assessing collaborative problem solving. In E. Care, P. Griffin, and M. Wilson (eds.), *Assessment and Teaching of 21st Century Skills* (pp. 75-91). Heidelberg, Germany: Springer Publishers.
Graesser, A.C., Hu, X., & Sottilare, R. (2018). Intelligent tutoring systems. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, and P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 246-255). New York: Routledge.
Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137.
Graesser, A.C., Rus, V., Hu, X. (2017). Instruction based on tutoring. In R.E. Mayer and P.A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 460-482). New York: Routledge Press.
Hanks, S., Pollack, M.E. & Cohen, P.R. (1993). Benchmarks, test beds, controlled experimentation, and the design of agent architectures. *AI Magazine*, *14* (4), 17-42.
Johnson, W.L. & Lester, J.C. (2016). Twenty years of face-to-face interaction with pedagogical agents. *International Journal of Artificial Intelligence in Education, 26,* 25-36.
Krathwohl, D.R., Bloom, B.S. & Masia, B.B. (1964). *Taxonomy of Educational Objectives: Handbook II: Affective Domain*. New York: David McKay Co.
Lepper, M. R., Drake, M. & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds), *Scaffolding learner learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.
Litman, D. (2013). Speech and language processing for adaptive training. In P. Durlach & A. Lesgold (Eds.), *Adaptive technologies for training and education*. Cambridge, MA: Cambridge University Press.
Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*(1), 98–129.
Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In Murray, T.; Blessing, S.; Ainsworth, S. (Eds.), *Authoring tools for advanced technology learning environments* (pp. 491-545). Berlin: Springer.
Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. The Journal of the Learning Sciences, 2(3), 277-305
Nkambou, R., Mizoguchi, R. & Bourdeau, J. (2010). *Advances in intelligent tutoring systems*. Heidelberg: Springer.
Nye, B.D., Graesser, A.C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24,* 427–469.

OECD (2013). *PISA 2015 collaborative problem solving framework*. Paris: France: OECD.
OECD (2017). *PISA 2015 Results (Volume V): Collaborative Problem Solving*. Paris: OECD Publishing.
Patil, A. S. & Abraham, A. (2010). Intelligent and Interactive Web-Based Tutoring System in Engineering Education: Reviews, Perspectives and Development. In F. Xhafa, S. Caballe, A. Abraham, T. Daradoumis & A. Juan Perez (Eds.), *Computational Intelligence for Technology Enhanced Learning. Studies in Computational Intelligence* (Vol 273, pp. 79-97). Berlin: Springer-Verlag.
Person, N. K., Kreuz, R. J., Zwaan, R. A. & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13*(2), 161–188.
Psotka, J. & Mutter, S.A. (1988). *Intelligent Tutoring Systems: Lessons Learned*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Rus, V. & Graesser, A.C. (Eds.) (2009). The Question Generation Shared Task and Evaluation Challenge. Retrieved from http://www.questiongeneration.org/.
Salas, E. (2015). Team training essentials: A research-based guide. Routledge.
Simpson, E. (1972). The classification of educational objectives in the psychomotor domain: *The psychomotor domain*. Vol. 3. Washington, DC: Gryphon House.
Sleeman D. & J. S. Brown (Eds.) (1982). *Intelligent Tutoring Systems*. Orlando, Florida: Academic Press, Inc.
Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*(1), 40-62.
Sottilare, R. & Gilbert, S. (2011). Considerations for tutoring, cognitive modeling, authoring and interaction design in serious games. *Authoring Simulation and Game-based Intelligent Tutoring workshop at the Artificial Intelligence in Education Conference (AIED) 2011*, Auckland, New Zealand, June 2011.
Sottilare, R., Holden, H., Brawner, K. & Goldberg, B. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *Interservice/Industry Training Systems & Education Conference*, Orlando, Florida, December 2011.
Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: U.S. Army Research Laboratory Human Research & Engineering Directorate (ARL-HRED).
Sottilare, R. (2012). Considerations in the development of an ontology for a Generalized Intelligent Framework for Tutoring. *International Defense & Homeland Security Simulation Workshop* in Proceedings of the I3M Conference. Vienna, Austria, September 2012.
Sottilare, R. (2013). Special Report: Adaptive Intelligent Tutoring System (ITS) Research in Support of the Army Learning Model- Research Outline. *Army Research Laboratory* (ARL-SR-0284), December 2013.
Sottilare, R., Ragusa, C., Hoffman, M. & Goldberg, B. (2013). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2013.
Sottilare, R., Brawner, K., Sinatra, A. & Johnston, J. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT). Orlando, FL: *US Army Research Laboratory*. May 2017. DOI: 10.13140/RG.2.2.12941.54244.
Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H., & Gilbert, S.B. (2018). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education, 28,* 225-264. DOI: 10.1007/s40593-017-0146-z.
Van Berio, M. P. (1997). Team Training vs. Team Building and Cooperative Learning: Defining the Field of Research (Team training vs team building en cooperatief leren: afbakening van het onderzoeksterrein) (No. TNO-TM-97-B019). HUMAN FACTORS RESEARCH INST TNO SOESTERBERG (NETHERLANDS).
VanLehn, K. (2006).The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education. 16*(3), 227-265.
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.
Woolf, B.P. (2009). *Building intelligent interactive tutors*. Burlington, MA: Morgan Kaufmann P

SECTION I - TEAM MODELING

Dr. Xiangen Hu, Ed.

Chapter 1 – Introduction to Team Modeling

Xiangen Hu
University of Memphis

Core Ideas

Intelligent Team Tutoring Systems (ITTSs) are a natural extension of traditional Intelligent Tutoring Systems (ITSs) used to instruct individual learners. This section examines the essential elements in team models. What should an ITTSs know about a team and its members in order for the ITTS to make effective instructional decisions? ITTSs are much more complicated than conventional ITS although one can see parallel research methodologies between ITSs and ITTSs. A natural and productive starting point for ITTSs research is to consider best practices from ITS research and examine opportunities to extend these practices to ITTSs.

We began our exploration of the different aspects of ITTSs using the Generalized Intelligent Framework for Tutoring (GIFT) as a model. GIFT has been one of the most comprehensive frameworks for ITS design, has a sizeable research community of interest and an extensive history of documented research and development during the past decade.

In traditional ITSs and in GIFT-based tutors, we are concerned primarily with the interaction between their four common components (learner model, instructional model, domain model, and user interface). For example, when considering learner modeling, how should we characterize (represent) the behaviors and progress toward learning objectives for individual learners in ITTSs. For teams, what mechanisms are required to collaborate either synchronously or asynchronously? Each learner in a team may have shifting roles and responsibilities in an ITTS as opposed to static objectives in a traditional ITS with a single individual learner. When roles are included in team modeling, we need to extend the competence measures such to include "collaborative" aspects (e.g., teamwork), but we also need to consider how the model that optimizes team organization and performance.

In addition to the learner's collaborative function and team organizational structure in ITTS, models need address other ITTS specific issues such as workload (distribution of tasks), socio-cultural dynamics, collaborative teaching/coaching models, and technological/interface models. The five chapters collected in this section demonstrate various aspects of teamwork (e.g., attitudes, behaviors, cognition, dynamics), explores how teamwork moderates learning, processes for developing shared mental models, and interpreting team dynamics.

Teams may take on many forms, but structure (how responsibilities, roles, and work are allocated) is only one important aspect of a team model. Team models include:

- Norms or standards to drive behaviors and expectations within the team
- Shared goals and objectives
- Processes to plan projects, share information, and coordinate collaborative work
- Processes to make decisions and criteria to evaluate decisions
- Motivators to positively influence team behaviors

The chapters in this first section exemplify the extension of traditional ITS models for individual learners to ITTSs models for teams and collaborative learning. In supporting this transition from individual learners to teams, we must also consider the impact of this transition on tutoring system components as suggested below:

- **Domain model** for ITTSs should consider knowledge/skill that is only valid/understood when more than one learners work together (e.g., collaborative problem solving skills
- **Pedagogical model** for ITTSs should consider cases where learners enhance their knowledge and skills by simply observing or interacting with other learners
- **Interface model** for ITTSs should consider multi-channel/multi-modal interaction with mixed (human and computers) in a shared learning environment.
- **Learner model** for ITTSs should consider how to separate individual learning objectives from those of the team

Individual Chapters

The Chapter by **Kim, Gorman,** and **Sottilare**, "Team learning and retention curves", proposes to use team learning and retention curves as components of a *learner model* to characterize of team learning. Given that teamwork involves multiple members, one research focus is to identify suitable ways to combine individual learning curve to form a team learning curve. The authors proposed a Bayesian hierarchical model to aggregate individual learning curve, and introduced several tools to implement these types of modeling. In ITTSs, these learning curve-based representation can inform the intelligent agents to provide better feedback to maximize the team learning gain.

The Chapter by **Dede, Grotzer, Kamarainen, Metcalf, Olney, Rus, Sottilare**, and **Wang**, "Graphical supports for collaboration: constructing shared mental models", describes several graphical supports, such as concept maps, 3d cognitive mapping, and self-visualization, which help team members to develop shared mental models. In particular, a collaborative construction of these graphical supports by all team members has been proven to be very effective for improving team learning. These new approaches extend the *domain* and *pedagogy* models of ITS, and ITTSs can leverage these findings by creating co-construction scenarios and tailoring the intelligent agents' interaction mechanisms to support these co-construction activities to promote team collaboration and learning.

The Chapter by **Ruis, Hampton, Goldberg,** and **Shaffer**, "Modeling processes of enculturation in team training" introduces network models of teamwork. Epistemic network analysis (ENA) provides good representations of the interactions among the team members but not directly encompasses the content information contained in the communication among the team members. The authors extended the traditional ENA framework by including the analysis of the communication contents and proposed a multilevel network model for teamwork. Such a modeling framework substantiates the *domain* and *pedagogy* models in collaborative environment and inform ITTS to better support team learning.

The Chapter by **Carlin, Rus,** and **Nye,** "Bayesian model of team training and measures", describes a framework for team training using a Partially Observable Markov Decision Process (POMDP) model, which is a good example of *pedagogy model* in ITTS. In this model, the estimated skill level is compared to the training objectives specified in the reward function and the difference informs proper selection of training actions to optimize the paths to achieve the training goal. The model can be directly integrated into the interaction mechanism of intelligent agents in ITTS to optimize the agent interaction in collaborative training tasks.

The Chapter by **Carlin, Perry,** and **Ostrander**, "Dynamic task selection for team task training using wearable sensors and multi-agent planning models" describes how a combination of the TEAMS modeling framework and new information from wearable sensors can help to optimize the task selection for team task training. Wearable sensors can provide team interactivity information in a less intrusive way, and the information allows more accurate and real-time knowledge of the state of the team interaction, which is an example of multimodal channels in the *interface model*. The authors show, through an example, that incorporating information from wearable sensors into the TAEMS, a framework that models complex computational task environment, will help to optimize the task selection for team training. Such an integration in ITTS allows intelligent agents to identify the state of teamwork more accurately and thus better support the team learning.

Chapter 2 – Team Learning and Retention Curves

Jong W. Kim[1], Jamie C. Gorman[2] & Robert Sottilare[1]
U.S. Army Research Laboratory[1], Georgia Institute of Technology[2]

Introduction

Performance change can be visualized by learning and retention curves. They are useful for the learner assessment. The learner can be assessed *during the course of training*, rather than at the end of a training course, which is called a formative assessment. The formative assessment approach would be particularly useful to see the progress of skill development and to identify when/how to disrupt the learning with tailored instructions and feedback in an intelligent tutoring system.

It is noted that two views can be associated with the understanding of team learning and retention curves. Measuring and analyzing team cognition can be considered as an aggregate of individual members, or they can be treated as an emergent property of interactions across team members (Cooke, Gorman, Duran, Myers, & Andrews, 2013). If we simply aggregate individual learning and performance to measure and analyze a collective of individuals' learning, we might leave out important, interaction-based features of team learning and performance.

People in a team perform both individual and interdependent tasks to achieve a team goal, which is usually accomplished through a series of multiple subtasks, each of which has its own history of learning and retention. Team learning is a complex problem that resembles a multithreaded computer architecture. An improved understanding of team learning and retention—i.e., an understanding from an individual to a collective of individuals, from a single task to a collective of decomposed subtasks, and from learning to forgetting—should be given consideration.

In this chapter, we start to look at these multi-level issues in team learning curves, and suggest directions for the design of adaptive instructional systems. We chose GIFT (Generalized Intelligent Framework for Tutoring) to explain the issues and suggestions related to team learning and retention curves since it provides a generalized framework for an intelligent tutoring system (Sottilare, Goldberg, Brawner, & Holden, 2012).

Learning Curves for Visualizing Performance Change

Learning curves have been used in industry in an attempt to account for the production time or cost (Jaber, 2016). In Cognitive Science and Education, learning curves have also played an important role to investigate practice effects in human memory systems (Anderson, Fincham, & Douglass, 1999). In such cases, learning curves follow a type of mathematical form that is generally known as a power function or an exponential function (Newell & Rosenbloom, 1981; Rosenbloom & Newell, 1987). The visualization based on the mathematical function exhibits an elegant way to represent the progress of learning. It helps our understanding of human learning. But, in a practical sense, it is getting harder and more complicated to fit this model due to the adaptive feature of the system and the hierarchical complexity of the task—e.g., different (or adaptive) contents in repeated practices, the large task sizes, and the complex tasks with multiple subtasks.

The progress—e.g., performance improvement by deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993)—can be visualized and summarized as a mathematical form, generally known as a power law of learning curve (e.g., Anderson, Fincham, & Douglass, 1999; Newell & Rosenbloom, 1981; Rosenbloom & Newell, 1987). A cognitive architectural process model that usually focuses on an individual learning process has provided useful information about how knowledge is acquired and stored in memory, and is retrieved to perform a task (Anderson, Boyle, Corbett, & Lewis, 1990).

Similarly, a team is considered as a collective of individuals, and a team also learns and forgets knowledge and skills over extended periods of time. Team learning and retention curves are expected to provide useful information about performance improvements/decrements in terms of practice and disuse of skill. These curves can thus be used to make strategic decisions on training interventions—i.e., what, how, and when to train. As shown in Figure 1, the task completion time is a dependent measure in assessment, and it also follows a power law of learning representing a speed-up effect (Kim & Ritter, 2015).

There are identified problems in using learning curves. For example, this learning curve can be obtained only if the same task should be repeatedly performed. It has been noted that team member interaction is one promising factor to account for team performance decay and retention rather than individual competency, arguing team performance is more than the sum of individual team member performance (Cooke, Gorman, Duran, Myers, & Andrews, 2013). We, then, can infer that collective competency of a team would not be necessarily accounted for by aggregating all individual competencies. A model of individual competency is important, but it is definitely necessary to have an advanced understanding from the team perspective. Among various aspects of a team (e.g., team cohesion, coordination, team cognition, and team communication, etc.), team cognition plays important roles in team effectiveness, and one form of team cognition is team mental model (Klimoski & Mohammed, 1994). As to team learning, it is also necessary to consider the roles in a team task. It is also difficult to keep experimental teams together long enough to measure skill retention over periods of time (Cooke, Gorman, Duran, Myers, & Andrews, 2013). Thus, a simple comparison of learning (retention) curves in a large task and in a team (a collective of individuals) seems to be insufficient and challenging. Also, one of the reported pitfalls of learning curves is that a larger domain model or a large student sample size is likely to exhibit a better fit than a smaller one, even if the system does not operationally teach the students any better (Martin, Mitrovic, Koedinger, & Mathan, 2011). Furthermore, a near-term assessment by comparing learning curves would not be related to the long-term stability of learning (Schmidt & Bjork, 1992).

Considerations for Team Learning Curves

Concerns about Aggregate Data

Figure 1 shows a learning curve from a dataset which was collected from 30 human participants performing a set of spreadsheet task (named the Dismal task) consisting of subtasks including: open a file, save a the file, calculate normalization, calculate frequency, calculate length, insert rows, save as a printable format (Kim & Ritter, 2015; Paik, Kim, Ritter, & Reitter, 2015). It is an instruction-following task by human participants and an ACT-R model. The learning curve shows a steep decrease at the beginning in the task completion time over trials, and a asymptote-like plateau at the later trials, which appears to follow a power law of learning. The curve was generated by averaging all data points.

Theoretically, if individuals do the same tasks in a given time frame, then performance data can be averaged to represent performance change and learning, which is considered as a summary of the learning experience. Usually, tasks are large and complex in real world. They would be also hierarchically structured. That is, tasks can be decomposed to a number of sub-levels (Lee & Anderson, 2001). Each level of tasks (or subtasks) can have a learning curve under the assumption that the learner performs the same tasks repeatedly. The learning curves by decomposed subtasks can be plotted, showing subtasks would be learned differently (Kim & Ritter, 2016). In this situation, there would be an issue of data aggregation. Aggregation of different aspects of learning data would be getting harder and more challenging in terms of task complexity.

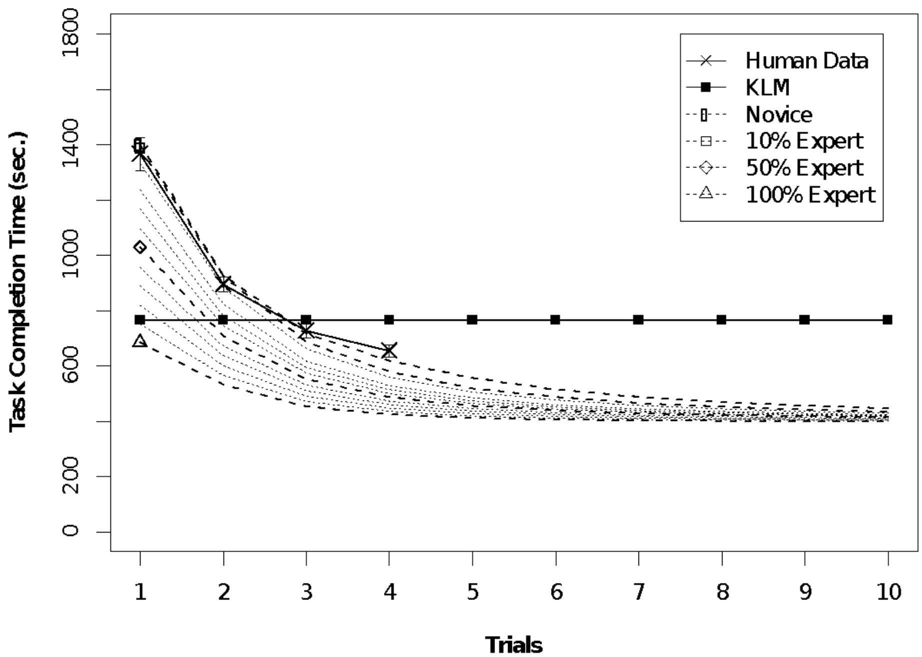


Figure 1. ACT-R models of the Dismal task (dashed lines, from fully novice to previously practiced expert), along with human aggregate data (X's and SEM error bars, N=30), and the KLM prediction in solid line (taken from Paik, Kim, Ritter, & Reitter, 2015).

Aggregation of the data refers to numerical or non-numerical information that is collected from multiple sources and on multiple measures, variables, or individuals, and compiled into a data summary. In a statistical analysis, the aggregation problem has been noted that information can be lost when the microlevel individual data is substituted for the aggregate, macro level data (Clark & Avery, 1976). A team can be decomposed into a group of individual members, and a task performed by the team can be decomposed into subtasks, and taskwork (i.e., working on a specific duty of one's job) and teamwork (i.e., coordination or communication by individuals to achieve a mutual goal) (Salas et al., 2015. pp. 5) can be decomposed by individuals. Thus, it is necessary, with regard to "team learning and retention curves", to consider the macro and micro level features of a team level performance.

Features of Team Learning Curves

Averaging is the most common statistical technique but it can potentially mislead our understanding about the learning data. The representativeness of averaged learning curves has been repeatedly an issue in Psychology (Brown & Heathcote, 2003). Averaging the learning data to understand an individual learning and performance can be distorted because of its nonlinearity and variable features of the parameters. Regarding the team level, the concern about averaging across and within individuals to understand the team may be more complex and challenging. That is, a team, as a collective of individuals, would have different roles and taskwork to do, and coordination and communication among individuals are essential to accomplish a shared goal (Salas et al., 2015). Understanding such features would be useful to guide us to walk through the unprecedented territory of team learning and its visualization.

Team knowledge is described as an emergent phenomena resulting from team members' interaction (Klimoski & Mohammed, 1994). This perspective places a greater emphasis on team processes and allows for the possibility that team knowledge is something different than the sum of the team's individual member's knowledge. Interactive team communications can be understood by a verbal and textual transfer of team knowledge (Gorman, Foltz, Kiekel, Martin, & Cooke, 2003).

Measurement of team knowledge can be observed in a study of the team communication analysis. Semantic content analysis is one example in team communications. For example, the UAV (unmanned aerial vehicle) knowledge for use can be analyzed by using a Latent Semantic Analysis (Gorman, Foltz, Kiekel, Martin, & Cooke, 2003). In the study by Gorman and his associates, the semantic space is decomposed in terms of geometric interpretation: (a) the length vector which is represented by the sum of all the words in a sentence, and (b) the cosine value to specify how two statements (or utterances) are semantically related in the semantic space. One of the metrics used in the study is communication density indicating the average task relevance of a team's communications—the ratio of the length vector to the number of words spoken during the mission. In this context, the learning effect can be derived by observing differences in semantic contents by experienced and inexperienced teams—how a novice team learns to communicate, leading to increase in communication density. This notion gives us one aspect of multiple facets of a team learning curve. Similar to the understanding of the team communication that can be decomposed to a semantic space, the multiple facets of team features is worth being decomposed to a domain for an improved analysis and visualization. As another example, a team success can be predicted by power spectral density analysis that decomposes the task-oriented dialogue into the frequency domain in time (Xu & Reitter, 2017).

A Statistical Model for multilevel Learning curves

Different tasks and subtasks at a team level with different roles would be differently learned and retained in our memory (i.e., declarative and procedural memory). A meaningful decomposition can be useful to analyze the phenomenon. Figure 2 shows the varying performance of the decomposed tasks shown in Figure 1. The subtasks have different task completion times and mean/median values both by tasks and by the aggregated participants. This understanding would affect the perspectives of adaptive instructional sciences, and training in various professions.

However, there have been rather sparse investigations to address the question of how consistently (or inconsistently) tasks or subtasks are learned at a team level. A recent investigation tested the inconsistency of the individual-level subtask learning, providing an understanding of how complex tasks can be decomposed, and of how a probabilistic model of individual level of subtask learning can be estimated using a Bayesian hierarchical modeling approach (Anglim & Wynton, 2015). The knowledge components in the domain and learner model at a team level should support performance assessment with team learning (and retention) curves. A statistical modeling approach of multi-level facets of a task can be useful to comprehend the team. In this section, we introduce a statistical model of learning and performance both by tasks (subtasks) and by a team (individuals).

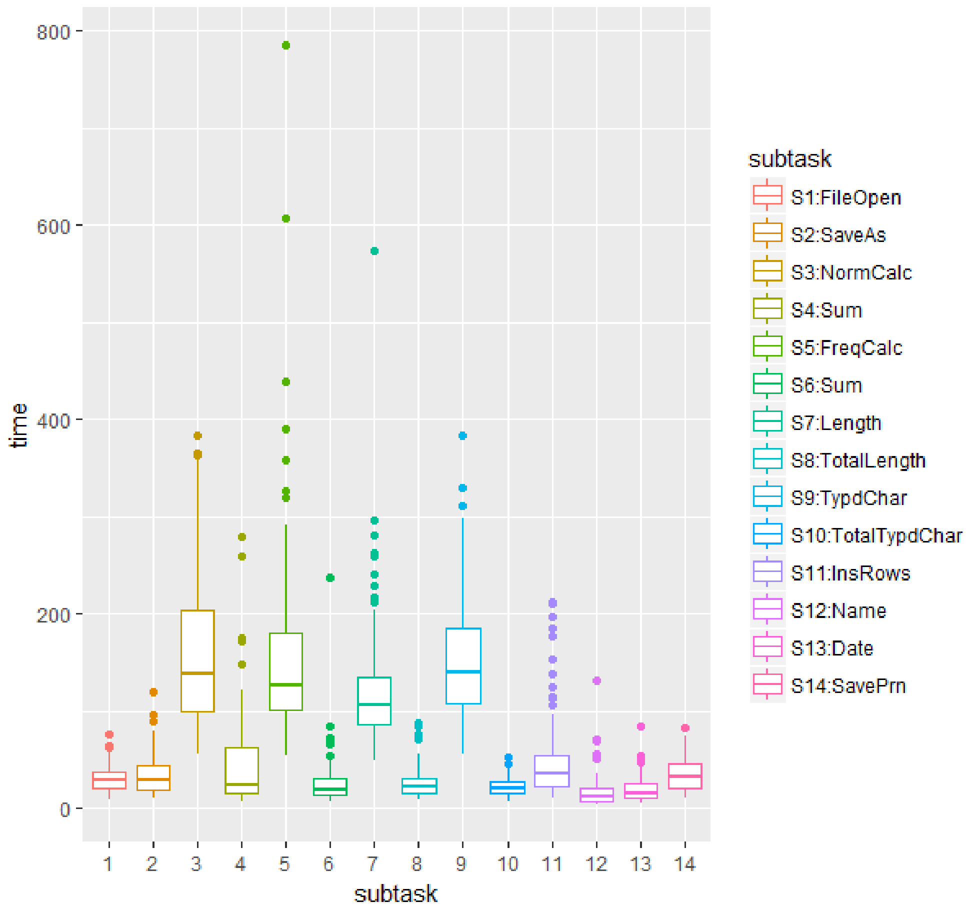


Figure 2. A box plot to describe the distributions of the task completion times, showing minimum, maximum, median values. Among the subtasks, S3, S5, S7, and S9 have the median task completion times above 100 s.

A Statistical Model for Aggregating Learning Curves

A Bayesian hierarchical modeling approach can be applied to investigate performance change by tasks and by individuals by teams (as a collective of individuals). Hierarchical models are interchangeably used with mixed models and random effect models, indicating that individual subtasks and individual participants have their own learning rates (e.g., amount learned, rate of learning, final performance, and variability around expected performance; (Anglim & Wynton, 2015). The Bayesian hierarchical modeling approach is particularly suited to the data analytics of a repeated measures design (Averell & Heathcote, 2011), which can investigate learning. A likelihood-based inference, for small sample sizes, can be generally unreliable with variance components that are particularly difficult to estimate. Bayesian modeling would be beneficial if there is not enough data for inferential statistics when we are interested in differences between individual participants and items (random effects and random slopes) (Sorensen, Hohenstein, & Vasishth, 2016). This approach can resolve a concern about the number of sample sizes (Lavine, 1999).

The aforementioned Dismal spreadsheet task data was collected in a repeated measures design, and the data should be analyzed to identify differences between individual subjects and tasks (random intercepts and random slopes), and correlations between variance components (random effects). Thus, a Bayesian hierarchical modeling approach is appropriate for the understanding of the task with its subtasks. Similarly, this approach can be applied to a team consisting of a collective of individuals, and the taskwork performed by those individuals. In this approach, the posterior distribution of each parameter of interest is derived, given some data and prior knowledge about the distributions of the parameters.

There are a number of tools available that support this type of analysis. One such tool is *R*[1], a computational statistical language, and a probabilistic programming language, *Stan*[2]. Previously, the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) in R was used to conduct a linear mixed effects analysis of the relationship between the response variable and the covariate predictors including fixed and random effects (Kim & Ritter, 2016). To utilize the capability of the Monte Carlo Markov Chain (MCMC) sampling, *Stan* can be used with *R* together (McElreath, 2016).

Based on a simple fixed effect model, one can proceed to implement complex linear mixed effect models by adding varying intercepts and slopes. This model will support an analysis of the correlation between the varying intercepts and varying slopes, which is noted as the maximal model (Barr, Levy, Scheepers, & Tily, 2013). Example models are shown in Table 1. These probability models describe how the dependent variable (e.g., the task completion time or other performance measures) can be generated in a team task performance, which allows us to derive the posterior probability distribution of the model parameters from a prior distribution and the likelihood function. The model in Table 1 is summarized that the dependent variable by-subtasks and by-individuals (or team members) are varying in terms of practice trials. Multiple measurements from each individual team member and each task (or subtask) can cause correlation between errors. The proposed statistical model can be examined with consideration of varying intercepts and varying slopes, and the correlation between varying intercepts and slopes.

The estimated parameters in the fixed effects model do not vary from an individual to an individual (i.e., team members) and from subtask to subtask (i.e., different tasks by the team). The dependent variable (i.e., the task completion time) is represented by Yij. The index i represents the i-th row in the data frame ($i \in \{1, \dots, N\}$), and the term εij represents the error in the i-th row, and a practice day j ($j \in \{1, \dots, J\}$). This model is a fixed effect model since the parameters of β_0, and β_1 do not vary from individuals in a team (pID) and multi-level tasks (St). The model is represented as a simple linear model with a predicator (X_j) of practice trials and the task completion time (Y_{ij}) as the dependent variable. We assume that the error term (ε_{ij}) is normally distributed with mean zero and unknown standard deviation, σe, and that it is identically distributed, indicating there is no correlation between errors.

$$Y_{ij}=\beta_0+\beta_1 X_j+ \varepsilon_{ij} \qquad \text{Model (1)}$$

It is necessary to check the normality assumption of the data. The Q-Q plot of residuals shows whether the data are normally distributed. If the data is not normally distributed, a log-transformation of the data would be necessary. To assess the significance of practice trials (day) as a predictor, the t-value of the fixed effects can be used. If the t-value of the slope estimate is large enough, you can estimate that the predictor is significant. Comparisons of the models can be performed and the Bayesian Information Criterion (BIC) can be used as a test statistic. In this manner, a team relevant statistical model can be built to visualize performance change in terms of varying tasks and individuals.

Table 1. A summary of the Bayesian hierarchical models for a team.

1 https://cran.r-project.org/

2 http://mc-stan.org/

No.	Model	*lmer* model syntax
1	$Yij=\beta 0+\beta 1Xj+\ \varepsilon ij$	
2	$Yij=\beta 0+pID0i+\beta 1Xj+\ \varepsilon ij$	Y~X+(1|participant)
3	$Yij=\beta 0+pID0i+(\beta 1+pID1i)Xj+\ \varepsilon ij$	Y~X+(1+X|participant)
4	$Yij=\beta 0+pID0i+St0i+(\beta 1+pID1i)Xj+\ \varepsilon ij$	Y~X+(participant)+(1|Subtask)
5	$Yij=\beta 0+pID0i+St0i+\beta 1Xj+\ \varepsilon ij$	Y~X+(participant)+(1|subtask)
6	$Yij=\beta 0+St0i+(\beta 1+pID1i)Xj+\ \varepsilon ij$	Y~X+(participant)+(1|subtask)

Discussion and Conclusions

We explored what to consider regarding team learning and retention curves. We note that data aggregation would cause distortion of our understanding toward individual and team learning (Brown & Heathcote, 2003). Thus, features of team should be considered to better deal with the representativeness of learning curves. We introduced a statistical model for team, which would be suitable to incorporate the aforementioned team features characterized by hierarchical and multi-level tasks with varying roles by a collective individuals. The model also supports a Bayesian comparison (Sorensen, Hohenstein, & Vasishth, 2016), indicating increase in statistical plausibility of comparing learning and retention curves. Therefore, the presented statistical model can reduce the data aggregation problem, and help to identify how team learning and retention curves would be different from individual ones.

Our exploration can be applied to improve the capability of GIFT (Generalized Intelligent Framework for Tutoring). GIFT is a module-based architecture, and its modules (e.g., Domain, Learner, Pedagogical, Gateway, and Sensor) are interchangeable and linked together by a message bus (Sottilare, Brawner, Sinatra, & Johnston, 2017; Sottilare, Goldberg, Brawner, & Holden, 2012). GIFT offers an operational mechanism of the real time assessment. The learning curves can be derived from the interactive operations of the modules. In GIFT, the Domain Knowledge File (DKF) plays an important role in the assessment structure by tracking performance of an individual.

Assessment is important to determine an appropriate intervention of training. Visualizing such assessments can be useful to quantify the progress, and it reciprocally helps to determine remediation of adaptive learning. Based on the individual assessment structure, GIFT can be extended to afford multiple simultaneous DKFs, called the team DKF (Brawner, Sinatra, & Gilbert, 2018). The team DKF can track a collective of individuals and multi-level tasks with varying roles. This would improve the capability of team assessment in GIFT. That is, team learning and performance are measured while the team is engaged in a training scenario during a lesson. In general, the assessment element contains rules and configurations for the Domain module to assess the learner actions. A concept is the lowest level performance node that is reported between modules and stored in a database. A concept is assessed by conditions that contains a logic to assess the learner's performance specified in the Domain module. The concepts and conditions are hierarchically organized. The team training framework needs to support variable domains, content areas, multiple team configurations, and multiple types of team performance assessments (Brawner, Sinatra, & Gilbert, 2018). The GIFT's modular structure of design alternatives for team has been proposed with an understanding of the hierarchical inheritance approach (Brawner, Sinatra, & Gilbert, 2018), and this structural taxonomy to a team modeling would generate interactions among the modules.

Acknowledgements

Research described herein was supported through a post-doctoral fellowship sponsored by the US Army Research Laboratory under contract to the Oak Ridge Associated Universities (ORAU). The authors wish to thank Kaitlyn Ouverson for her useful comments on this chapter.

References

Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, *42*, 7-49.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1120-1136.

Anglim, J., & Wynton, S. K. A. (2015). Hierarchical Bayesian models of subtask learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(4), 957-974.

Averell, L., & Heathcote, A. (2011). The form of the forgetting curve and the fate of memories. *Journal of Mathematical Psychology, 55*(1), 25-35.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Brawner, K., Sinatra, A., & Gilbert, S. (2018). Lessons learned creating a team tutoring architecture: Design reconsiderations. In *Design Recommendations for Intelligent Tutoring Systems: Teams* (Vol. 6). Orlando, FL: U.S. Army Research Laboratory.

Brown, S., & Heathcote, A. (2003). Averaging learning curves across and within participants. *Behavior and Research Methods, Instruments, & Computers*, *35*(1), 11-21.

Clark, W. A., & Avery, K. L. (1976). The effects of data aggregation in statistical analysis. *Geographical Analysis, 8*(4), 428-438.

Cooke, N. J., Gorman, J. C., Duran, J., Myers, C. W., & Andrews, D. (2013). Retention of team coordination skill. In W. Arthur Jr., E. A. Day, W. Bennett Jr. & A. M. Portrey (Eds.), *Individual and team skill decay* (pp. 344-363). New York, NY: Routledge.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363-406.

Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of Latent Semantic Analysis-based measures of team communications content. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (pp. 424-428). SAGE Publications.

Jaber, M. Y. (2016). *Learning curves: Theory, models, and applications*: CRC Press.

Kim, J. W., & Ritter, F. E. (2015). Learning, forgetting, and relearning for keystroke- and mouse-driven tasks: Relearning is important. *Human-Computer Interaction, 30*(1), 1-33.

Kim, J. W., & Ritter, F. E. (2016). Microgenetic analysis of learning a task: Its implications to cognitive modeling. In F. E. Ritter & D. Reitter (Eds.), *Proceedings of the 14th International Conference on Cognitive Modeling* (pp. 21-26). University Park, PA: Penn State.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*(2), 403-437.

Lavine, M. (1999). What is Bayesian statistics and why everything else is wrong. *The Journal of Undergraduate Mathematics and its applications, 20*(2), 165-174.

Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex? A study in learning decomposition. *Cognitive Psychology*, *42*(3), 267-316.

Martin, B., Mitrovic, A., Koedinger, K. R., & Mathan, S. (2011). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction, 21*(3), 249-283.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan* (Vol. 122): CRC Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum.

Paik, J., Kim, J. W., Ritter, F., & Reitter, D. (2015). Predicting user performance and learning in human-computer interaction with the Herbal compiler. *ACM Transactions on Computer-Human Interaction, 22*(5), 1-26.

Rosenbloom, P., & Newell, A. (1987). Learning by chunking: A production system model of practice. In D. Klahr, P. Langley & R. Neches (Eds.), *Production system models of learning and development* (pp. 221-286). Cambridge, MA: MIT Press.

Salas, E., Benishek, L., Coultas, C., Dietz, A., Grossman, R., Lazzara, E., & Oglesby, J. (2015). *Team training essential: A research-based guide*. New York, NY: Routledge.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology, 12*(3), 175-200.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). *An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: US Army Research Laboratory.

Sottilare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference* (pp. 1-13).

Xu, Y., & Reitter, D. (2017). Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 623-633). Vancouver, CA.

Chapter 3 – Graphical Supports for Collaboration: Constructing Shared Mental Models

Christopher Dede[1], Tina A. Grotzer[1], Amy Kamarainen[1], Shari J. Metcalf[1], Andrew M. Olney[2], Vasile Rus[2], Robert A. Sottilare[3], and Minhong Wang[4]
Harvard Graduate School of Education[1], University of Memphis[2],
U.S. Army Research Laboratory[3], University of Hong Kong[4]

INTRODUCTION

Graphical images are useful in helping teams develop shared mental models, particularly when digital tools enable team members to co-construct these representations. For example, the use of concept maps as external representations of knowledge, in a form that can be manipulated and reasoned with, can clarify thinking, focus a task, facilitate collaboration, and reduce cognitive load. This chapter describes several types of graphical supports—concept maps, 3-dimensional cognitive mapping, and self-visualizations—as ways to enable the collaborative construction of shared mental models.

Concept Mapping as a Representational Structure that Enables Shared Mental Models

Concept mapping as a representational structure is an effective mechanism for students to express their conceptual understanding (Novak, 1990; Rice, Ryan, & Samson, 1998; Rosen & Tager, 2014; Toth, Suthers, & Lesgold, 2002). The use of concept maps as external representations of knowledge, in a form that can be manipulated and reasoned with, can clarify thinking, focus a task, facilitate collaboration, and reduce cognitive load (Cox, 1999; Jonassen, 2003). In particular, computer learning environments for inquiry-based science learning have the opportunity to include electronic concept mapping as a knowledge construction tool. Students engaging with computer based simulations and virtual worlds may develop a deeper understanding about the dynamic systems represented in these environments through engaging in concept mapping of the system as a map of causal relationships between its factors. Michael Zeilik's website (http://archive.wceruw.org/cl1/flag/cat/conmap/conmap1.htm)

EcoXPT is a multi-user virtual environment (MUVE)-based middle school science curriculum that supports learning about the causal dynamics of ecosystems through observation, exploration, and experimentation in a virtual world (Dede, 2017; Grotzer, 2017). It builds and expands upon earlier research with EcoMUVE (Metcalf et al., 2011, Grotzer et al., 2013). EcoXPT supports situated learning – students conduct scientific inquiry while immersed in the richly represented ecological setting, interacting with virtual people and organisms, collecting and analyzing data. As a culminating activity, student teams engage in construction of a concept map as a representation of the components, processes and relationships relevant to the phenomena identified in the ecological scenario.

In prior research with EcoMUVE, students were instructed to draw concept maps on paper. For EcoXPT, it was hypothesized that students might be better supported with an integrated computer-based concept map tool. In particular, it was considered that providing a structured tool for concept mapping might scaffold students in concept map construction to support deeper learning about the causal relationships in the system. Additionally, by integrating the tool with the virtual environment, it might support students making more connections between data collection, data analysis, and hypothesis constructing activities.

The EcoXPT curriculum consists of a two week, inquiry-based unit centered on a virtual pond and the surrounding watershed. Students explore the pond, learn about the plants and animals in the ecosystem, and travel in time to see changes over the course of a virtual summer. They discover on one day that all of the large fish have died, and are given the inquiry task of figuring out why it happened. The system models an eutrophication scenario in which fertilizer runoff induces excessive algae growth. Then the algae die off and, as bacteria decompose the dead matter, the bacterial population surges. This combination of factors--coupled with weather conditions of warm temperatures, cloudy days, and low wind--lead one evening to dramatically lowered dissolved oxygen in the pond that causes the death of all of the bluegill and large-mouth bass in the pond, although the minnows, which can survive in relatively low dissolved oxygen conditions, do survive.

EcoXPT includes a range of integrated tools that support students in learning about the pond and its surroundings. Students can make observations, photograph organisms in and around the pond, shrink to view microscopic organisms, and travel in time. They collect measurements about the water (e.g., phosphates, temperature, dissolved oxygen), weather (e.g., wind speed, cloud cover), and populations of organisms (including three species of fish, two types of algae, and bacteria); then the view graphs showing trends in the data over time. They use reference tools such as an online field guide and an atom tracker, and gather testimony from characters in the world. The pilot version of the curriculum used in this study also included a lesson on drawing concept maps to represent the causal relationships in an ecosystem.

Students collaborate in teams to solve the mystery, working together to collect and analyze data in the virtual world. As a culminating activity, each team of students constructs and presents to the class a poster that represents their hypotheses to explain why the large fish in the pond died. The team posters are required to include a concept map, a written summary of their hypothesis, and printouts of the evidence they used to support their ideas.

This study piloted a new electronic concept mapping tool that was designed to scaffold students' concept mapping activities. The tool (Figure 1) provided a pre-defined palette of factors designed to represent all of the variables that student were able to observe or measure in the virtual world. Students drag factors out of the palette to place them as nodes in the concept map, and drag a link from one node to another to create connecting arrows representing causal relationships between factors.

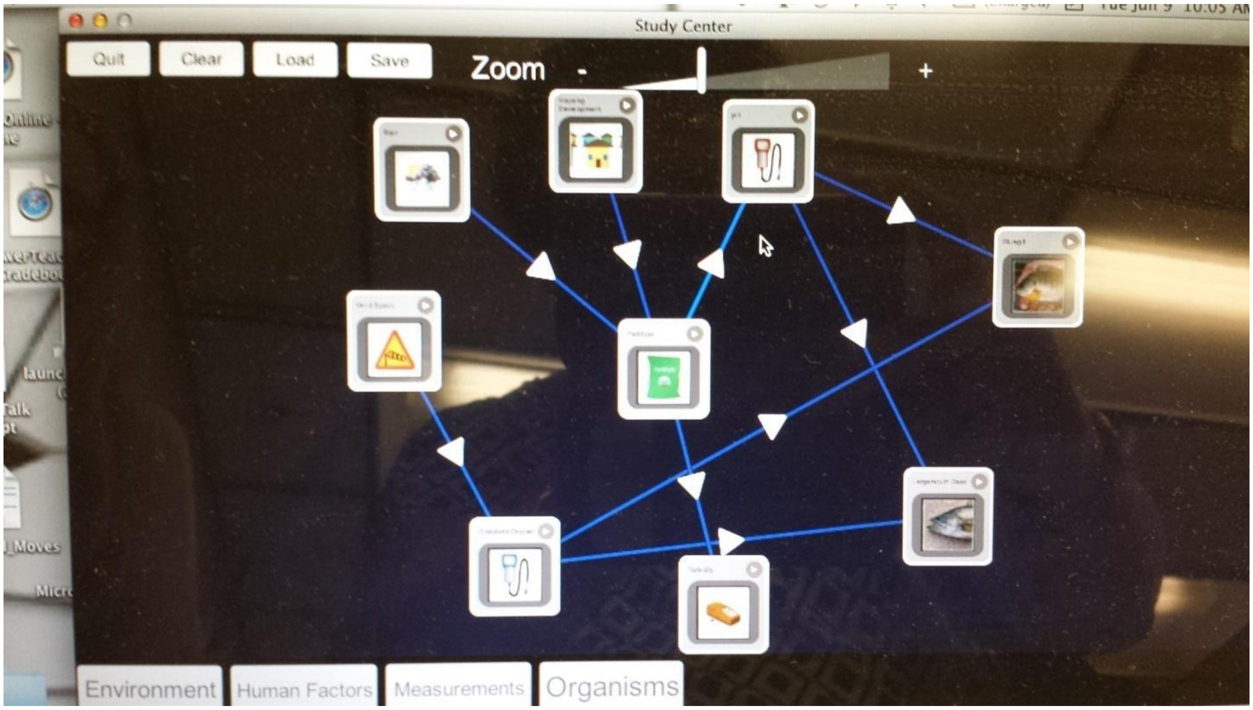


Figure 1: Photo of concept map being constructed on screen

In a pilot study, students who used the electronic concept map tool constructed larger and more complex concept maps than similar teams doing concept mapping on paper. This finding appears to support the proposal that providing external structure for concept mapping can scaffold collaborative activity by teams of students. In particular, the increase in number nodes in the electronic concept maps is likely caused by the fact that the software tool provided students with a large set of pre-defined nodes to use, potentially fostering convergence.

3-Dimensional Cognitive Mapping as a Representational Structure that Enables Shared Mental Models

In inquiry-based learning contexts, many students experience difficulties managing the complex inquiry process and engaging in fruitful inquiry learning. The inquiry process often involves iterative cycles of gathering information through observation or experiments, generating hypotheses, reasoning based on the collected information, and drawing conclusions. Many students find it cognitively demanding to integrate problem data with subject knowledge and to reason with intricately intertwined data. It is therefore necessary to guide students through the complex inquiry process to help them become accomplished problem-solvers.

To facilitate complex inquiry without undermining the nature of student-centered learning, indirect instructions such as prompts, hints, and scripts are used to bring learners' attention to important issues (e.g., what to do next) during the task, or a complex task is decomposed into a set of main actions or key questions. Recent research highlights the importance of making cognitive processes visible in complex problem or task situations. Related work involves the use of mental models for high-order thinking and in-depth learning, such as concept maps representing the relationships between concepts, causal maps representing the relationships of cause and effect, and evidence maps linking evidence with claims or hypotheses. In view of the need for integrating multiple aspects of cognitive processes in exploring a problem, integrated cognitive maps for example by representing the problem-solving process and the knowledge underlying the problem-solving process have shown promising effects (Wang Wu, Kinshuk, Chen, & Spector, 2013; Wu, Wang, Grotzer, Liu, & Johnson, 2016). In this section, we introduce a novel three-dimensional cognitive mapping (3DCM) approach, which makes complex inquiry visible and accessible to students by allowing them to externalize the information on a problem, the subject knowledge underlying the problem, and the hypothesizing and reasoning process involved in exploring the problem in a single image for effective thinking, action, and reflection (Chen, Wang, Dede, & Grotzer, 2017). As shown in Figure 2, the integrated cognitive map consists of three parts: a concept map, a data table, and a reasoning map. The concept map represents the subject knowledge underlying the problem in a set of interrelated concepts. The data table outlines the problem information in a set of key variables and their changes over time. In the reasoning map, each hypothesis is supported ("for") or rejected ("against") by evidence from the problem data or subject knowledge. To examine the root cause of the problem, the hypothesis is further explained by other hypotheses explicating deeper causes of the problem.

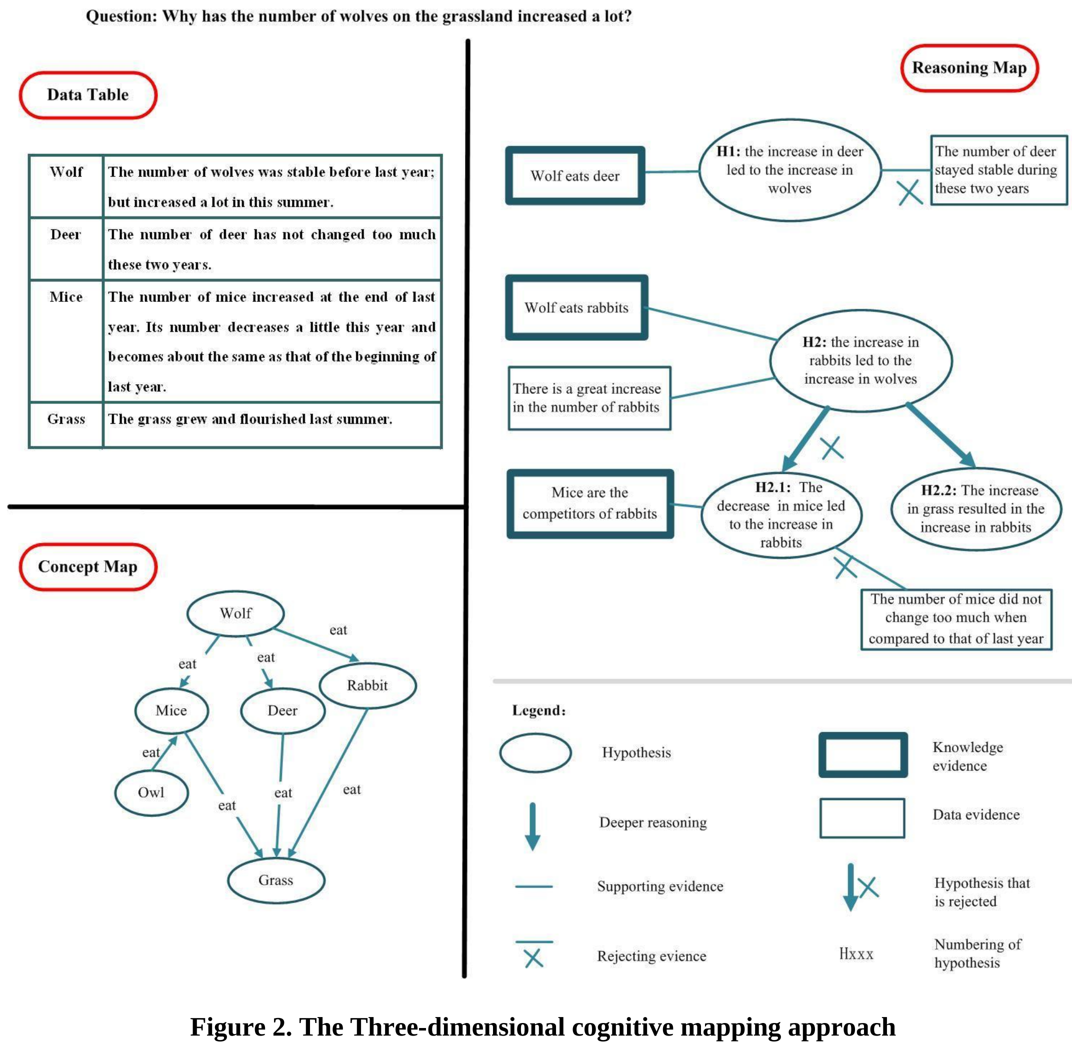


Figure 2. The Three-dimensional cognitive mapping approach

Forty-eight students (24 males and 24 females) from one 11th grade high school class participated in the study. They were classified into three categories of academic ability according to their pre-test scores: high, medium, and low, with each category having 16 students. Students were randomly divided into 16 small groups of 3 (i.e., one high-level, one medium-level, and one low-level student). They explored a fish die-off problem (why many large fish in a pond ecosystem had suddenly died) by performing causal reasoning and construct logical and scientific explanations. To do so, the students interacted with a virtual pond system to collect relevant information and observe changes in multiple variables over time. They discussed and solved the problem in small groups by evaluating and compiling the collected information, formulating and justifying hypotheses, and making conclusions. Students were asked to create a three-dimensional cognitive map to assist their inquiry, and submitted an inquiry report including hypotheses, reasoning, and conclusions.

Pre- and post-knowledge tests were administered to assess students' knowledge of the learning subject. A post-test questionnaire was used to measure students' attitudes toward inquiry learning, anxiety level, and confidence level. The results show that the participants displayed a high level of knowledge gain, positive attitudes, low anxiety, and medium levels of confidence. The interview records reveal that the 3DCM approach provided learners with a holistic view of the inquiry task, and guided them in generating hypotheses step-by-step and developing evidence-based reasoning based on relevant data and knowledge. Moreover, a post-hoc test indicated that the students at a low academic level had acquired significantly more knowledge than either the high-level or medium-level students, thus narrowing the academic gap between low-level, medium-level, and high-level students. Taken together, these findings show promising effects of the 3DCM approach in supporting inquiry learning.

Group Construction of Concept Maps as an Aid to Collaboration

Group construction of concept maps is highly effective for learning. According to a meta-analysis, concept maps have greater learning gains when constructed in groups (d = .96) than when constructed individually (d = .82) or studied individually (d = .37) (Nesbit and Adesope, 2006). The advantage for group construction over individual construction or study may derive from the fact that group construction implicitly combines both of these activities: each learner has both the opportunity to extend the group concept map and the opportunity to study what other members have contributed. The kind of support afforded by group construction appears similar to, but distinct from, the kind of scaffolding provided by so-called expert skeleton concept maps (Novak and Canas, 2006), which are partially specified concept maps with unlabeled nodes and/or edges. Learning gains with skeleton maps appear mixed (Chang, Sun, & Chen, 2001; Wang et al., 2015) and there is some evidence suggesting that skeleton maps assess a different kind of understanding than regularly constructed concept maps (Ruiz-Primo et al., 2001).

Group concept maps conceptually represent the collective individual "mental maps" of the group. Recent work has proposed a framework of "concept landscapes" to analyze collections of individual maps on a shared topic (Muhling, 2017). Two kinds of aggregation are proposed to create concept landscapes. The first is accumulation, a process by which individual maps are related by similarity, shared nodes, or shared edges. A landscape of accumulated maps is perhaps more appropriately considered as a high-dimensional space, e.g., node-space or edge-space, though a low dimensional landscape could also be formed by transforming the map similarity matrix with a technique like multidimensional scaling. The second concept landscape aggregation is called amalgamation. Amalgamation combines individual maps into a single map by merging nodes and edges, weighting them by their frequency across maps, and then pruning the combined map using a technique like thresholding, minimal spanning trees, or the Pathfinder algorithm (Schvaneveldt, 1990).

Group construction of concept maps appears to follow a process of amalgamation, with the important difference that the individual maps are not created before combining. Instead, individual maps are amalgamated incrementally. Through the processes of conversation and shared group map editing, individual mental maps become aligned, nodes and edges are drawn, and less important (or contentious) elements pruned. When considered as amalgamation, group construction of concept maps need not happen synchronously or with a stable group. However, an amalgamation-only view ignores the social context and common ground created by synchronous group construction, and these components would need to be replaced in a group construction paradigm operating asynchronously. Recent work has investigated asynchronous group construction by using a learning companion intelligent agent in place of a human peer (Olney & Cade, 2015). In this work, the learning companion and human student iteratively grow and expand a concept map that is based on previous student interactions. The learning companion effectively replaces the larger group of students by presenting elements of their maps as its own and providing a social presence for the student to engage with. Whether using a learning companion to simulate synchronous group construction is as effective as true synchronous group construction is an ongoing issue for research.

Self-Visualizations as Graphical Representations of Mental Models

Self-visualizations, i.e., graphical representations of one's mental model, have been used in science and computer programming education both as instructional tools and for assessing learners' understanding. For instance, generating a graphical representation during problem solving in conceptual Newtonian Physics plays a similar role as self-explanations, allowing learners to reflect on their understanding of the target concepts as well as enabling the instructor, e.g., a computer-based intelligent tutoring system, to assess learner's understanding and provide hints, for instance, by highlighting elements of the visual representation. That is, the graphical representation is used by the tutor-tutee team to advance their mission of maximal knowledge transfer. We argue that these self-visualizations play a similar role in understanding as self-explanations do. Furthermore, they could help team members understand each other's contributions to, for instance, the team's mission which could be maximal learning or developing a high-quality, i.e., bug-free, complex software product. The latter case offers a unique opportunity to investigate the role of visualizations for collaborative work as large software development projects usually involve very large teams comprising of hundreds or thousands of members with different time, space, and cultural characteristics.

We present next a brief summary of using graphical representation to externalize one's mental models in science learning and computer programming. Examples of such self-visualizations are free-drawings, which allow learners to freely express visually their thinking and understanding, and Control structure diagrams (CSD; Hendrix, Cross II, & Maghsoodloo; 2002). We review briefly previous work on self-visualizations next, that is, we focus on visualizations generated by a target individual as opposed to visualizations generated by an expert through interviewing the individual. Due to space reasons, we do not present visualizations at higher levels of granularity such as system diagrams highlighting the high-level organization of a complex software product. It should be noted that in large software development teams, visualizations play an important role for many aspects of this gargantuan collaborative effort including tracking changes to code, highlighting one's role in the overall team, and training newcomers on the current state of the project (Ellis, Wahid, Danis, & Kellogg, 2007).

The use of visualizations, i.e., free-body diagrams (see Figure 3), as an instructional strategy for improving students' ability to explain and predict Physics situations has been reported by Mualem and Eylon (2010). Mualem and Eylon reported significant learning gains (pretest-posttest) when this strategy was used to coach 9th graders on qualitative problem solving. In another study, Larkin and Simon (1987) showed that translating a propositional problem description into a visual representation is essential in Physics problem solving. Biswas, Leelawong, Schwartz, and Vye (2005) proved the usefulness of using well-structured visual representations (concept maps) in a learning-by-teaching environment. Similarly, visualizations of computer program structure and behavior could help with source code comprehension (Hendrix, Cross II, Maghsoodloo, 2002).

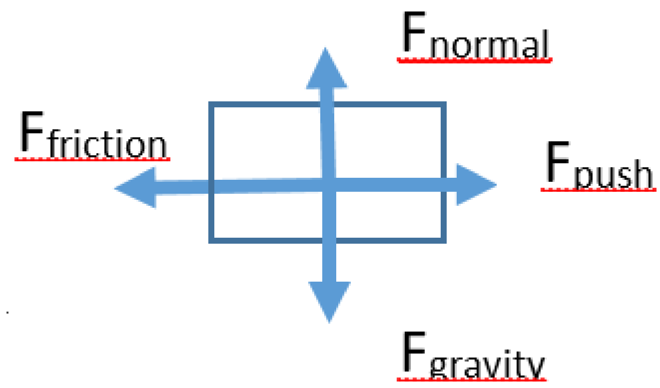


Figure 3. An example of a free-body diagram used in Physics.

Using visual expressions of one's situation model can also be useful for assessing the accuracy of such models. For instance, Chi and colleagues (1994) analyzed mental models from verbal input and drawings. Other research has assessed mental models by asking learners to draw and explain diagrams or visual images that demonstrate an overall function or system (Butcher, 2006; Gadgil et al., 2012).

Control Structure Diagrams (CSDs; Hendrix, Cross II, & Maghsoodloo; 2002) are graphical representations that capture the control structure and modular organization of a computer program. CSDs have the advantage of acting as a companion to source code because CSDs elements are attached to chunks of source code (see Figure 4) as opposed to being a separate representation, which is the case for flowcharts. Hendrix, Cross II, and Maghsoodloo (2002) showed that CSDs are more helpful than other visual representations, e.g., flowcharts, for source code reading and comprehension. The companionship aspect of CSDs with respect to a source code could play a role similar to the use of subgoal labels which have been shown to reduce cognitive load and increase performance while students learn programming (Margulieux, Guzdial, & Catrambone, 2012).

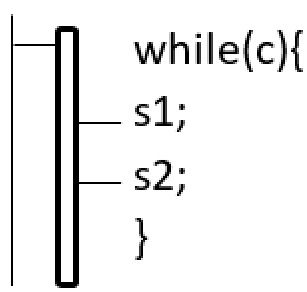


Figure 4. Example of a control structure diagram used in visualizing the organization of computer code.

In sum, self-generated graphical representations of mental models can be used both as an instructional strategy to help one's understanding (Hendrix, Cross II, and Maghsoodloo, 2002; Mualem & Eylon, 2010), for assessing the quality of mental models (Chi et al., 1994; Butcher, 2006; Gadgil et al., 2012), and for expressing a team member's understanding of a larger task that the team must tackle and which can then be used as a starting point to generate a team mental model that encompasses all members' understanding (Ellis, Wahid, Danis, & Kellogg, 2007).

Shared Mental Models for Adaptive Team Training

Cannon-Bowers and Bowers (2011) reason that the most stressful demands on individuals in work/operational environments arise from their participation as a member of a team. These performance demands may be primarily due to the development of teamwork skills that team members must acquire for the team to perform optimally and the emergence of divergent goals within the team, but they may also be due to the complex, and dynamic nature of teamwork and the constant need to adapt to emergent teamwork processes and phenomena (Kozlowski and Ilgen, 2006). Grand et al. (2016) described team dynamics as the modeling of team cognition and shared knowledge of team tasks that underlie team development and thereby affect team learning and performance. To overcome these barriers to learning, the successful team is able to understand goals and construct shared mental models of the processes for teamwork and taskwork.

Shared mental models include organized common knowledge about a system (e.g., instructional domain) that enables individual team members to understand its basic processes and then form predictions and expectations about its future states (Rouse & Morris, 1986). If Intelligent Tutoring Systems (ITSs) are to be effective tools for adaptive team training, Fletcher & Sottilare (2017) advocate a close-coupling of ITS instructional strategies, shared mental models, and teamwork. Sottilare et al. (2017) reinforce this with their findings on the importance of teamwork behaviors, attitudes, and cognition.

In any domain, the successful adaptive team tutor optimizes instruction by adjusting the presentation of content (e.g., text, graphics, and active media like serious games) to maintain team member engagement and thereby optimize learning opportunities within the team. Collaborative graphical activities within adaptive team instruction support the development of or reinforce shared mental models by providing a mechanism leading to better common understanding of the domain, the team, and their learning. These activities may reduce stress and cognitive workload by synchronizing goals and activities, and thereby increase team learning and performance.

Kay, Yacef & Reimann (2007) observed that learners, especially those in leadership roles, found visualizations useful and that a significant number of learners modified their behaviors based on the visualizations provided. Visualizing shared-knowledge awareness, the perception of shared knowledge learners have while working in a collaborative learning context, can also enhance group learning (Collazos, Guerrero, Redondo & Bravo, 2011).

The most effective type of media to support the development of shared mental models and improve team performance may depend on the type of activity (e.g., team taskwork, collaborative learning, or collaborative problem solving) in which the team is engaged. During taskwork activities, graphics that provide the status of team performance or achievement (e.g., dashboards) can level understanding within the team leading to better performance. Widgets within dashboards that represent leaderboards, activity streams, and concept maps are common and useful. During collaborative learning activities, graphics that model concepts or processes (e.g., free body diagrams) can reinforce individual learning or identify shortfalls in knowledge, misconceptions or diverging objectives of team members. In attempting to collaboratively solve problems, graphics that visualize data or allow learners to share and vet information are also valuable.

Conclusion

The graphical supports described in this chapter (concept maps, 3-dimensional cognitive mapping, and self-visualizations) have proven to be effective vehicles for the collaborative construction of mental models. Each fulfills the necessary condition of organized common knowledge about a system (e.g., instructional domain) that enables individual team members to understand its basic processes and then form predictions and expectations about its future states). As advances in technology enable increasingly sophisticated types of visual representations, such as virtual and mixed realities (Liu, Dede, Huang, & Richards, 2017), insights from the graphical supports described here will aid in designing effective vehicles for collaboration.

In GIFT, the Domain Course file and the Domain Knowledge file are good components in which to implement graphical/visual representations for shared mental models. Digital tools for creating individual and group graphical supports have the advantage of providing a logfile record of the steps involved in producing these mental models, so that both instructors and students can review the processes of creation by individuals and synthesis/collaboration by the group. This is not only valuable in producing a shared mental model, but also in elucidating strengths and weaknesses of the collaborative actions involved.

References

Biswas, G., Leelawong, K., Schwartz, D., Vye, N. & The Teachable Agents Group at Vanderbilt (2007). Learning by teaching: a new agent paradigm for educational software. Applied Artificial Intelligence, 19:3-4, 363-392, DOI:10.1080/08839510590910200.

Butcher, K. R. (2006). Learning from text with diagrams: promoting mental model development and inference generation. Journal of Educational Psychology, 98(1), 182-197.

Cannon-Bowers, J. A., & Bowers, C. (2011). Team development and functioning. In S. Zedeck (Ed.), APA handbook of industrial and organizational psychology, Vol 1: Building and developing the organization (pp. 597–650). Washington, DC: American Psychological Association.

Chang, K. E., Sung, Y. T., & Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid. Journal of Computer Assisted Learning, 17(1), 21–33. https://doi.org/10.1111/j.1365-2729.2001.00156.x

Chen, J., Wang, M., Dede, C., & Grotzer, T. A. (2017). Design of a Three-Dimensional Cognitive Mapping Approach to Support Inquiry Learning. Educational Technology & Society, 20(4), 191-204.

Chi, M. T. H., de Leeuw, N., Chiu, M. H. and La Vancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18, 439-477.

Collazos, C., Guerrero, L., Redondo, M., & Bravo, C. (2007). Visualizing shared-knowledge awareness in collaborative learning processes. Groupware: Design, Implementation, and Use, 56-71.

Cox, R. (1999). Representation construction, externalised cognition and individual differences. Learning and instruction, 9(4), 343-363.

Dede, C., Grotzer, T., Kamarainen, A., & Metcalf, S. (2017). EcoXPT: designing for deeper learning through experimentation in an immersive virtual ecosystem. Educational Technology & Society, 20(4), 166–178.

Ellis, J.B., Wahid, S., Danis, C., Kellogg, W.A. (2007). Task and social visualization in software development: evaluation of a prototype, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, April 28-May 03, 2007, San Jose, California, USA

Fletcher, J.D. & Sottilare, R.A. (2017). Shared mental models in support of adaptive instruction for team tasks using the GIFT tutoring architecture. International Journal of Artificial Intelligence in Education. DOI: 10.1007/s40593-017-0147-y.

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. Learning and Instruction, 22(1), 47-61.

Grand, J. A., Braun, M. T., Kuljanin, G., Kozlowski, S. W., & Chao, G. T. (2016). The dynamics of team cognition: a process-oriented theory of knowledge emergence in teams. Journal of Applied Psychology, 101(10), 1353.

Grotzer, T.A., Metcalf, S.J., Tutwiler, M.S., Kamarainen, A.M., Thompson, M. & Dede, C. (2017, April). Teaching the systems aspects of epistemologically authentic experimentation in ecosystems through immersive virtual worlds. National Association for Research in Science Teaching (NARST). San Antonio, TX.

Grotzer, T.A., Kamarainen, A., Tutwiler, M.S, Metcalf, S, & Dede, C. (2013). Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. BioScience, 63(4), 288-296.

Hendrix, T. D., Cross, J. H., & Maghsoodloo, S. (2002). The Effectiveness of Control Structure Diagrams in Source Code Comprehension Activities, IEEE Transactions on Software Engineering, Vol. 28 No. 5, pp. 1-15.

Jonassen, D. (2003). Using cognitive tools to represent problems. Journal of Research on Technology in Education, 35(3), 362-381.

Kay, J., Yacef, K., & Reimann, P. (2007, July). Visualisations for team learning: small teams working on long-term projects. In Proceedings of the 8th international conference on Computer supported collaborative learning (pp. 354-356). International Society of the Learning Sciences.

Kozlowski, S. W. J., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. Psychological Science in the Public Interest, 7(3), 77–124.

Larkin. J. & Simon, H, (1987). Why a diagram is (sometimes) worth 10.000 words. Cognitive Science, 11, 65-99.

Liu, D., Dede, C., Huang, R., & Richards, J. (Eds.). (2017). Virtual reality, augmented reality, and mixed reality in education. Hong Kong: Springer.

Margulieux, L. E., Guzdial, M., & Catrambone, R. (2012). Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In Proceedings of the Ninth Annual International Conference on International Computing Education Research, 71-78.

Metcalf, S.J., Kamarainen, A., Tutwiler M.S., Grotzer, T.A. & Dede, C. J. (2011). Ecosystem science learning via multi-user virtual environments. International Journal of Gaming and Computer-Mediated Simulations, 3(1), 86-90.

Mualem, R., & Eylon, B.S. (2010). Junior high school physics: Using a qualitative strategy for successful problem solving. Journal of Research in Science Teaching, 47(9):1094-1115.

Muhling, A. (2017). Concept landscapes: aggregating concept maps for analysis. Journal of Educational Data Mining, 9(2), 1-30.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: a meta-analysis. Review of Educational Research, 76(3), 413–448.

Novak, J. D. (1990). Concept mapping: A useful tool for science education. Journal of research in science teaching, 27(10), 937-949.

Novak, J. D., & Canas, A. J. (2006). The Theory Underlying Concept Maps and How to Construct Them. Institute for Human and Machine Cognition.

Olney, A. M., & Cade, W. L. (2015). Authoring intelligent tutoring systems using human computation: designing for intrinsic motivation. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), Foundations of Augmented Cognition (Vol. 9183, pp. 628–639). Springer International Publishing. https://doi.org/10.1007/978-3-319-20816-9_60

Rice, D., Ryan, J., & Samson, S. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? Journal of Research in Science Testing, 35(10), 1103-1127.

Rosen, Y., & Tager, M. (2014). Making student thinking visible through a concept map in computer-based assessment of critical thinking. Journal of Educational Computing Research, 50(2), 249-270.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: prospects and limits in the search for mental models. Psychological Bulletin, 100, 349–363.

Ruiz-Primo, M. A., Schultz, S. E., Li, M., & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques*. Journal of Research in Science Teaching, 38(2), 260–278. https://doi.org/10.1002/1098-2736(200102)38:2<260::AID-TEA1005>3.0.CO;2-F

Schvaneveldt, R. W. (Ed.). (1990). Pathfinder associative networks: studies in knowledge organization. Ablex Publishing Corp. Retrieved from http://portal.acm.org/citation.cfm?id=119801

Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H. & Gilbert, S.B. (2017). Towards a Design Process for Adaptive Instruction of Teams: A Meta-Analysis. International Journal of Artificial Intelligence in Education. DOI: 10.1007/s40593-017-0146-z.

Toth, E. E., Suthers, D. D., & Lesgold, A. M. (2002). “Mapping to know”: The effects of representational guidance and reflective assessment on scientific inquiry. Science Education, 86(2), 264-286.

Wang, S., Walker, E., Chaudhry, R., & Wylie, R. (2015). Personalized expert skeleton scaffolding in concept map construction. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June 22-26, 2015. Proceedings (pp. 808–811). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-19773-9_121.

Wang, M., Wu, B., Kinshuk, Chen, N. S., & Spector, J. M. (2013). Connecting problem solving and knowledge-construction processes in a visualization-based learning environment. Computers & Education, 68, 293–306.

Wu, B., Wang, M., Grotzer, T. A., Liu, J., & Johnson, J. M. (2016). Visualizing complex processes using a cognitive-mapping tool to support the learning of clinical reasoning. BMC Medical Education, 16(1), 216.

Chapter 4 – Modeling Processes of Enculturation in Team Training

A. R. Ruis[1], Andrew J. Hampton[2], Benjamin S. Goldberg[3], David Williamson Shaffer[1,4]
University of Wisconsin–Madison[1], University of Memphis[2], U.S. Army Research Laboratory[3], Aalborg University Copenhagen[4]

INTRODUCTION

Both military and civilian work is increasingly organized around small, dynamic teams rather than large bureaucratic frameworks. Such teams combine individuals with highly specific skill sets and extensive training to solve complex, non-standard problems, often under extreme pressure. Importantly, these teams are not typically composed of interchangeable members but are formed and trained *as expert teams* to function semi-autonomously. To accomplish mission objectives in the face of complex challenges, the United States military needs to develop teams that consistently exhibit high levels of taskwork and teamwork skills. Implementing principles of team training and maintaining team efficacy are critical in the armed forces, where teams are often widely dispersed and consequences for underperformance can be severe, though many civilian teams—such as hospital trauma teams or flight crews—face similar challenges. Establishing and maintaining high levels of team performance in these contexts requires the creation of practical and effective team development interventions, including team training, as well as systems for ongoing assessment of team function.

We argue that one critical component of training, monitoring, and maintaining high-functioning teams is the ability to *model team performance*.

Communication, cognition, coordination, collaboration, and coherence in teams are critical for predicting team performance. To improve our ability to enhance and maintain team performance, we need to develop a better understanding of these components. Specifically, we need to understand how the components of team dynamics influence team performance in complex problem-solving situations (Fiore et al., 2010; O'Neil, Chuang, & Baker, 2010; Paris, Salas, & Cannon-Bowers, 2000; Salas, Cooke, & Rosen, 2008). However, current tools and methods lack the capacity to assess these components of teamwork, to the dismay of stakeholders. This is true in the armed forces, industry, government, and civic organizations, and it has motivated national and international assessments of teamwork and collaborative problem solving as a 21st-century skill.

In what follows, we outline a program of research on the science of teamwork, based on a theoretical framework for analyzing the decision-making processes and effectiveness of teams. Drawing on our prior work developing critical constructs and mechanisms to measure team performance using multilevel network analysis, we argue that one critical advance in the science of teamwork is using these tools to build predictive models that consider teams as complex systems. To model team performance effectively, we need to understand teams as multilevel networks comprised of three main components: (a) the *social network* that structures team interactions; (b) the *conceptual networks* that guide the actions of individuals on the team; and (c) the *communication network* by which that action is accomplished. A critical step in creating a system to monitor and support team performance, then, is the development of multilevel network analysis techniques for assessing teamwork during complex problem solving.

HIGH-PERFORMANCE TEAMS IN MILITARY AND CIVILIAN WORK

Two factors have made analysis of the performance dynamics of small teams critically important to current and future U.S. military endeavors. First, the nature of armed conflict has shifted in the last 50 years from large-scale machine warfare organized around regiments of conventional forces, in which combat and intelligence gathering present relatively standard problems (*materiel domain*), to small-scale asymmetric warfare organized around small teams of special forces, in which counter-insurgency and other elements of unconventional combat present non-standard, highly complex problems (*human domain*). Second, budget cuts, the current geopolitical landscape, and changes in military priorities have led to reorganization of operations around small, dynamic teams engaged in counter-terrorism, cyberwarfare, drone operations, foreign internal defense, peacekeeping, and humanitarian aid (Knoke, 2013; Odierno, Amos, & McRaven, 2013; Ressler, 2006; Stewart, 2013; Thomas & Dougherty, 2013; Tucker & Lamb, 2007; Turnley, 2011).

Today's small military teams, such as four-member Navy Special Warfare squads or twelve-member Army Operational Detatchment-A teams (Feickert, 2010; McRaven, 1996), combine individuals with highly specific skill sets and extensive training to solve complex, non-standard problems under extreme pressure. Importantly, teams such as these are not composed of interchangeable soldiers embedded in rigid bureaucratic frameworks—as in the squads of conventional military organization. Rather, these teams are formed and trained *as expert teams* to function semi-autonomously (Thomas & Dougherty, 2013).

Of course, this approach is not limited to military organizations. Civilian work, too, is increasingly structured around small, high-performance teams (Buchholz, Roth, & Hess, 1987; Katzenbach, 1993; Lehman & DuFrene, 2010). Hospital trauma teams, emergency response teams, flight crews, design teams, and research teams, among others, are core organizational units in many contexts. While civilian teams may not face the same pressures as military teams, performance is often equally dependent on the extent to which the members function *as a team*.

TEAMS AS MULTILEVEL NETWORKS

A central challenge in understanding team performance is to integrate understanding of *how* a team collaborates with information about *what* they are collaborating on. For example, high-functioning teams "communicate clearly", but in order to assess whether a team is performing well, we need to know more than just *that* they are communicating well. We also need to know that they are communicating effectively about *particular aspects of the specific problem* they are working on, and that their approach to the problem is *appropriate for the specific circumstances* in which they are working.

That is, to measure team performance, we need a technique that can measure critical aspects of team performance, including those articulated in the PISA 2015 Collaborative Problem Solving Framework (Greiff, 2012; Organization for Economic Co-operation and Development, 2013) and various frameworks of 21st-century skills (Griffin, 2012; Koenig, 2011; Kozma, 2009; Trilling & Fadel, 2012). These aspects include (a) how well a team collaborates in terms of social and cognitive alignment; (b) how well a team functions in a problem-solving context, including alignment with organizational factors and team outcomes; (c) how well each individual contributes to creating cognitive, social, and organizational alignment and team outcomes; and (d) the relationships among and integration of these factors.

Put another way, we need to understand, simultaneously, the social network of the team, the conceptual networks that guide the actions of the individuals on the team, and the communication network by which those actions are accomplished. Thus, we argue that a critical step in creating a system to monitor and support team performance is the development of network analysis techniques for assessing teamwork during complex problem solving. Specifically, we suggest that the science of teamwork needs research into methods for constructing *multilevel network models* of high-volume *discourse* data (Frank, 1998, 2011; Penuel, Riel, Krause, & Frank, 2009; Wang, Robins, Pattison, & Lazega, 2013). By "discourse", we refer to spoken interactions, but also more broadly to any actions or interactions of team members and others in the problem-solving context (Gee, 1999). Our goal is to develop network analysis techniques that *account simultaneously for the cognitive, social, and communications networks* that comprise team activity and within which a team functions.

NETWORK MODELS OF TEAMWORK

Our approach is to start with *epistemic network analysis* (ENA), a technique that we have developed to analyze records of individual and team problem solving (Shaffer, 2017; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017). A fundamental claim in this work is that *it is essential to consider the semantic and conceptual content of what gets said during social interactions in addition to tracing the patterns of who talks to whom in a social network*. Social network models devoid of content are doomed to fail because team interactions are never "content neutral" (Maroulis & Gomez, 2008). It is impossible to evaluate the quality of team interactions by examining who is talking to whom without knowing what they are talking about. This work is thus unique in combining deep analyses of both content and social network processes.

Specifically, ENA models team activity by identifying categories of action, communication, cognition, and other relevant features and characterizing them with appropriate coding schemes into smaller sets of domain-relevant nodes. The weights of the connections among network nodes (i.e., the association structure of key elements in the domain) are then computed and visualized. Critically, ENA models team actions and interactions in such a way that it is possible to *extract information about each team member's contributions to team performance*. ENA uses statistical and visualization techniques to enable comparison of the salient properties of different networks, including networks generated by different teams or by teams at different points in time, teams in different spatial locations, or teams engaged in different activities. These salient properties are modeled not just in terms of the general structure of the networks, as is often revealed by other network analysis techniques (changes in density or betweenness centrality, for example), but ENA also extracts properties relevant to the actual content of the network.

In other words, ENA can analyze what teams are doing, how they are thinking about what they are doing, what role individuals are playing in team performance, and how teams compare to one another in the context of real problem solving. Using ENA, we have been able to identify critical patterns of interaction in expert and novice teams, as well as successful and unsuccessful teams and individuals (Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015; Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Chesler et al., 2015; Quardokus Fisher, Hirshfield, Siebert-Evenstone, Arastoopour, & Koretsky, 2016; Shaffer, 2017; Sullivan et al., in press).

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

A critical next step in developing a science of teamwork is to extend ENA from modeling team interactions in a single modality (communication among team members) to account for the different layers of activity that influence a team's work. Factors we believe can and should be modeled include: (a) cognitive and psychological factors of team members and of the problem being solved; (b) modes of communication, including synchronous and asynchronous interactions; (c) the organization and structure of the team, as well as the influence of role and hierarchy within the team; (d) social relationships and interactions among team members; and (e) organizational and other external influences on the team's activities.

To accomplish this, we hypothesize that we will need to make three significant advances in the network science of teams. First, we will need to extend our existing natural language processing coding algorithms for text dialog. That is, we will need to develop machine learning and other techniques to quickly and reliably develop, calibrate, and implement coding schemes for a wide array of discourse data, including both text and spoken dialogue. Building on existing computational linguistics technologies, such as Coh-Metrix (Graesser et al., 2014; Graesser, McNamara, & Kulikowich, 2011) and LIWC (Pennebaker, Booth, & Francis, 2007), we will augment the existing natural language processing capabilities of ENA to develop a version of the tool that can code diverse data streams for multilevel analyses.

Second, we will need to examine how we can most appropriately model the social, cognitive, and communicative processes by which connections in those networks are constructed. That is, we will need to develop a more robust scientific understanding of how to identify and model the links between ideas and between people that individuals make during team problem solving (Dillenbourg, 1999).

Third, we will need to develop multilevel ENA (mENA) network models. We conceptualize these mENA models as the network science analog of hierarchical linear modeling, where the effects of each layer of the model are analyzed, but critically, those analyses account for the interactions between the different layers of the model. For example, prior work has looked at integrating hierarchical linear modeling and social network analysis to examine how social factors influence students' school achievement (Frank, 1998). Similarly, mENA would be able to model the network of cognitive relations for each member of the team, but also account for the nesting of these individual cognitive models in the team setting. mENA would also be able to model the impact of individuals' cognitive and affective states on social interactions and relations among team members.

As part of this work, we will also explore ways to integrate mENA into systems for team training and development, such as training modules developed with GIFT, the Generalized Intelligent Framework for Tutoring (Sottilare, Brawner, Goldberg, & Holden, 2012). For example, mENA could be used to model expert teamwork and behavior based on observations gathered across an array of high-performing teams. In these instances, aggregating relevant mENA features across problem-solving scenarios can produce rich network models that will organize the actions, communications, and contextualized decision points that need to be explicitly defined within GIFT's domain module. Designating relationships between information, communication, and action across roles within a team provides rich inputs for structuring and configuring an associated Domain Knowledge File that manages assessment and pedagogical requests during a GIFT-managed scenario event. Through these techniques, one can determine whether high-performing teams execute tasks in similar ways (i.e., they have similar mENA networks), which could warrant more general claims about team performance, or whether there are unique differences across different teams or different teamwork scenarios. The same techniques can be applied to novice and low-performing teams to identify common challenges.

Following the development of an mENA expert model, mENA could be used to assess specific team activities in comparison with the expert model. This will enable functional evaluations of teams that are in training against representations of ideal behavior; as discussed above, mENA is particularly well suited to make such comparisons, both statistically and visually. Differences in mENA models that reflect critical cognitive, communicative, or enactive behaviors could thus assist in establishing granular assessment methods that can inform coaching decisions. With respect to GIFT, this involves making strategy selections that associate with directed feedback delivered in real-time, scenario adaptations that focus on adjustments in difficulty and complexity, and post-event scenario selections to target specified skill sets that require additional training or practice.

Ultimately, our goal is to produce a system for training and maintaining high-performance teams that (a) enables easy creation of training modules that (b) provide teams with realistic simulations of problem-solving scenarios and (c) generate mENA models that give team members and coaches actionable feedback.

ACKNOWLEDGEMENTS

This work was funded in part by the National Science Foundation (DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DRL-1661036, DRL-1713110, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, REC-0347000), the MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

REFERENCES

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology*, *6*(1016).

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501.

Buchholz, S., Roth, T., & Hess, K. M. (1987). *Creating the high performance team*. Wiley.

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, *137*(2), 024701:1-8.

Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches*. New York, NY: Elsevier.

Feickert, A. (2010). *US Special Operations Forces (SOF): Background and issues for Congress*. Washington, D.C.: Congressional Research Service.

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K., Salas, E., Letsky, M., & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting processes in complex collaborative contexts. *Human Factors*, *53*, 203–224.

Frank, K. A. (1998). Quantitative methods for studying social context in multilevels and through interpersonal relations. *Review of Research in Education*, *23*, 171–216.

Frank, K. A. (2011). Social network models for natural resource use and extraction. In Ö. Bodin & C. Prell (Eds.), *Social networks and natural resource management: Uncovering the social fabric of environmental governance* (pp. 180–205). Cambridge, UK: Cambridge University Press.

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. London, UK: Routledge.

Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, In press.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234.

Greiff, S. (2012). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, *19*(2), 111–121.

Griffin, P. (2012). *Assessment and teaching of 21st century skills*. Springer. Retrieved from http://books.google.com/books?id=mcH47qFrJasC&lpg=PP1&pg=PP1#v=onepage&q&f=false

Katzenbach, J. R. (1993). *The wisdom of teams: Creating the high-performance organization*. Harvard Business School Press.

Knoke, D. (2013). "It takes a network": The rise and fall of social network analysis in US Army counterinsurgency doctrine. *Connections*, *33*(1), 1–10.

Koenig, J. A. (2011). *Assessing 21st century skills: Summary of a workshop*. Washington, D.C.: National Research Council.

Kozma, R. (2009). Transforming education: Assessing and teaching 21st century skills. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (Vol. 13–23, pp. 13–23). Luxembourg: European Commission.

Lehman, C., & DuFrene, D. (2010). *Building high-performance teams*. Cengage Learning.

Maroulis, S., & Gomez, L. M. (2008). Does" connectedness" matter? Evidence from a social network analysis within a small-school reform. *Teachers College Record*, *110*(9), 1901–1929.

McRaven, W. H. (1996). *Spec Ops: Case studies in special operations warfare: Theory and practice*. New York, NY: Presidio Press.

Odierno, R. T., Amos, J. F., & McRaven, W. H. (2013). *Strategic landpower: Winning the clash of wills*. US Army, US Marine Corps, and US Special Operations Command.

O'Neil, H. F., Chuang, S. S., & Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem solving. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 261–279). New York, NY: Springer.

Organization for Economic Co-operation and Development. (2013). *PISA 2015: Draft collaborative problem solving framework*.

Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: A review and analysis. *Ergonomics*, *43*(8), 1052–1075.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). LIWC2007: Linguistic inquiry and word count. Austin, TX: LIWC.net.

Penuel, W. R., Riel, M., Krause, A., & Frank, K. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record*, *111*(1).

Quardokus Fisher, K., Hirshfield, L., Siebert-Evenstone, A. L., Arastoopour, G., & Koretsky, M. (2016). Network analysis of interactions between students and an instructor during design meetings. In *Proceedings of the American Society for Engineering Education* (p. 17035). ASEE.

Ressler, S. (2006). Social network analysis as an approach to combat terrorism: Past, present, and future research. *Homeland Security Affairs*, *2*(2), retrieved-from http://www.hsaj.org/?article=2.2.8.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*(3), 540–547.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9–45.

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). *The generalized intelligent framework for tutoring (GIFT)*. Orlando, FL: U.S. Army Research Laboratory.

Stewart, R. (2013). *Human domain operations: Institutionalising eating soup*. Retrieved from http://smallwarsjournal.com/print/14814.

Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., … Jung, H. S. (in press). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*.

Thomas, J., & Dougherty, C. (2013). *Beyond the ramparts: The future of US Special Operations Forces*. Washington, D.C.: Center for Strategic and Budgetary Assessments.

Trilling, B., & Fadel, C. (2012). *21st century skills: Learning for life in our times*. San Francisco, CA: Jossey-Bass.

Tucker, D., & Lamb, C. J. (2007). *United States special operations forces*. New York, NY: Columbia University Press.

Turnley, J. G. (2011). *Cross-cultural competence and small groups: Why SOF are the way SOF are*. MacDill AFB, FL: JSOU Press.

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, *35*(1), 96–115.

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

Sottilare, R., Goldberg, B., Brawner, K., & Holden, H. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.

Sottilare, R., Holden, H., Goldberg, B., & Brawner, K. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottilare, R. (Eds.) *Fundamental Issues in Defence Simulation & Training*. Ashgate Publishing.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., et al., (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*, 15(3), 147–204.

VanLehn, K. (2011): The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, Educational Psychologist, 46:4, 197-221.

Vygotsky, L.S. (1978). Mind in Society: The development of higher psychology processes. Cambridge MA: Harvard University Press.

Chapter 5 – Bayesian Model of Team Training and Measures

Alan Carlin[1], Vasile Rus[2], Ben Nye[3]
Aptima, Inc.[1], University of Memphis[2], University of Southern California Institute for Creative Technologies[3]

Introduction

A large amount of research has been conducted recently on adaptive training and personalized learning, and from the modeling perspective, conferences such as AIED (Artificial Intelligence in Education), EDM (International Conference on Educational Data Mining), and LAK (Learning Analytics & Knowledge Conference), have driven advances in the state-of-the-art. However, extension of research models from individual training to team training presents a new set of challenges. Different team members may have different training requirements, and in some cases team members may have different training needs; a training exercise that is best for one team member may either be boring to another team member, or too advanced. Furthermore, individual training measures may apply to different team members, to different extents. In this paper we address the following questions: 1) How can one select a training exercise that best trains all of the individual team members?; 2) How can the team training measures be related back to individual assessments, and thus to selection of future team training exercises?

We propose a generative model to reason about instructor decisions in the adaptive team training process. The model includes: 1) A representation of individual and team training objectives, 2) A representation of the Knowledges, Skills, and Experiences (KSE's) required to fulfill the training objectives, 3) A representation of individual and team exercises that a human or computer-based instructor can assign to team member(s), 4) For each exercise, a representation of which KSE's are trained by the exercise, and 5) For each exercise, a representation of its outcome measures and how each measure relates to KSE's.

The model can be used to produce optimal adaptive sequences of individual and team exercise selection. Furthermore, it can be parameterized based on performance data obtained from a Learning Records Store (LRS) or a Learning Management System (LMS). In this paper, we formalize the model, and we also describe how it can be integrated into frameworks such as the Generalized Intelligent Framework for Tutoring (GIFT).

Individual Training Models

The approach described in this paper formalizes the team training process as a Partially Observable Markov Decision Process (POMDP; Smallwood and Sondik, 1973). A POMDP is related to the more widely used and simpler Hidden Markov Model (HMM), such as that used in Bayesian Knowledge Tracing (BKT; Anderson 1995). The team training variant we describe in this chapter (POMDP-TT) combines Bayesian Knowledge Tracing, Item Response Theory, and previous individual-trainee versions of a POMDP model.

Bayesian Knowledge Tracing

A Hidden Markov Model contains *States*, *Transitions*, and *Observations*. In the BKT formulation individual components of knowledge are tracked as either learned or unlearned, and a variable called state, which we designate S_n, where n identifies the skill being learned. Table 1 summarizes typical elements of BKT, with a renaming/updating of variables and descriptions in Table 1 to conform to the notation in the rest of this chapter.

Table 1: Bayesian Knowledge Tracing Definitions

$p(S_i)$	The probability that KSE i is in the learned state prior to the first opportunity to apply it
$p(T)$	The probability a KSE will make a transition from the unlearned to the learned state
$p(G)$	The probability a student will guess correctly if a KSE is in the unlearned state
$p(Slip)$	The probability the student will make a mistake if the KSE is in a learned state

The probability $p(T)$ can be used to form a probability transition table. For example a value of $p(T)$ generates the table below, where the probability of transitioning from an unlearned to a learned state is 70% if the student is presented with an item that uses the KSE. In the table there is zero probability of skill decay, that the skill is unlearned once learned.

Table 2: Transition Table in Hidden Markov Model for Bayesian Knowledge Tracing

	Unlearned	Learned
Unlearned	.3	.7
Learned	0	1

The probability $p(G)$ and $p(Slip)$ represent the probability of guessing the item correct if the skill is unlearned (i.e. a “lucky guess”) and making a mistake despite the skill being learned, respectively. Together they compose the observation table of the Hidden Markov Model. Table 3 shows an example of the observation table for the case where $P(G) = .1$ and $P(Slip) = .2$

Table 3: Observation Table in Hidden Markov Model for Bayesian Knowledge Tracing

	Correct	Incorrect
Unlearned	.1	.9
Learned	.8	.2

In the above formulation, there is a separate model for each skill that contains its own $p(T)$, $p(G)$, and $p(Slip)$. Typical applications of a BKT approach are to (1) Learn these model parameters, given data, and

(2) assess individual student progress, given the model parameters and that individual student's training history.

Item Response Theory

Whereas Bayesian Knowledge Tracing focuses on the evolution of a skill over time, Item Response Theory (IRT) is a family of models that focuses on item performance (Lord, 1980), in that skill level can be related to task difficulty, each can support many or continuous skill levels, and the values in the table can be determined from data if applicable. A simple variant models item performance as:

$$\boldsymbol{p(correct) \propto c_i + \frac{1 - c_i}{1 + e^{-a_i\,(s_i - d_i)}}}$$

Where $\boldsymbol{s_i}$ is a parameter representing trainee ability, $\boldsymbol{d_i}$ is a number representing item difficulty, $\boldsymbol{a_i}$is a discrimination parameter, and $\boldsymbol{c_i}$represents chance. With IRT, these parameters can be learned for multiple trainees performing on multiple items. By adding more variables to the IRT model the shape of the Item Characteristic Curve (ICC) can be varied. Typical IRT parameters are item difficulty, item discrimination parameter (governing how rapidly an increase/decrease in student skill level results in a change of the probability of correctness), and probability of getting items correct by guessing. The IRT model can also be extended to a Partial Credit Model (PCM) that generates ICC's for gradations of correctness of response, and it can also be extended to account for multiple different skills.

POMDP

A POMDP (Partially Observable Markov Decision Process) represents an extension from an HMM in that it is a *Decision Process*. In Levchuk et al., the authors introduced multiple *training actions* that were each able to train team skill (Levchuk, 2007, 2012). Each training action is associated with a different $p(T)$. The team was modeled as being in a high-skill, medium-skill, or low-skill state. The high-skill state was associated with a reward. Unlike the BKT approach above, in this work the model parameters were supplied by Subject Matter Experts. The problem addressed by the POMDP was action selection, to select the action that maximizes reward over time. Formally, the components of a POMDP model are:

- S: a finite set of states
- A: a finite set of control actions
- Z: a finite set of observations
- $\tau(S \times A \times S)$: a state transition function
- $O(Z \times S \times A)$: an observation function for each action.
- $R(S \times A)$: a reward function for each state and action.
    - γ: A discount factor over future time steps.
- $b_0(S)$: An initial distribution that assigns a probability to each state, referred to as a belief state, at time zero
- γ: A discount factor over future time steps.

As in an HMM, a POMDP updates its assessment of student belief state after each action.

$$Pr\,(s^1 |\, s^0, a^0) \propto Pr\,(s^0)\,\tau(s^0, a^0, s^1)$$

Time unfolds over a series of discrete time steps, $t = 1,2, .. \infty$. s^t and a^t represent the trainee state and training action taken at time step t. As opposed to an HMM, in which only one action is considered per model, a single POMDP model allows for explicit comparison between actions within the same model. Thus, a POMDP model can be used for the problem of apples-to-apples comparisons for action selection. This involves assigning each action a value for the current belief state, and selecting the action with the highest value. Value differs from Reward in that the reward function R is by definition for a single step, and the value function (usually denoted V, $V(s)$ or (s, a)) is defined for multiple steps:

$$V = \sum_{t=0}^{\infty} R(s^t, a^t)\gamma^t$$

The value of a belief state can be computed as an expectation over its component states.

$$V(b, a) = \sum_{s} b(s)V(s, a)$$

Where $b(s)$ is a function that maps a state s to a probability.

The value of an action for a state is initialized to be the direct reward.

$$V(s, a) := R(s, a) \text{ and } V(s) = max_a \,(V(s, a))$$

But the above yields only the immediate reward. Reward over more than one time step can be recomputed using a Bellman backup.

$$V(b, a) = \sum_{s'} \sum_{s} b(s)\tau(s, a, s')O(o|a, s'))V(s', \pi_{\{b,a,o\}})$$

Where $\pi_{\{b,a,o\}}$ is the action selected after taking action a and receiving observation o. That is, the value of a belief state is decomposed into the product of the probability of each state, the probability of transitioning from that state to a new state s', the probability of making an observation from that new state, and the value of continuing the adaptive policy from that new state, given the observation.

A variety of POMDP and MDP representations for training have been constructed (Brunskill, 2012; Folsom-Kovarik, 2012; Rowe, 2015). In this chapter, we build upon modern factored single-trainee models to build a model to reason about team training (Carlin, 2016). We term the resulting model, POMDP-for-Team Training (POMDP-TT).

POMDP-for-Team Training (POMDP-TT)

In this section we define a POMDP-for-Team-Training model (POMDP-TT) based on a theory of deliberate practice (Ericsson, 1993). *Figure 1* shows the construct. At its core, a POMDP-TT contains a standard POMDP model from the literature, that is a tuple $< S, A, \tau, \omega, O, R, \gamma >$. Supporting information about the training domain is used to construct the parameters within this tuple. This includes information about the domain (Team Roles), and information about available training content.

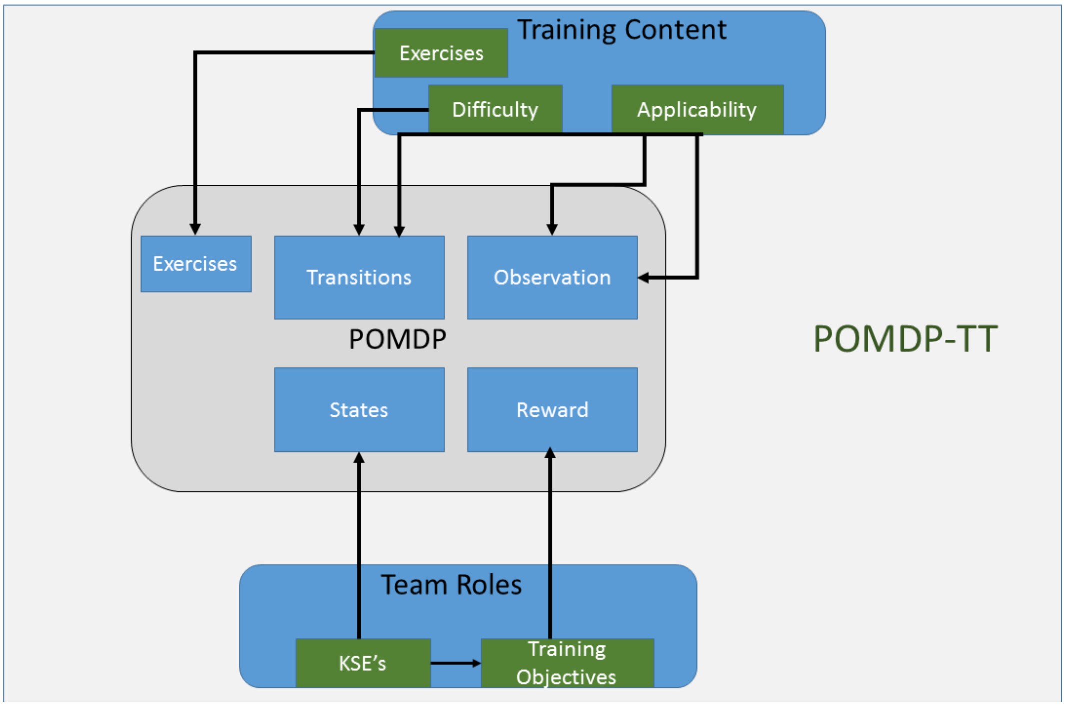


Figure 1: POMDP-TT. At the core is a POMDP model, surrounding information about team training is used to define the POMDP parameters.

POMDP-TT parameters

Information about the POMDP-TT parameters can be obtained by eliciting the information from subject matter experts via a user interface, or alternatively by data mining a Learning Management System for information. Regardless, information elicited should include:

Team Roles:

Information about the positions in the team being trained. Information about the team roles includes the **KSE's** (Knowledges, Skills, and Experiences) relevant to each role, and the **Training Objectives** for the role.

KSE's be elicited from instructors directly via a Scenario Authoring Tool (SAT), or in some domains such as in military domains, there is an already-existing list of training standards. Another method of eliciting KSE's is by correlating item performance, such as data mining a Learning Management System using Principal Component Analysis (Carlin, 2013).

Training Objectives are derived from the KSE's. It is not always the case that the KSE's represent all the training objectives, for instance, some training domains have a complete listing of KSE's for a position, but at initial levels of qualification many of the KSE's may not be relevant, or the KSE's may not need to be trained to an expert level of competence. Therefore Training Objectives contain the list of relevant KSE's for the position being trained, as well as the level of desired performance.

The training planner then needs to only optimize training for the selected skills at the desired level of performance, and no further.

Training Content

Training Content includes information about the training exercises. This includes a list of the exercise names and personnel, which is extracted in a SAT when the exercises are created, or else can be extracted from records in the LMS. Exercises are also tagged for Applicability and Difficulty of each skill. Typically, applicability is specified by tagging the exercise for the KSE's trained. Sometimes, a KSE is featured prominently in the exercise whereas other times, a KSE plays a supporting role. Applicability is a specification of which is the case. In some domains, applicability is specified on a scale (e.g. 1-5), and in others, KSE's are specified as either "primary" or "secondary". Difficulty is the $\boldsymbol{d_i}$ skill in Item Response Theory (see above section).

Applicabilities and difficulties can either be extracted by a SAT, or alternatively a Machine Learning Algorithm such as Markov Chain Monte Carlo can be applied to LMS data to learn these parameters (Carlin, 2016).

(Inner) POMDP parameters

The POMDP-TT parameters, which correspond to the training domain, can be translated into standard POMDP parameters. Below we specify the construction.

States

The state space S continues to be a set of possible team states. However, POMDP-TT describes this set using a factored representation. S is factored into individual components, so that $S = \Pi(S_k)$. Each S_k represents a team member's state in a single Knowledge, Skill, or Experience (KSE). For each $s_k \in S_k$, $s_k \in (0, |max|)$, where s_k represents trainee skill level on that KSE, and max represents the maximum possible skill level. Thus, by the above description member $s \in S$ can be described by a vector $< s_1, s_2, .. s_k >$

Team Members: Under this model, it is possible for many KSE's to apply to a single team member, and/or for many team members to apply to the same team KSE. To specify this relationship more clearly, we can optionally specify:

- The set of team members $I = \{I_1, I_2, .. I_{\{|team|\}}\}$
- A mapping function $\Lambda(S_i) \rightarrow U, U \subset I$ that maps KSE S_i to the team members that $S_i$applies to.
- A related mapping function $\Lambda(I_i)$ that identifies the KSE's related to team member i.

Actions

The set of Actions A continues to be a set, where each $a \in A$ represents training content. Each member $a \in A$ is described as a tuple $(< d_1, .. d_k >, < app_1, .. app_k >)$ where d_i and app_i represent the difficulty and applicability of training content a with respect to KSE i.

Rewards

A reward for each state R(s) is defined by designating a subset of the KSE's as *training objectives*, and a function f that maps progress on the training objectives to a value. Thus, $R(s)$ is defined as $f(s_1 \ldots s_k)$

where each s_i is the student state with respect to a training objective. Note that by this construction, the training objectives may include training an individual team member on multiple skills. Furthermore, rewards allow training objectives to be configurable; reward can be specified for state components at an intermediate level (e.g., $R(<3,3,3>) = R(<5,5,5>)$), or some state components can be ignored in the reward function (e.g., $R(<5,0,5>) = R(<5,5,5>)$), to specify the true training objectives.

Transition Function

Define an individual transition function for each S_i using the applicability and difficulty, and the concept of the Zone of Proximal Development (ZPD; Vygotsky 1978). The transition probability between state s and state s' given action a is based on the following principles:

- The transition probability is proportional to applicability app_i.
- The transition probability is inversely proportional to the difference in skill level between s and s'. (i.e., smaller jumps in skill are more probable than large jumps).
- The transition probability is inversely proportional to the difference in difficulty level of the item and the current student skill level. (enforces ZPD).

An equation that summarizes these three principles is below, where $s_i' > s_i$, d and s_i are always positive, ϵ is a positive constant close to zero, and $w_1$and w_2 are model parameters. (In deployed applications, these have been set to $w_1 = 2$ and $w_2 = s_{\{max\}}$, the maximum possible skill level.)

$$\tau(a =< d_i, app_i >) \propto e^{\frac{-(w_1)(|d-s_i|+1)(s_i'-s_i+1)}{(w_2)(app_i)+\epsilon}}$$

Observations

The set of observations ω refers to the set of measures that are obtained during (or collected after) each training exercise. In Figure 1, $\omega = \{correct, incorrect\}$, but this can be extended into further gradations to account for partial credit, e.g. $\{incorrect, somewhat\ correct, mostly\ correct, correct\}$.

Observation Function

In POMDP-TT, observation probabilities are governed by similar principles to Item Response Theory (IRT; Lord 1980). However to account for multiple skills having different applicabilities, we vectorize a 2-parameter model.

$$p(correct) = \frac{1}{1 + e^{\sum_k \ app_i^t(d_i^t - \theta_j^t)}}$$

This model is easily extended to include further parameters, or into a Partial Credit Model (Masters, 1982).

Application

When student performance is known (e.g. when $p(correct)$ is known to be 0 or 1), and furthermore when item difficulty is known, the observation and transition functions can be used to derive a probability distribution for student skill level θ. This distribution is compared to the training objectives specified in the reward function, which in turn can specify skills and skill levels for each task that correspond to "training goals". Different training exercises (Actions) will lead to different paths to the goal, and a POMDP solver can optimize the path to the goal, by selecting the optimal action for the determined probability distribution (belief state) of the individual student.

Conclusions and Recommendations for FUTURE research

In this paper we have described a framework for team training using a POMDP model. The framework includes cataloguing the skills involved in the domain, the training objectives, the training exercises available, the expected effect of each exercise on team skill levels, the set of possible measures, and the relation of these measures to trainee skill level. The approach uses principles from Bayesian Knowledge Tracing, Item Response theory, and Markov Processes, but vectorizes each of these models so that multiple team members are considered in the model, not just a single trainee.

These constructs imply several paths for future research. One avenue is the construction of a standard to communicate the above information, by extending GIFT's Domain Knowledge File (DKF) or through similar constructs. A standard such as xml or json could support specification of exercises and the applicability and difficulty of each exercise with respect to team skills. A second avenue is introducing optimization constraints into the model and to exercise selection based on real-world considerations, for example only a subset of the team members may be available on a given day, or only certain exercises may be available at certain times. If implemented, these extensions could be used to support GIFT in measuring, assessing, and recommending team training.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. The journal of the learning sciences, 4(2), 167-207.

Brunskill, E., & Russell, S. (2012). RAPID: A reachable anytime planner for imprecisely-sensed domains. arXiv preprint arXiv:1203.3538.

Carlin, A., Ward, D., & Freeman, J. (2016). Representation, Selection, and Scheduling of Training in a Lifelong Learning Context. ModSim World.

Carlin, A., Oster, E., Nucci, C., Kramer, D., & Brawner, K. (2016). Educational Data Mining Using GIFT Cloud. GIFTSym 5.

Carlin, A., Dumond, D., Dean, C., & Freeman, J. (2013). Higher Automated Learning. Short paper at *Conference Artificial Intelligence in Education 2013 (AIED).*

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological review, 100(3), 363.

Folsom-Kovarik, Jeremiah, "Leveraging Help Requests In POMDP Intelligent Tutors" (2012). *Electronic Theses and Dissertations*. 2262.

Levchuk, G.M., Gildea, K., Freeman, J., and Shebilske, W. A Benchmarked Experiential System for Training (BEST). (2007). Proceedings of the 12th International Command and Control Research and Technology Symposium, Newport, RI.

Levchuk, G., Shebilske, W., & Freeman, J. (2012). A Model-Driven Instructional Strategy. *Adaptive Technologies for Training and Education*, 303.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

Rowe, J. P., & Lester, J. C. (2015). Improving student problem solving in narrative-centered learning environments: A modular reinforcement learning framework. In International Conference on Artificial Intelligence in Education (pp. 419-428). Springer, Cham

Smallwood, R. D. & Sondik, E. J. (1973). The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, *21*, 1071-1088.

Sottilare, R., Holden, H., Goldberg, B., & Brawner, K. (2013). The Generalized Intelligent Framework for Tutoring (GIFT). In Best, C., Galanis, G., Kerry, J. and Sottilare, R. (Eds.) *Fundamental Issues in Defence Simulation & Training*. Ashgate Publishing.

Vygotsky, L. (1978). Interaction between learning and development. Readings on the development of children, 23(3), 34-41.

Chapter 6 – Dynamic Task Selection for Team Task Training Using Wearable Sensors and Multi-Agent Planning Models

Alan S. Carlin[1], Samantha K. B. Perry[1], Alec G. Ostrander[2]
Aptima, Inc.1, Iowa State University2

Introduction

In this paper we discuss how information from wearable sensors can be used to represent, measure, improve, and train team task performance in real-time. To accomplish this, we combine elements from frameworks in separate literatures, which in turn involves combining models of tasks with models of data.

Regarding task models, the AI-planning literature is focused on choices made by a decision-maker (human or automated). As such, the models in this literature have focused on representing the tasks themselves, and the near-term and long-term ramifications of each decision. Task models produced in this literature include the TAEMS model (Decker, 1995; Horling, Lesser, Vincent, Wagner, Raja, Zhang, et al., 1999), the DEC-POMDP model (Bernstein, Zilberstein & Immerman, 2000), distributed constraint optimization models (DCOP), and many others. In this chapter, we focus on examples that are loosely based on the TAEMS model, in which tasks are decomposed into a hierarchy of subtasks. The agent executing the tasks can use this information to choose which task to perform at any given time. In practice, one issue with the use of TAEMS is the level of effort involved in building the task structure, which involves specifying each of the subtasks in the model as well as their relations to each other. The more fine-grained the subtasks, the more information team executing the tasks has to plan out future tasks. On the other hand, a coarse granularity of subtasks results in fewer decision-points for the team, defeating the purpose of the representation. In this chapter, we will describe how to mitigate this effect by using during-task data and machine learning to improve agent reasoning mid-task.

Specifically, we use unobtrusive sensors data acquired during the task. Organizational analysts are increasingly turning to wearable sensors to understand and assess performance on a task. One reason is the nearly continuous data these sensors can capture. Another is the opportunity to capture interesting phenomena without interrupting the individual performing the task. To that end, multiple literatures in a variety of fields have focused on topics varying from how to generate better hardware to collect the data, to how to process the raw data from this hardware into meaningful measures, to how to assemble these measures together in a human-readable format. In training applications, these data have been found to inform real-time performance feedback through visualizations of the face to face communication pattern of the team members (e.g., the Command Operations Dashboard; DeCostanza, Orvis, Perry, & Brown, 2017) as well as real-time cognitive workload assessments through the analysis of wireless EEG data (e.g., Durkee, Pappada, Ortiz, Feeney, & Galster, 2015).

In this chapter, we examine how wearable sensors can be used to improve team task performance during execution of tasks that require online planning. First, we review the TAEMS planning model with an eye toward its strengths and limitations[3]. In the following section, we review the capabilities of a few types of wearable sensors. After introducing the relevant frameworks, we then propose a mechanism by which

[3] Our use of TAEMS in this chapter is loosely based. We maintain the core concepts involved in the framework, as reviewed in this chapter, but the examples may not be conformant to a release of exact specifications.

nearly continuous measurements from these unobtrusive wearable sensors can be used to assess task progress, which can, in turn, be used to re-plan future tasks.

The capabilities described in this chapter have relevance both operationally and in training. Operationally, better real-time predictions of task performance can lead to higher quality performance. In training, better assessments have benefits for both the trainee and trainer. For the trainee, real-time, individualized feedback on task performance can lead to better in-task decisions. The trainer can use this same information to provide expert feedback and guidance, and potentially make adjustments to the task as needed.

Multi-Agent Planning Models

In this section, we review a TAEMS-based (Task Analysis, Environment Modeling, and Simulation) framework as it pertains to modeling and planning team taskwork.

TAEMS

TAEMS "is a framework with which to model complex computational task environments that is compatible with both formal agent-centered approaches and experimental approaches" (Decker, 1995). Notably, an extension of TAEMS, called C-TAEMS was the framework used in the DARPA Coordinators program where human teams were instrumented with AI agents to participate in a mock disaster rescue study (Barbulescu, Rubinstein, Smith, & Zimmerman, 2010). We note three critical properties of the framework below, using an example from the medical domain (see Figure 1).

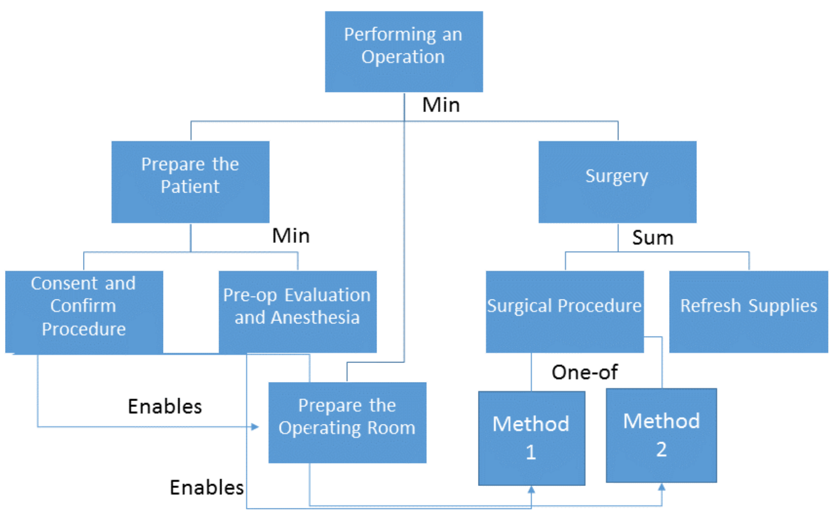


Figure 1. Example of Task Hierarchy from the Medical Domain

Tasks: Tasks have properties associated with them, including a distribution on the *duration* associated with task performance and a *quality* of how successfully a completed task has been performed.

Hierarchy: Tasks are decomposable into sub-tasks. Figure 1 below shows a task that is decomposed into three critical components of an operation: Prepare the Patient, Prepare the Operating Room, and Surgery subtasks. Parent nodes in the hierarchy are linked to their child-tasks through Quality Accumulation Functions (QAFs). In the figure, the quality of the Prepare Patient Task is the lesser ("Min") of the qualities achieved for the Consent and Confirm Procedure and Pre-Op Evaluation and Anesthesia subtasks (a "Min" QAF). That is, failing at any of the subtasks causes failure in the parent task since performance on the preparation task is driven by the worst performance among its subtasks. The quality of the Surgery subtask is the sum of the Surgical Procedure and the Refresh Supplies subtasks. And the Surgical Procedure can be performed via two methods, so only "One of" the subtasks will achieve quality.

Task Relationships: Tasks can have Non-Local Effects (NLE's) on each other, where completion of one task affects the distributions of quality and duration of another task. The Consent and Confirm Procedure task must precede the Pre-Op Evaluation and Anesthesia subtask less the patient be put under anesthesia prior to their final consent to the procedure, so we could draw an Enables relationship between these tasks (see Figure 1). Other types of NLE's include Facilitates, Disables, and Hinders.

Using this framework, it is possible then to define problems to maximize the quality of a root-level task in the hierarchy, within a duration. We will assume the presence of one or more agents, each of whom is capable of executing one or more of the subtasks at a time at the lowest level of the tree. At any given point in time, the agent must decide which task to continue along, or whether to assign itself a new task.

From a representation standpoint, the more detailed the subtasks are, the more complex the model and the more potential decision points there are for the agents to determine the best course of action; conversely, if the subtasks are too simplistic in nature, then there may be too few decision opportunities, defeating the purpose of the model. In practice, however, the generation of elaborate task models is onerous and can potentially discourage adaptation. Before delving into striking an appropriate balance, we first discuss ways in which tasks can be captured to the appropriate level of detail and data granularity. Wearable sensors are an excellent example of that.

Wearable Sensors

Wearable sensors have become a popular avenue for measuring individual and team psychology phenomena through a variety of different sources. While the sensors themselves are variable and have different capabilities, it is the quality, and in some cases quantity, of the signals that tend to be of greatest importance when considering their ability to capture constructs. The next sections will delve into two categories of sensors – those that focus on capturing interpersonal interactions, and those that measure an individual's physiological state.

Most typically used in team training tasks are sociometric sensors, often referred to as "badges". These are designed to capture the social interaction patterns of individuals in teams. While several different badges have been developed over the years (e.g., Michigan State University's - see Baard, Kozlowski, DeShon, Biswas, Braun, Rench, et al., 2012; Sociometric Solutions Inc.'s (SSI) - see Kim, McFee, Olguin, Waber & Pentland, 2012), they share some common elements. A sociometric badge is an electronic device, typically worn around a user's neck like a name tag that is used to gather data about interactions between members of a group. They are designed to be relatively small and inexpensive enough to be deployed at up to an organizational scale. The badges are equipped with a variety of sensors. All have some sort of interaction capturing signal (e.g., infrared, Bluetooth and or RFID) and an accelerometer to capture motion. Some also have a microphone to capture frequency and amplitude of sounds. Although each signal by itself has limited utility, through multimodal analysis methods, researchers have been able to uncover emergent features that give numerous insights into the way members of a team interact. For example, a time series of the volume data assesses how often and for how long an individual speaks. From the audio frequency, one can also estimate the speed and energy with which the user speaks (Lederman, Calacci, MacMullen, Fehder, Murray, & Pentland, 2016). By examining the audio from multiple users in tandem, prior work has obtained a high-level picture of the conversation through metrics such as how much turn-taking was done in the course of a conversation (Kim et al., 2012). Other work has analyzed badge accelerometer data to assess movements such as gestures and changes in posture, while the infrared sensor and the relative signal strength can paint a picture of how the group is physically distributed (Baard et al., 2012; DeCostanza et al., 2017; Kim et al., 2012). Recent work has used face to face and proximity signals to capture the social networks and boundary spanning behaviors of a multi-team system (Brown, Perry, Braun, McCormack, Orvis & DeCostanza, 2017) and even applied this information to real-time training feedback during live Army exercises through their innovative Command Operations Dashboard (DeCostanza et al., 2017).

Physiological sensors are another prevalent type of relatively unobtrusive measurement device. Although not a comprehensive list, some of the most common signals captured are: electroencephalography (EEG), electrocardiography (ECG), heart rate variability (HRV), respiratory sinus arrhythmia (RSA), photoplethysmography (PPG), electromyography (EMG), galvanic skin response (GSR), and eye-tracking systems. These systems have seen advances not only in the research but also in their transportability and unobtrusiveness. For example, EEG devices have burgeoning capabilities such as dry-electrode configurations, longer battery life, wireless data transfer, reduced weight and increased aesthetics. Research advances have continued with reference to understanding the individual; however, some researchers have extended the work to capture team states from an aggregation of the physiological states of individuals. For instance, Baard (Perry) and colleagues used HRV to understand the cognitive state and subsequent performance of teams (Baard et al., 2012) while others have used HRV and RSA to measure the performance of a team in a high-stress task (Elkins, Muth, Hoover, Walker, Carpenter & Switzer, 2009). Further research has also taken these physiological measures and applied it to real-time feedback during training simulations. Durkee and colleagues have developed cognitive workload measure using wireless EEG sensors and their functional state estimation engine (FuSE[2]) to provide the team with a real-time assessment of the relative workload of the individuals in the team during a task, allowing the team an opportunity to correct instances of excessively high workload across appropriate team members (Durkee et al., 2015).

The next section will tie the TAEMS model with the wearable sensor data output to discuss how to use sensor data to predict the outcome of tasks in the model, and in turn how to use the predicted outcome to make decisions about whether to continue tasks or select new ones.

Blending Task Model and Wearable Sensors

In this section, we bring together the TAEMS model with wearable sensor information in the description of a two-step process for updating task planning during execution. The steps involved are:

1. Predict quality and duration distributions based on sensor data.
2. Update task selection or training needs based on the existing quality and duration distributions.

Step 1: Quality and Duration Distributions

The step of predicting quality and duration distributions can be cast as a supervised learning problem over wearable sensor data. To do so, one can gather a data set labeled by task. One can format this data into a matrix of relevant information, as shown below in Table 1.

A supervised learning algorithm[4] can learn the rule by which Output Quality and Duration labels are predicted, and the feature values can be extracted from sensors. That is, the algorithm would take all available

[4] Many supervised learning algorithms are suited to operate on data formatted in the format of Table 1. Examples of suitable algorithms are regression (with data manipulation to turn the nominal variables into ordinal), decision trees, support vector machines, or deep learning networks.

data from the wearable sensors, populate the table, and complete the table with the expected quality and duration metrics when those are measured. This prediction can be used by the instructor when deciding whether to provide hints or other instruction.

The columns of Table 1 contain processed sensor data. One area of interest during training is determining whether an individual in a team is being overtasked and whether the other individuals should work to mitigate that issue through backup behaviors. Through measuring the cognitive activity level of each individual (the sensor being the wireless EEG), an algorithm can determine the relative cognitive state of each individual on the team. As seen in Figure 2, networks not only present a visual means for investigating the team interaction patterns given that each dot – or node – is an entity (e.g., individual) and line is the relationship between those individuals (e.g., pieces of information transmitted), but it can also provide meaningful analytics. For example, network centrality, as depicted by the highlighted individual in the center of the network, indicates that this individual plays a critical role in the phenomenon displayed in this team. Therefore, Figure 2 may show the most central individual as a hub of information if the lines indicate the number of pieces of information communicated. Through capturing the network centrality metric (Freeman, 1979) of interaction patterns (the sensor being the Bluetooth or infrared features of a badge), one can understand whether the individual is engaged in the task with others, or off in an entirely different part of the workspace. Through combining the metrics of cognitive workload with interaction patterns, the system can determine whether a particular individual is being overworked during a task or during a point in time and which individuals may be best able to assist that person based on location and task familiarity.

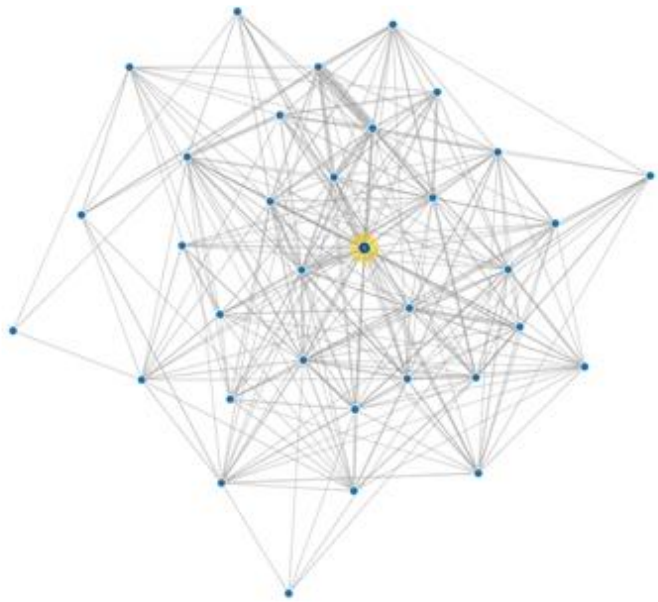


Figure 2. A Network Diagram Highlighting Centrality

This can be seen in Table 1 through understanding the cognitive workload (via EEG data) and network centrality (via interaction patterns interpreted from badges data) of the surgeon during the task. In the first row the surgeon may have been inexperienced or thought that the task was extremely easy, so he took his time and worked with relatively few other individuals. The output may be acceptable, but given the length of time the surgeon took to do the task, he and his team need additional training. In contrast, the surgeon who had very high workload and did not have a high network centrality suggests that this individual needed assistance. For example, a rule could pre-process by changing the categorical variables into numbers, and then learn a function: e.g.

$$Output\ \ Quality = f(Cognitive\ Workload, Network\ centrality)$$

The rule that is learned from the data can be used to predict the outcomes of future procedures. The last row of the column shows a hypothetical new surgical procedure, and the machine learning rule learned from the rest of the table would be used to predict the entries listed as “To be predicted”.

Table 1. Example of Step 1 Data

Task	Time since task start	Surgeon Cognitive Workload (EEG)	Surgeon Network Centrality (Badges)	Output Quality of Surgery (min 0: max 100)	Output Duration of Surgery	Team Needs More Training
Surgical procedure	15 min	Low	Low	70	Too long	Yes
Surgical procedure	10 min	High	Low	80	Too long	Yes
Surgical procedure	5 min	Moderate	High	100	Excellent	No
Surgical procedure	10 min	Moderate	Moderate	70	Acceptable	No
Surgical procedure	2 min	High	Low	To-be predicted	To-be predicted	To-be predicted

Step 2: Update task selection

The predictions in the previous section will result in an updated estimate of duration and quality for that task and potentially other tasks. These revisions will, in turn, be used to task or re-task trainees. For example, the data captured in the final row of Table 1 suggests that without intervention, the team will likely continue to let the surgeon be overloaded. Therefore, the system can recommend that another team member, such as a nurse, temporarily switch from the resupplying during the surgery task to assist in the surgical procedure.

To support this, the tutor should maintain a table of expected values for completing each unfinished task in the hierarchy. For tasks in the leaf nodes of Figure 1, this involves a direct computation of the expected quality of completing each task. The changed value would propagate up the hierarchy; changes in the distribution of child nodes trigger a re-computation of the value of the parent nodes. The trainee selects the task with the largest impact on the root node. Although this example strategy has the property of being *myopic* (it only considers the next action and does not plan ahead), it can be expanded to a non-myopic strategy by using a Markov Decision Process to compute the values.

For an example, reference Table 2 below. The additional breakdown of the operation procedure into its component elements shows that the respective qualities of the components (e.g. the preparation of the patient, the preparation of the room, and the surgery itself) have an impact on the overall operation. If the patient has not consented and the procedure is not confirmed, then there would be a negative impact on the tasks of preparing the patient and preparing the room. Additionally, refreshing the supplies is dependent on the room state at the beginning of the surgery as well as the quality of the surgery (i.e., if there is considerable blood loss or an unexpected change, more supplies would be needed). Specifically, it would be unwise to skip the Consent and Confirm Procedure task, because of the diminished quality it propagates to the Prepare the Patient and Prepare the Operating Room tasks, which in turn propagates to the Performing an Operation task.

Table 2. Example of computation of expected quality contribution with and without the Consent and Confirm Procedure subtask

Task	Expected Quality Contribution with Consent and Confirm procedure	Expected Quality without Consent and Confirm Procedure	Comment
Performing an Operation	95	0	Min of Prepare the Patient and Surgery
Prepare the Patient	95	0	Assigned lesser value of Consent and Confirm and Pre-Op Evaluation
Consent and Confirm Procedure	100	0	
Pre-Op Evaluation and Anesthesia	95	95	Suppose by hypothesis for this example that these preparations are not affected by Consent and Confirm
Prepare the Operating Room	100	0	Enabled by Consent and Confirm Procedure (Enables link not shown in Figure)
Surgery	95	10	Sum of Surgical Procedure and Refresh Supplies
Surgical Procedure	95	0	Without Consent and Confirm, inherits 0 quality from either method. With Consent and Confirm, Method 1 will be selected and quality achieved will be 85.
Refresh the Supplies	10	10	Proceeds with or without Consent and Confirm
Method 1 Surgery	85	0	Consent and Confirm enables Method 1. Cannot proceed without it.
Method 2 Surgery	80	0	Consent and Confirm enables Method 2. Cannot proceed without it.

Conclusions and Recommendations for future research

With the growth of research on and applications for wearable sensor data, the mechanisms available to researchers with interest in pursuing complex analytical solutions to difficult and dynamic problems is ever expanding. Advances in both sociometric and physiological sensors have benefitted both individual and team level initiatives, and some have had success providing real-time training feedback (see DeCostanza et al., 2017 and Durkee et al., 2015 for examples); however, there has not been a link between these measurement strategies and the AI literature.

This AI-planning perspective on the issue of training recommendations provides scientists and engineers with a common communication platform. In this paper we discussed how information from wearable sensors can be used to represent, measure, improve, and train team task performance in real-time. The TAEMS model provides particular insight through its ability to use task structures to provide agent logic for the determination of which tasks should be pursued next.

This chapter covered, with a broad brush, the applications of wearable sensor data in the TAEMS models using one example of a complex task from the medical field. Future research in this vein would benefit from a more detailed analysis of the TAEMS framework with regard to the details of complex and dynamic tasks to determine how task sequencing and interdependence would play a role in the model development. Additionally, as researchers continue to pursue team and multi-team systems work, future studies on the TAEMS model may benefit from an examination on how feedback can be effectively given to individuals within a multi-team system within the context of the complex task planning model and whether there are any limitations associated with the size of the model or the task distribution of the individuals in the team.

The inclusion of a task model would levy some requirements on any tutoring framework. For example, to implement these ideas, the Generalized Intelligent Framework for Tutoring (GIFT) would need to represent the tasks being assigned, including their duration, quality, and relation to each other. Depending on the application, the system could need to be capable of tracking which tasks have been completed, which tasks are outstanding, which tasks have been assigned to whom, and how the completion rolls up into overall quality. Furthermore, to implement the ideas proposed in this chapter regarding associating sensor data to task quality, the distributed data would need to be collected and timestamped appropriately.

The capabilities described in this chapter have relevance both operationally and in training. Operationally, performance can be enhanced by better predictions of what tasks individuals should be performing, even though recommending an individual switch tasks mid-way through in such cases where one teammate is overloaded and unable to maintain their workload on a critical task, for instance. In training, not only do enhanced measurements provide more reliability in the assessments, but the enhanced assessments from the proposed models will provide both the trainer and trainee with valuable and timely feedback that will facilitate the growth of knowledge and skills, resulting in enhanced operational performance.

References

Baard, S. K., Kozlowski, S. W. J., DeShon, R. P., Biswas, S., Braun, M. T., Rench, T. A., Pearce, M., Bo, D., & Piolet, Y. (2012, April). Assessing team process dynamics: An innovative methodology for team research. In G. F. Goodwin and A. H. DeCostanza (Co-chairs), Get out of the way! Unobtrusive measures of team constructs. Symposium conducted at the annual conference for the Society for Industrial and Organizational Psychology, San Diego, CA.

Barbulescu, L., Rubinstein, Z. B., Smith, S. F., & Zimmerman, T. L. (2010, May). Distributed coordination of mobile agent teams: the advantage of planning ahead. In Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1 (pp. 1331-1338). International Foundation for Autonomous Agents and Multiagent Systems.

Bernstein, D. S., Zilberstein, S., & Immerman, N. (2000, June). The complexity of decentralized control of Markov decision processes. In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (pp. 32-37). Morgan Kaufmann Publishers Inc.

Brown, T. A., Perry, S. K. B., Braun, M. T., McCormack, R., Orvis, K. L., & DeCostanza, A. H. (2017, July). A dynamic exploration of multi-team system face-to-face boundary spanning. In T. Brown & S. Perry (Co-Chairs), MTSs: Dynamics at the boundaries is what matters. Symposium presented at the 2017 Interdisciplinary Network for Group Research Conference, Saint Louis, MO.

Decker, K. (1995). TAEMS: A framework for Environment Centered Analysis & Design of Coordination Mechanisms. In Foundations of Distributed Artificial Intelligence. G. O'Hare and N. Jennings, eds. New York: John Wiley and Sons.

DeCostanza, A. H., Orvis, K. L., Perry, S. K. B., & Brown, T. A. (2017, April). Real-Time Feedback for Teams using Unobtrusive Assessment Techniques. In S. K. B. Perry & G. F. Goodwin (Co-chairs), Beyond Unobtrusive Methodologies: The Intrusive Component of 'Big Data' Research. Symposium Presented at the 32nd Annual Conference for the Society for Industrial and Organizational Psychology, Orlando FL.

Durkee, K. T., Pappada, S. M., Ortiz, A. E., Feeney, J. J. & Galster, S. M. (2015). System decision framework for augmenting human performance using real-time workload classifiers. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision, pp.8-13.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. Applied Ergonomics, 40, 997-1003.

Freeman, L. (1979). Centrality in social networks: Conceptual clarification. Social Networks 1(3) 215–234.

Garguilo, M., & A. Rus (2002). Access and mobilization: Social capital and top management response to market shocks. INSEAD.

Horling, B., Lesser, V., Vincent, R., Wagner, T., Raja, A., Zhang, S., Decker, K., and Garvey, A. (1999). The TAEMS White Paper.

Kim, T., McFee, E., Olguin, D. O., Waber, B. & Pentland, A. (2012). Sociometric badges: Using sensor technology to capture new forms of collaboration. Journal of Organizational Behavior, 33, 412–427.

Lederman, O., Calacci, D., MacMullen, A., Fehder, D. C., Murray, F. E., & Pentland, A. S. (2016). Open Badges: A Low-Cost Toolkit for Measuring Team Communication and Dynamics. Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS), Washington, D.C.

Section II - Team Assessment Methods

Dr. Arthur Graesser, Ed.

Chapter 7 – Assessment in Team Tutoring

Arthur C. Graesser
University of Memphis

Core Ideas

The chapters in this section focus on the ***assessments*** of teams and individual team members. Team science has identified characteristics of well-functioning teams in different contexts that range from coordinated psychomotor tasks to the design of sophisticated digital artifacts. The members of a team *ideally* are aware of the goals they attempting to accomplish, have a shared vision of how to accomplish the goals, have an understanding of the team organization and each of the members' roles, have frequent updates on who does what and when, and have reliable communication among team members. A team *ideally* can adjust when confronted with obstacles to goals and conflicts that invariably occur, and can dynamically replan the agenda and roles when necessary. However, most teams are not ideal. Sometimes individual team members do not deliver on their assigned tasks, as in the case of social loafers, incompetent team members, or saboteurs. There can be serious repercussions from breakdowns in team performance that can be attributed to a dysfunctional team member or the team as a whole. Assessments of team and individual performance are therefore necessary.

Summative, formative, and stealth assessment are often contrasted in the assessment literature. *Summative* assessment is a score (or set of scores) collected at some point in time for the purpose of measuring the overall performance or capability of an individual or team. A summative assessment may have practical consequences, such as personnel selection, job classification, certification, graduation, or promotion. *Formative* assessment is a score collected during learning or executing a task that is used as feedback to the learner or instructor. The feedback is designed to help the individual improve learning and performance during the learning process. Formative feedback is particularly important in intelligent tutoring systems, including the Generalized Intelligent Framework for Tutoring (GIFT), to improve the performance of teams or individual team members. The feedback can be directed to the team as a whole, to individual team members, or to subgroups. *Stealth* assessment also includes scores collected during learning or task execution, but explicit feedback on performance is not directly given to teams or individuals. Instead, the assessment guides the dynamic selection of pedagogical activities by the digital tutor without the team or individual being aware of the assessment. Stealth assessment is also part of the GIFT architecture.

The chapters in this section collect assessments throughout team activities by a fine-grained analysis of log files that record the stream of actions, events, processes, timing, self-reports, and sometimes physiological, neurophysiological, and emotional responses via sensing modules. The data may consist of open-ended behavior, such as verbal reports, answers to questions in natural language, conversational chat, collaborative interaction, and actions during problem solving. The raw data and derived measures are stored in a learning record store that continuously updates the learner model. The pedagogical module makes use of these data and adaptively computes appropriate pedagogical goals and tactics to improve learning and performance. These mechanisms follow the architecture of GIFT, but there is one important layer of additional complexity: The intelligent tutoring system needs to monitor, regulate, and provide feedback to the team, and sometimes subgroups, in addition to individuals.

Individual Chapters

The chapter by ***Stevens, Gorman, Zachary, Johnston, Dorneich,*** and ***Foltz***, "How is the Team doing, and why?" describes a number of projects that collect multiple, semi-independent, decision-making components that contribute to system dynamics and that cover macro to micro time scales. One project measures team neurodynamics whereas others measure team communication and responses to feedback in an intervention. The sensory and communication data are rich so there needs to be ways to abstract, compress, and transform the data, sometimes by taking advantage of the natural redundancy in biological signals or communication patterns. The assessments need to be computed quickly on-line so that feedback can be given to team members in a timely manner.

The chapter by ***Sinatra, Kim, Johnston,*** and ***Sottilare***, "Assessment of Team Performance in Psychomotor Domains" describes the processes, challenges, and potential solutions to assessing team performance in real time for psychomotor tasks. The authors specifically describe a military room clearing task, and discuss how it could be assessed in an intelligent tutoring system. Assessments required for team psychomotor tasks involve analyzing physical movement and physiological responses in order to compute the alignment of such data with a model of expert individual and team behavior. It is important to determine the unique sensors and technologies that need to be implemented in order to engage in assessments of both individual and team accomplishments in a psychomotor task.

The chapter by ***Goldberg, Nye, Lane,*** and ***Guadagnoli***, "Team Assessment and Pedagogy as Informed by Sports Coaching and Assessment" also focuses on the psychomotor domain, with a focus on sports (football and baseball). They review the pedagogical insights about sports coaching and assessment from three sources: published reports on sports training, first-hand accounts of team training, and a review of measures of team performance. Feedback is central to a coach's activity so the chapter identifies the different features, timing, and recipients of feedback. There are different roles of team members so team sports require coordinated interdependent activities and communication. The authors propose that a pedagogy for teams should be grounded in a situated understanding model.

The chapter by ***Zachary, Goldberg,*** and ***Hampton***, "The Role of Context in Team Performance and Team Training" examines a theory of context understanding and considers how context contributes to team performance and assessment. Context is defined as a "cognitive process that is representation-centric, constructive, pervasive, and strongly interconnected with domain expertise." Moreover, they argue that context can be framed and modeled as an explicit computational process to be carried out by computational devices, which would be essential for GIFT. Stories/narrative, situated awareness, perceptual features, and links to expertise are among the important considerations for team training in military domains, such as dismounted infantry and naval fleet air defense.

The chapter by ***Hu, Dowell, Cai, Graesser, Shi, Cockroft,*** and ***Shorter***, "Constructing Individual Conversation Characteristics Curves (ICCC) for Interactive Intelligent Tutoring Environments (IITE)" proposes a computational model that assesses conversational interactions among agents (humans or computer avatars) in intelligent tutoring environments. The model builds on group communication analysis, a methodology for quantifying the discourse dynamics and sequential interactions between agents in multi-party interactions. The model automatically analyzes emergent roles of team members based on each member's interaction profile (individual communication characteristic curves, based on semantic analysis and similarity). The profile has measures of each team member's participation, responsivity, internal cohesion, social impact, newness of information in the conversation, and semantic density.

Chapter 8 – How Is This Team Doing, and Why?

Ron Stevens[1], Jamie Gorman[2], Wayne Zachary[3], Joan Johnston[4], Michael Dorneich[5] & Peter Foltz[6, 7]
The Learning Chameleon, Inc.1, Georgia Institute of Technology2, Starship Health Technologies, LLC3, U.S. Army Research Laboratory4, Iowa State University5, University of Colorado-Bolder6, Pearson Education7

Introduction

Team and unit effectiveness is the primary goal of many military, healthcare and organizational training programs. Numerous military training studies have produced detailed guidance on the importance of teamwork behaviors in performing complex, tactical combat tasks (Salas, Benishek, Coultas, et al., 2015). Meta-analyses and other literature reviews indicate a common core of teamwork behaviors that are agnostic to the type of combat tasks being performed (Salas et al., 2015). They include exchanging relevant combat-related information, using proper communication protocols, taking the initiative to provide information without being asked, providing situation updates, and correcting team mate errors.

Training research to improve these common teamwork behaviors has demonstrated that following a training exercise providing actionable feedback to teams about their behaviors reduces tactical errors (Salas, Diaz-Grandos, Klein, et al., 2008). For example, Smith-Jentsch, Cannon-Bowers, Tannenbaum & Salas (2008) demonstrated that Simulation-Based Team Training (SBTT) using After Action Review (AAR) protocols for team self-correction resulted in improving the quality and quantity of information passed among team members.

Much of this research has had to rely on subjective evaluations of voice communications, but there are considerable logistical problems (i.e., it is labor intensive and time consuming) for implementing in operational environments. Similar logistical arguments have limited extensive team performance assessment and AARs where best practice currently recommends spending two to three times longer debriefing compared to the actual time spent in the simulated scenario. A final problem is that the benefits of the enhanced AARs take time to accrue, and this time required is seen as an added burden for already heavily burdened training pipeline.

In response there have been increased calls for research to solve the logistical problem of implementing improved SBTT through the development of adaptive tutors (Gilbert, Slavina, Sinatra, et al., 2018; U.S. Army Human Dimension strategy). The tremendous advantage of adaptive tutors is they enable tailoring training to specific learning needs, enable assessing learning and performance in minutes vs. hours, can provide real-time scaffolding rather than waiting for post-scenario feedback, and over time, can be implemented iteratively to mitigate skill decay. However due to the costs and time required, team training research has been somewhat limited in the study of how teams change and develop over time, and how various interventions may be needed to accelerate development.

'How is this team doing?' is the most dynamic of the triad of questions 'How did this team do?' 'How is this team doing? and, 'How will this team do?' which are each key elements for understanding how teams develop over time and how their learning can be supported through feedback, scaffolding and prediction. Answering these questions require dynamic assessments with a past to draw inferences from and a future to extrapolate to. The measures generated for answering these questions should be objective, quantitative, able to be compared across team members and teams, and applicable to a variety of tasks. For incorporation into an Intelligent Tutoring System (ITS), they would need to be generated in near real-time to support scaffolding while also being able to be accumulated into aggregated team and team member values for feedback across sessions.

We are still a way from these capabilities, but as illustrated in this chapter pathways forward are becoming visible that leverage increased sensor and analytic capabilities and point to realistic sequences of tool development. Such tools will support empirical quantitative comparisons across teams, tasks and experience and help uncover the cognitive interactions between team members with the evolving context. This will enable cognitively-informed task designs and accelerate the rates of team and team-member learning by focusing on the cognitively-relevant properties of performance.

Even with refined models of each team member and the team, the tools will still be limited by larger issues like team process and outcome. It is possible for each individual action of each individual of a team to be correct and for the team to coordinate according to standard policies and procedures, and still arrive at an unsatisfactory outcome. Such was the case with the USS Vincennes, which shot down a civilian airliner in the Persian Gulf in 1988 (see Cannon-Bowers and Salas, 1998). If all the processes were correct yet the outcome was very bad, what should the assessment be – was the team doing well or terribly? Conversely, a single action by a single individual in a team can lead to a very negative state for the mission. But if everyone else was doing well, and if the team adapted to the negative situation and resolved it, was the team doing well or poorly?

This chapter draws on the experiences of researchers with a diversity of experimental, and experiential understandings of teams. Our goal is to identify paths that might help shape our current thinking about teaming in ways that will enable more rapid and directed incorporation of feedback and scaffolding into team training activities, and eventually into ITS architectures.

A theme that runs through the projects described is how to extract information from multi semi-independent decision-making components who are all contributing to system dynamics and do so in ways that capture the variables that the system uses to tune itself across macro to micro scales. Ways are needed to reduce the dimensionality of the data by abstracting, compressing, transforming and taking advantage of the natural redundancy in biological signals. Specific representations or models need to be constructed to re-create or predict aspects of a (or multiple) teamwork tasks. The number of variables involved are difficult to imagine across the timescales of teamwork, and many of the variables coordinate locally and their primary activities are only weakly connected with to behaviorally-relevant functions.

Expert decision makers universally develop mental models of the external context and use them for relevant decision and action planning (Zachary, Rosoff, Read & Miller, 2013). This requires the assessment of abstract process or outcome measures in current mission context (and path by which it got there) and also the abstraction of the unit of behavior that must be observed and measured. This behavioral abstraction need increases the time scale of observation and assessment. For example, the Advanced Embedded Training Systems (AETS) was originally designed to capture individual keystrokes, eye movements, and word-by-word speech of each operator (Zachary, Canon-Bowers, Bilazarian, et al.,, 1999). It turned out, however, that these units needed to be aggregated into larger units, called High Level Actions (HLAs), that represented both an abstracted intention with a specific intended action in the mission space, in order to assess either abstracted process or abstracted outcome measures.

Such abstractions are currently derived by cognitive task analysis of domain experts, who generally agree on the how they talk about the problems space in functional terms. These are then built up (or more accurately decomposed down to) the various combinations of keystrokes, word and eye-movement when needed to get to an identifiable HLA, In other words, it is important therefore that the most strategic variables for the changing contexts are captured.

The following sections describe different ways that investigators have used compression, abstraction and transformation to move the evaluation of teamwork towards more granular models, and perhaps towards ones that will be more compatible with the real-time notion of scaffolding in an ITS.

Example 1. Team Neurodynamics

People rapidly develop organizations in a group setting, and these structures can be extracted through interaction networks using speech flow and interaction technologies (Gorman Dunbar, Grimm & Gipson, 2017; Fishbach, Gloor, Lassenius, Olguin, et al., 2009). Once established, these inter-personnel relationships tend to be stable (Flack, 2012) until the relationships are perturbed by external events or internal conflicts. They must then reorganize their thinking, roles and/or configurations into corrective structures more appropriate for the situation. The neurodynamic properties of these reorganizations are poorly understood. For instance how do different organizational states relate to the established stability of well-performing teams? What are the likely changes to the neurodynamics of the team, and each team member, that might occur following unexpected disturbances?

Neurodynamic organization is the tendency of team members to enter into prolonged (minutes) metastable neurodynamic relationships as they encounter and resolve disturbances to their rhythms. Neurodynamic organization is often determined by using electroencephalographic (EEG) signals. EEG is the recording of electrical activity of the brain at different regions along the scalp. The rhythmic patterns in the electrical oscillations from different brain regions contain signals representing complex facets of brain activity.

When modeling the neurodynamics of teams, the goal is to develop representation(s) that use EEG amplitudes (in µ-volts) of the team members as inputs and then export higher level representations of team organization. In order to contribute to an effective theory of teamwork the representation developed by any combination of abstraction, compression and transformation must not only link to the microscopic, but also help explain the macroscopic. Coarse-grained descriptions are ones where some of the system fine details have been smoothed over, and some of the information we had about the system is lost (Flack, 2012). The description however remains true to the underlying systems.

Ideally such representations could bridge across the micro, meso, and macro levels of team dynamics (Fiore, Smith- Jentsch, Salas, et al., 2010) in that the EEG amplitudes could be used in a downward manner to understand the theoretical basis for the micro-dynamics of social neuromarkers while the organizational units could be used in an upscale manner to link to behavioral markers that contribute to team learning (Sottilare, Burke, Salas, et al, 2017). In this way a composite physical-information hierarchy was constructed that spans Marr’s hierarchy of analysis with function and computation at the highest level, algorithms in the middle level, and biophysical substrate at the lower level (Marr, 1982).

During teamwork the rapid EEG oscillations that emerged on the scalp are transformed into symbolic data streams which provide historical details at a second-by-second resolution of how the team neurodynamically perceived the evolving task and how they adjusted their dynamics to compensate for, and anticipate new challenges. The first modeling step separated the EEG amplitudes of each team member each second into high, medium and low ranges. This resulted in a 500-fold data compression from the raw EEG signals that still maintained important functional characteristics (i.e. active processing vs inhibition of alpha waves), averaged over larger time scales. For ease of visualization, these high, medium and low amplitude categories were assigned the values 3, 1 and -1. Within limits, and depending on the questions being asked, these values could be used either numerically as indications of EEG power or symbolically as any other symbol collection. This results in a three-element vector. When histograms of these vectors were symbolically combined for the three team members, they created a three-histogram symbol representing the neurodynamic state of the team at that second. The values of the three histograms in Figure 1A indicate that at this second, team member 1 had below average EEG power levels, team member 2 had above average and team member 3 had average EEG power levels; so the vector for this neurodynamic symbol (*NS*) was -1, 3, 1. The possible combinations of three persons with three EEG amplitudes created a twenty-seven symbol neurodynamic state space (*NSS)* that encompassed all of the possible team member-power combinations for the performance (Stevens et al, 2012). While developing the *NSS*, a topology was generated such that the symbols in the upper left corner (1-5, 10-14) represented times when most team members had low to average levels of EEG power, while the symbols on the right side and bottom row, numbers (9,18,24-27) represented times when most team members had above average power levels. Each *NS* in the symbolic state space therefore situates the EEG power levels of each team member in the context of the levels of the other team members and the context of the task. This assembly of the power levels of three persons into a single symbol further reduced the dimensionality by a factor of three.

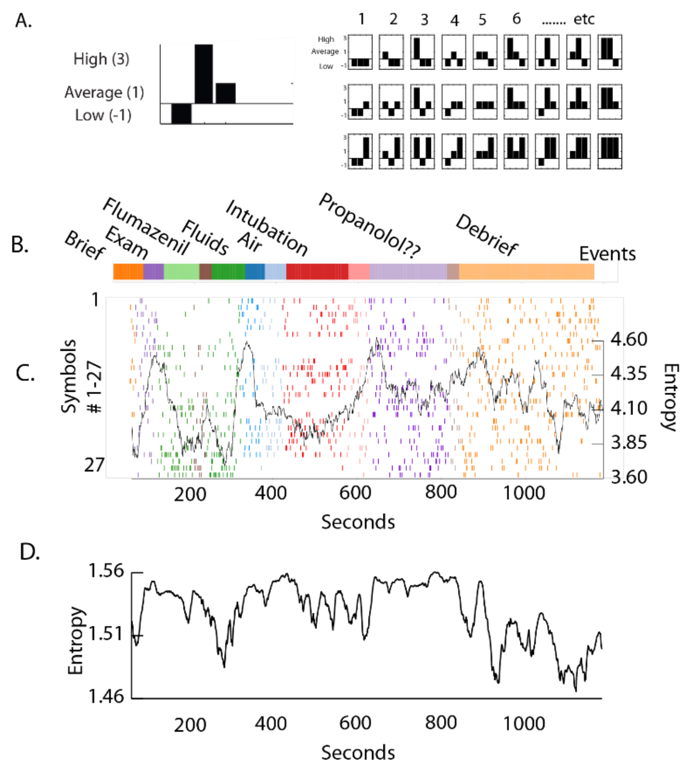


Figure 1. Neurodynamics of a healthcare team and their instructor. A) Composition of a Neurodynamic Symbol and a Neurodynamic Symbol Space. B) The segments highlight important sequences of events during a three-person training segment. C) The symbols (Y-axis) in the neurodynamic data stream of the team were sequentially plotted each second (X-axis). A quantitative profile of the variability in the symbol distribution over a time segment was obtained by calculating the entropy over a 60s moving window that was updated each second. The entropy represents the bits of information in the 1-40 Hz frequency bands of the P7 channel which had the lowest entropy during the training session. D) A similar entropy profile of the instructor observing the training.

The next processing step switched from dimension reduction to visualization where the structures in these symbolic data streams were revealed by sequentially plotting each symbol as it appeared in the sequence allowing changes in symbol expression to be linked with task events (Fig. 1B). An example of these dynamics are shown in Fig. 1C for a three-person healthcare team treating a simulated patient with a suspected drug-overdose.

In all teams studied the temporal distribution of symbols in the data streams has been non-uniform, with a limited subset of symbols being expressed for a minute or more, only to be replaced by another symbol subset when the task demands changed. These symbol concentrations produced local variations in the randomness, or entropy, of the neurodynamic data streams. Entropy is the average surprise of outcomes sampled from a probability distribution or density. A density with low entropy means that, on average, the outcome is relatively predictable, while a system with higher entropy would be less predictable. Entropy is therefore a measure of uncertainty. In this way a quantitative and dynamic profile was constructed which can be reported with a 1s granularity for real time modeling, or aggregated over a performance for comparisons across teams (Stevens & Galloway, 2017).

As performances with submarine navigation teams were accumulated a trend emerged whereby more junior officer teams showed lower entropy levels than did experienced submarine crews. This resulted from more frequent entropy fluctuations of greater magnitude and duration (see Figure 2). A follow-up study compared the team ratings, conducted by third-party instructors, with levels of neurodynamic entropy during the simulation training (Stevens, Galloway, Lamb, Steed, Lamb, 2017). A positive correlation was seen such that the higher the team entropy levels, the higher the ratings, i.e. the less neurodynamically organized the team, the better they performed. This neurodynamic organization-performance relationship provides an opportunity for delivering both feedback to the team (i.e. aggregated entropy levels for the performance) and scaffolding to the team (via when they have entered a persistent period of neurodynamic organization).

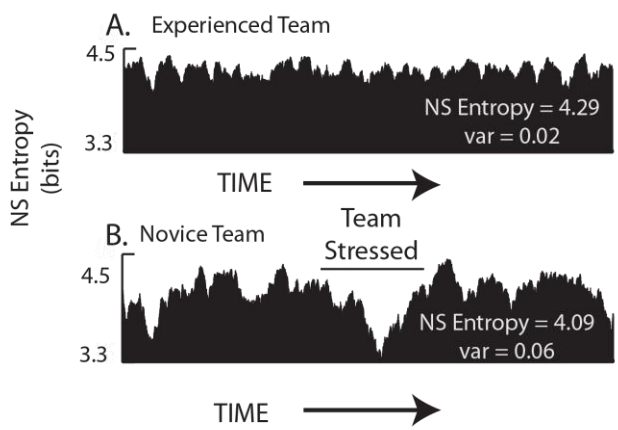


Figure 2. Sample neurodynamic entropy profiles for (A) an experienced and (B) a novice submarine navigation team.

The symbolic neurodynamic modeling described for three-person teams can also be applied to each team member within a team by using the -1, 1, and 3 values symbolically and then calculating the temporal entropy changes. The maximum entropy for three-symbol systems becomes 1.58 bits as opposed to 4.76 bits for three-person teams. This allows direct quantitative comparisons of the individual neurodyanmic organization of each person as well as the team as a whole.

Entropy, as measured by the changing distribution of symbols, is a measure of the organization and information in a system, but during teamwork, the entropy changes are over a large background of 'noise'. To better analyze and visually represent the dynamics, the experimental entropy was subtracted from the maximum entropy (4.76 bits for twenty-seven symbols), which resulted in neurodynamic organization becoming a positive value. With the individual and team information both calculated in bits of information, quantitative comparisons became possible between team members as well as their individual contributions to the team. (Stevens & Galloway, 2017; Stevens, Galloway & Willemsen-Dunlap, 2017). This is the first time such quantitative comparisons have been able to be made between team members and the team during complex, natural team training.

A comparison of the neurodynamics of a healthcare team and of the three team members is shown in Fig. 3. One finding that was consistent across submarine navigation, healthcare and high school problem solving teams is that the team neurodynamic organization/information is similar, but not equal to the sum of the individual information. The difference lies in the information that is not unique to an individual, but is shared by one or more other teammates (Fig. 3C) (Stevens & Galloway, 2017; Stevens, Galloway & Willemsen-Dunlap, 2017). The dynamics of each of the individuals is often quite different from one another, which is consistent with the idea that they each are acting semi-autonomously as they perform their required task work. To the extent that increased neurodynamic organization of individuals occurs during periods of uncertainty, stress and other measures of increased attention (Stevens, Galloway, Willemsen-Dunlap and Halpin, 2017), the individual neurodynamics may provide places to insert scaffolding or feedback triggers.

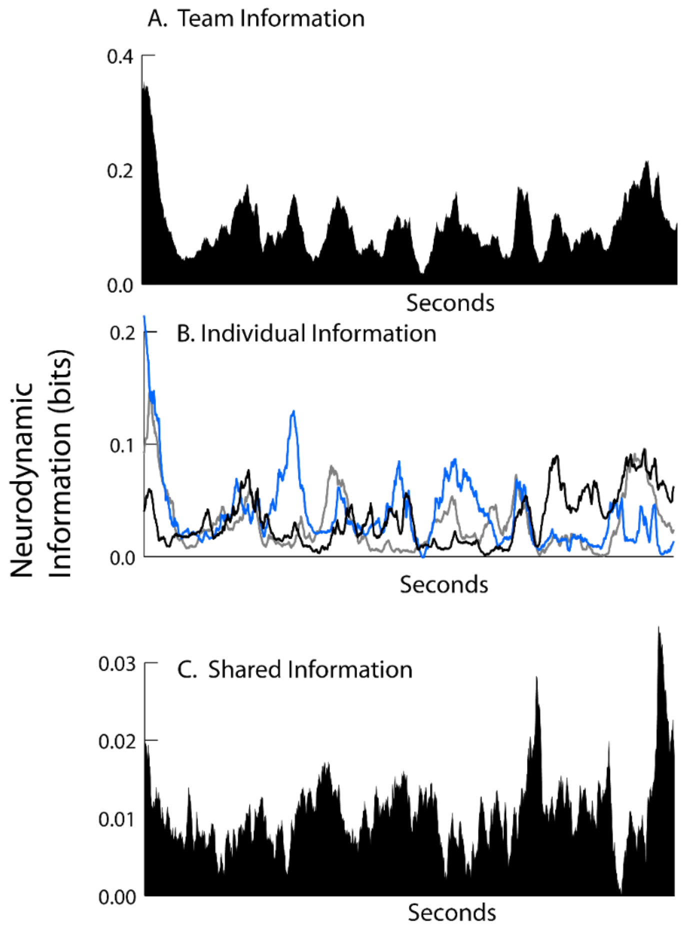


Figure 3. Quantitative comparison of team (A) and the individual (B) information of a three-person healthcare team. C. Sum of the shared information among the three team members.

Example 2. The On-line and Long-Memory Measures of Communication

To differing degrees, teams plan, think, decide, and act. Action-based teams coordinate across real-time perception-action links whereas decision-making teams coordinate across more cognitive, planning links

(DeChurch & Mesmer-Magnus, 2010). Many military, industry, and medical teams employ a combination of action-based and decision-making coordination. In this context, team cognition is the cognition that happens while team members are coordinating and interacting (Cooke, Gorman, Myers, & Duran, 2013). In this section, we focus on the fundamental medium of team cognition, namely team communication (Cooke, Gorman, & Kiekel, 2008). Specifically, we focus on two aspects of team cognition from this viewpoint: (a) On-line team cognition as measured through the real-time communication response of a team to events in the team environments and (b) Long-memory in team cognition as measured through the coherence of team communication on timescales beyond the capacity limits of individual human memory. For both of these aspects, we suggest potential links to assessment and remediation for team ITS.

On-line team cognition is based on the idea that "intelligent" teams can generate an effective response to environmental change in real time. A team with good on-line team cognition can communicate to assess the current situation and produce a flexible and adaptive response to the changing environment. This often requires a generative response that is perhaps similar to, but not identical with, responses used in the past. Put differently, a team with good online team cognition can have consistent behavior in similar (routine) situations but is good at changing its actions rapidly and appropriately as the situation requires (i.e., in novel, non-routine situations). From this perspective, on-line team cognition can be assessed and remediated through real-time team communication analysis (Gorman, Hessler, Amazeen, et al., 2012; Grimm, Gorman, Stevens et al., 2017). In this way, on-line team cognition occurs in the Cognitive band within Newell's timescales of human actions (Newell, 1990).

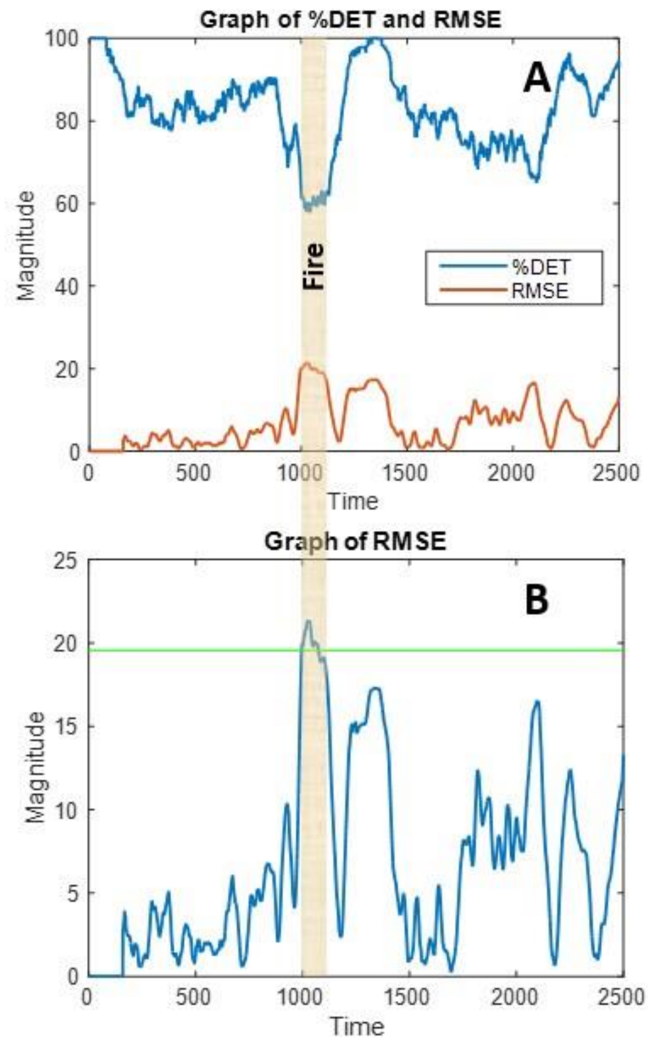


Figure 4. (a) Communication determinism (%DET) and root mean square error (RMSE) from the prediction model; (b) RMSE relative to a 99% confidence interval (green line) indicates a significant shift in communication (drop in %DET) in response to a fire in the OR (reprinted with permission from Gorman, Dunbar, Grimm, & Gipson, 2017).

Real-time communication analysis need not actually be performed as the team task unfolds, though it can be with the benefit of detecting changes and/or anomalies in team cognition in real time (Gorman et al., 2012). However, the essential feature of real-time, on-line team cognition is that we are measuring changes in team communication at sampling rate that is approximately the same as the characteristic timescale of team activity. Figure 4 shows a realtime analysis of team cognition in a simulated medical task (Stevens, Galloway, Gorman et al, 2016) sampled at 1 Hz, highlighting a significant environmental event (e.g., fire outbreak in the operating room), as measured by the predictability of their communication pattern over time (higher %DET corresponds to more predictable pattern). The graph of RMSE shows how quickly and adaptively the team responded to the event. The spike in RMSE corresponding to the fire indicates good on-line team cognition, whereas the spike just after the fire indicates a quick return to predictability. For team ITS, assessment of team communication at this timescale of analysis might provide guidance on the quality of on-line team cognition versus the need to remediate through interactive exercises aimed at quickly responding to and recovering from unexpected environmental events.

On relatively long timescales (days, weeks, months, years), team communication exhibits long-memory (Beran, 1994). In team cognition, long-memory is a systems-level memory that is not stored inside any individual team member that nevertheless informs team members' ongoing interpretations and productions of communication behavior. Rather than manifesting in the individual, long-memory manifests in the history of communication interactions in a particular team task. An example is the use of navigation language in the submarine navigation task (Gorman, Dunbar, Martin et al., 2016). Navigation language is not contained in any one individual, but is distributed across team members as they simultaneously absorb the existing language and imbue it with their own idiosyncratic properties unique to their team. In turn, subsequent crews will inculcate their language and evolve it for their own use. In this way, submarine navigation language has its own long memory separate from any individual. Moreover, unlike common mechanisms of human memory, long-memory structures communication patterns among existing and new team members' communication beyond the spans of human memory constraints.

Studies have demonstrated how long-memory develops in teams (see Gorman et al., 2017 for a review). In one study, Latent Semantic Analysis cosine (i.e, coherence or knowledge relatedness; Landauer, Foltz, & Laham, 1998) diminishes as the timescale (distance between utterances) was increased for different medical teams (these teams were described in a study by Stevens et al., 2016). The novice medical team had a shorter timescale of coherence (their communication had a "shorter memory" over about 15 utterances); by contrast, the experienced team had a longer timescale of coherence (their conversation had a "longer memory" over about 31 utterances). This between-team cross section demonstrates how long memory develops with experience. In another study by Gorman (2005), the LSA cosine method was used to show long-memory develops within uninhabited air vehicle (UAV) teams. Specifically, long-memory in UAV team communication increased over a series of missions.

Assessment Triggers: How do you know it is time to intervene?

Assuming that meaningful abstractions compressions and representations can be developed using comparable quantitative measurement scales, the next question is how to use them. Intelligent team tutoring systems (ITTS) must assess the current state of the team in real-time to best support learning. An ITTS, however, needs to decide when and where to intervene. Traditional ITS provide one-to-one feedback to enable learning by modifying instructional content, timing, and teaching strategies (Wenger, 1987; Murray, 2003; Koedinger & Tanner, 2013; Gilbert, Blessing, & Guo, 2015). Strategies include changing task difficulty (Harley, Lajoie, Frasson, & Hall, 2015), adjusting timing and difficulty of assessments (Arroyo, Woolf, Burelson et al., 2014), or providing additional examples and hints (Chaffar, Derbali, & Frasson, 2009; Woolf, Burleson, Arroyo et al, 2009).

In the domain of adaptive automation, assessments made in real-time have been referred to as *triggering mechanisms*, and form the basis of the assessment-intervention loop (Feigh, Dorneich, & Hayes, 2012). For instance, in traditional ITSs, feedback to the user has been triggered by conditions such as task performances or actions that indicate a learner misconception. Triggers are based on information that can be sensed, observed, or modeled to develop an understanding of context. The ability to answer the question, "How is the team doing?" remains limited because of the difficulty in assessing context; for example, assessment methods have often relied on static task models and user performance to gauge learner state indirectly. If an ITTS decides to intervene in the learning process, it will need triggers to identify when to engage an intervention, how long an intervention should persist, and when to disengage the intervention.

The field of adaptive automation provides a taxonomy (Feigh et al., 2012) that can be modified to classify ITTS assessment triggers into six broad categories: learner, system, environment, task, spatiotemporal, and contextual.

Learner-based triggers

ITS interventions can be triggered by the learner directly or by a system assessment of the learner state. Learner-based triggers can consist of both active and passive assessments. The system could simply ask the learner for their current state. The "current state" of the learner could be the state of the team and individual knowledge, skill, emotions, or physical state. More typically, traditional ITSs employ a range of assessment techniques to measure the current state of the learner. This could be through performance on the task, looking for the presence of absence of expected behaviors, or other observable demonstrations of understanding.

System-based triggers

Current or predicted states of the system can be used to trigger interventions. System-state triggers can consist of a system model that can include its structure, modes, internal states, anticipated future states, and range of potential actions (Feigh et al., 2012). For instance, a teaching progression may be in one of several modes (training, practice, test). Each mode may correspond to a set of specific system behaviors (Johnson, 1990).

Environment-based triggers

States of the environment or events external to the learner and the system can be used to trigger interventions. Environment-based triggers represent the relevant aspects of the world outside the immediate system and operator. In a team tutor-based training, this might include ambient conditions (e.g. light level) or external events that occur in the environment.

Task- and goal-based triggers

A learning session is typically composed of a coherent set of goals and sub-goals and accomplished by a set of tasks. Triggers can be based on task state or goal state. Goal-based triggers could be assessed by comparing current actions to expected actions based on a knowledge of the learning goals of the training. Learning may take place in phases, where one phase builds upon the previous one, forming a plan that can be used as the baseline for an assessment of progress through the training. Task state is the initialization, progress, and completion of tasks.

Spatiotemporal triggers

Both time and location can be used as adaptation triggers. Time-based triggers in ITS may be as straightforward as completion of the task in the allotted time. Location triggers can use absolute or relative locations to trigger interventions.

Contextual triggers

Patterns of information involving objects, concepts, and relationships across space and time can represent context-cues that can trigger ITTS interventions. Contextual triggers in an ITTS indicate places in the evolution of problem where multiple dimensions of past and present information are combined to form a context cue to which one or more team members should detect and respond. Contextual triggers identify opportunities to assess whether an important context pattern has been detected (based on recognitional learning) and translated into appropriate action (based on decision learning).

Example 3. Feedback-generated enhancement of team knowledge

The final example approaches the question 'How is this team doing?" starting with the assumption that an intervention has been developed that is delivered by a triggering process. An important question is how long instructors need to wait to determine if the intervention is having an effect?

Grand, Braun, Kuljamin, Kozolowski & Chao (2016) recently reported results from their study of the dynamic team cognition processes that take place within and between individuals to generate and sustain team knowledge. The study collected behaviors that related to learning the team task and sharing information about it, and examined the emergence of team knowledge outcomes. An experiment was conducted with 263 3-person teams in which they participated in a computer simulation of a naval task for crisis relief. Each team participated in a 2 hour session that included task training familiarization and then twelve 8-minute trials of task performance. The task had team members post their communications on an electronic information board they shared with other team members. While team members had some knowledge overlap, no team member could complete the task without sharing information with the others. Experimental condition teams were given 10 different types of automated visual prompts (contextualized guidance) that were triggered by behavioral errors or inefficiencies detected by the simulator. The display prompts were started at the third trial and provided feedback and recommendations to individual members for improving their performance (e.g., feedback on an incorrect post was triggered when a member posted information to the shared board that did not exist). Display prompts were stopped at trial 10. A log file containing every task relevant behavior performed by each individual was recorded for all trials. Team knowledge emergence was determined from behaviors that were automatically categorized and timestamped, making it possible to track when, what and how knowledge was being acquired and distributed during each trial.

Results showed that the experimental teams were more efficient at generating collectively held knowledge compared to control condition teams. A large portion of knowledge in the control condition teams was not shared because of inefficient communication processes that prevented information from reaching all three members. To explain this, Grand et al. found that experimental condition teams achieved total team knowledge coverage earlier than the control condition team. The control condition information exchanges flattened out at about the halfway point in the trials, whereas the experimental condition information exchanges continued to increase. However, it wasn't until the 9^{th} trial that the experimental condition's information exchanges were significantly greater in number than the control condition exchanges. Grand et al concluded that when teams need to rapidly build knowledge, even a single member who cannot keep up with the team can slow the dissemination and accumulation of collectively held knowledge. Grand et al recommended that for teams with a commonly shared goal and whose members are specialized, every effort should be made to promote knowledge sharing by all individuals as early and often as possible.

These results have positive implications for team tutor development. First, as long as information exchange processes for accomplishing the team task are the target, scaffolding can target individual team members with feedback during the task. Second, since improved performance in the Grand et al study was achieved within 9 trials (about 72 minutes), training simulation scenarios could potentially be segmented in such a way to achieve learning objectives much more quickly using a scaffolding approach than the typical 4 to 6 hour approach using post scenario AARs.

Conclusions and Recommendations for FUTURE research

As illustrated in this chapter there are many challenges to developing practical implementations to answer the question posed in this chapter. These challenges arise from the need to accurately perceive the team's cognitive state and to project forward to what the team will need given the evolving context of the task. Our ability to provide rapid and effective feedback during team training therefore depends on how well we can make sense of the parallel and complex information streams that are increasingly being generated about the team, team members and the environment. Another challenge is the lack of practical examples showing what the interactions would be between commonly described team behaviors and the underlying microdynamics of cognition. This last section describes first efforts at developing such practical models by providing evidence for neurodynamic and speech entrainment during simulation debriefings (Stevens, Willemsen-Dunlap, Gorman, et al., 2018).

Neural and Speech Entrainment during After Action Reviews

The scenario part of simulation training has received the most attention as it is here the planning and action phases of experiential learning occur. There have been fewer studies examining team dynamics during the debriefing. It is in this training segment that team members are supported by facilitators in developing a reflective understanding of their own and the team's performance (Fanning & Gaba, 2007; Husebo, Dieckman, Rystedt,et al., 2013).

There is general agreement that some form of structure is needed during the debriefing process, but aside for general guidance for sequencing topics (Sawyer, & Deering, 2013; Cheng, Morse, Rudolph, Arab, Runnacles & Eppich, 2016) there have been few examples of what that structure looks like from the behavioral, communication and neural perspectives. There is a sense in the research community that more dynamic models and theories of debriefing are needed (Fanning & Gaba, 2007; Bowe, Johnson & Puscas, 2017). As described earlier, neurodynamic and communication models might help bridge this gap by showing how information is passed across systems and temporal scales.

As described previously (Stevens & Galloway, 2017; Stevens, Galloway & Willemsen Dunlap, 2017), the neurodynamic organizations of teams can be decomposed into the neurodynamic organizations of each team members plus the information that is shared among team members. This allows comparisons to be made between the different topic discussions (macro-scale dynamics) speech determinism (meso-scale dynamics) and individual and team neurodynamic organizations (micro-scale dynamics) during the debriefing.

In the example shown in Fig.5B there were five major discussion topics during the debriefing: 1) Why was the patient experiencing seizure? 2) The possibility of obstructions during intubation; 3) Relating stories of the use of anesthesia in the operating room; 4) Team situation awareness; and, 5) Fidelity of the simulation / room setting. Within each of these segments there was increased %DET suggesting coherence between speech dynamics and the structure of event discussions. There were also periods of elevated team neurodynamic organization within each segment resulting in a positive correlation between %DET and team *NO* ($r = .48$, $p < .005$) (Fig. 5C). The positive correlation indicates that team neurodynamic organization was greatest when few team members were speaking.

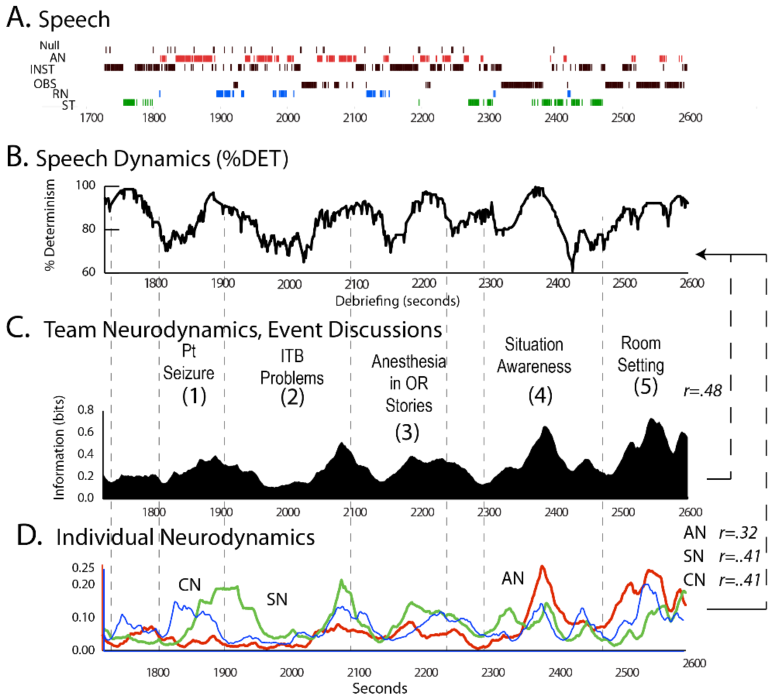


Figure 5. Multi-modal dynamics of a team debriefing event. A) This plots the second-by-second speech of each member of the team. B) This shows the % determinism of the speech. C) The five main discussion topics are shown along with the profile of the team neurodynamics. D) The individual neurodynamics of the Anesthesiologist (AN), Circulating Nurse (CN), and Scrub Nurse (SN) are plotted in red, blue and green respectively. The correlations of the individual and team neurodynamic data streams with the % DET are shown to the right.

Each team member showed different levels of neurodynamic organization during the discussion topics (Fig. 5D). For instance the Anesthesiologist (AN) showed little *NO* except during the Situation Awareness discussion (~2400s) while the Circulating Nurse (CN) and the Scrub Nurse (SN) had earlier periods of *NO* as well as at this time. The neurodynamic organization of each of the team members (Fig. 5D) were moderately correlated with %DET.

These data illustrate how representations like neurodynamic and communication organization can be flexibly used to visualize the structures during debriefing and develop a better understanding of what different information means in the context of teamwork. The quantitative nature of these organization-information analyses will allow direct comparisons between teams and different debriefing protocols; i.e. self-debrief, facilitated, video review, co-debriefing, etc. which will provide guidance for how debriefings can best be included in the GIFT environment.

It is tempting to speculate that in five years or so adaptive agents integrated into the GIFT platform will be periodically asking the question ‘How is this team doing?’ and synthesizing answers using libraries containing the temporal dynamics of neural, physiological, speech characteristics of teams, team members and contextual cues based on examples like the one shown in Fig. 5.

The development pathway toward these agents will most likely begin using only dynamical changes in measures like %DET that might be associated with the emergence of leadership (Gorman et al, 2017), or neurodynamic entropy decreases associated with team stress and uncertainty (Stevens & Galloway, 2015;

2017). The usefulness of the dynamics is illustrated in Fig. 6 which annotates the team neurodynamic entropy profile shown in Fig. 1. The three major structures for assessment are the Magnitude (Fig. 6A), Duration (Fig. 6B) and Frequency (Fig. 6C) of the decreases in entropy. The fewer, shorter and smaller entropy decreases there are during simulation training the better the team performance (Stevens, Galloway, Lamb, Steed, Lamb, 2017; Stevens & Galloway, 2017). The idea is that experienced teams dwell less on details of team taskwork, and communicate situation understanding more rapidly than less efficient teams.

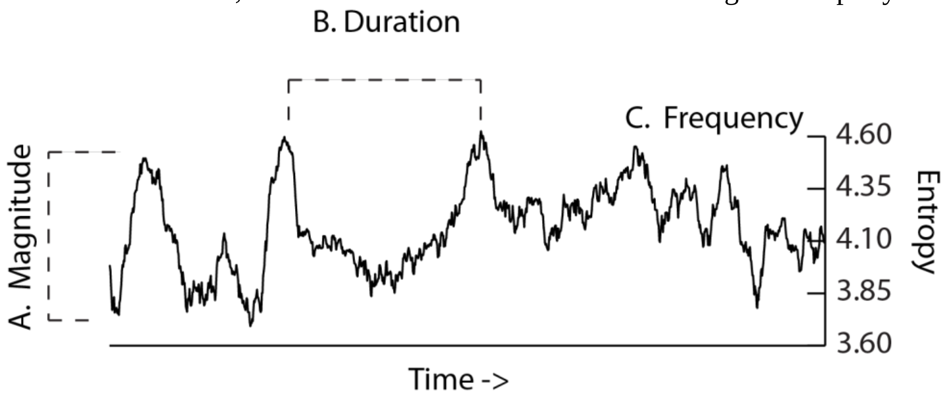


Figure 6. Dynamical features of neurodynamic entropy fluctuations (Stevens & Galloway, 2015).

An interesting feature of neurodynamic entropy decreases is that their magnitude is often proportional to the duration of the decrease suggesting it may be possible to predict how long the team will remain in a highly organized state from the magnitude of the entropy decrease. While this figure illustrates dynamical metrics for team neurodynamic organizations, it would equally apply to the %DET dynamics or heart rate variability, or most other data stream dynamics that have an information-organization basis.

Setting assessment triggers to different change magnitudes, durations and frequencies would be one pathway for insertion of team dynamics into the GIFT environment. More advanced triggers are also possible and would likely be based around the structural bases responsible for the dynamical traces, whether they be the dynamics resulting from structures associated with EEG frequency / sensor combinations, or enhanced %DET measures based on semantic subsets of words. Here the involvement of phenomena like entrainment will be an important supporting component as entrainments represent repeating and persistent structures with possible memory components. These structures within the neurodynamic and communication data streams may be useful therefore for longer-term prediction of team function.

References

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

Beran, J. (1994). *Statistics for long-memory processes* (Vol. 61). New York, NY: Chapman & Hall.

Cannon-Bowers, J. and Salas, E. (1998). *Making Decisions under Stress.* Washington, D. C.: Am. Psych. Assn.

Bowe, S., Johnson, K. & Puscas, L. (2017). Facilitation and Debriefing in Simulation Education in *Otolaryngologic Clinics of North America, Vol. 50, Issue 5*, 989-1001.

Chaffar, S., Derbali, L., & Frasson, C. (2009). Inducing positive emotional state in Intelligent Tutoring Systems. In AIED (Vol. 2009, No. 200, pp. 716-718).

Cheng, A., Morse, K., Rudolph, J., Arab, A., Runnacles, J., & Eppich, W. (2016). Learner-centered debriefing for health care simulation education: Lessons for faculty development. *Simul Healthc, 11*, 32-40.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science, 37,* 255-285. doi: 10.1111/cogs.12009.

Cooke, N. J., Gorman, J. C., & Kiekel, P. A. (2008). Communication as team-level cognitive processing. In M. Letsky, N. Warner, S. Fiore & CAP Smith (Eds.), *Macrocognition in teams (pp. 51-64).* Burlington, VT: Ashgate.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95,* 32-53. doi: 10.1037/a0017328

Fanning, R., & Gaba, D. (2007) The role of debriefing in simulation-based learning. *Simul Healthc, 2,* 115-125.

Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors, 54(6), 10081024.

Fiore, S. M., Smith-Jentsch, K. A., Salas, E., Warner, N., & Letsky. M. (2010). Towards an understanding of macro cognition in teams: Developing and defining complex collaborative process and product. *Theoretical Issues in Ergonomics Science 11(4),* 250-271.

Fischbach, K., Gloor, P. A., Lassenius, C., Olguin, D., Pentland, A. S., Putzke, J., & Schoder, D. (2009). Analyzing the flow of knowledge with sociometric badges. *Procedia- Social and Behavioral Sciences, 2,* 6389-6397.

Flack, J. C. (2012). Multiple time-scales and the developmental dynamics of social systems. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367,* 1802–1810. doi:10.1098/rstb.2011.0214.

Foltz, P. W., Laham, D. & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. *Proceedings of EdMedia* '99.

Gilbert, S. B., Blessing, S. B., & Guo, E. (2015). Authoring Effective Embedded Tutors: An Overview of the Extensible Problem Specific Tutor (xPST) System. International Journal of Artificial Intelligence in Education, 1-27.

Gilbert, S., Slavina, A., Sinatra, A. M., Bonner, D., Johnston, J., Holub, J., MacAllister, A., Dorneich, M., Winer, E. (2017). Creating a Team Tutor Using GIFT. International *Journal for Artificial Intelligence in Education*.

Gorman, J. C. (2005). The concept of long memory in assessing the global effects of augmented team cognition. In *Proceedings of the 11th International Conference on Human-Computer Interaction* (pp. 22-27).

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modeling teams as dynamical systems. *Frontiers in Psychology, 8,* 1053.

Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Detecting perturbations to team communication. *Ergonomics, 55,* 825839. doi: 10.1080/00140139.2012.679317

Gorman, J. C., Martin, M. J., Dunbar, T. A., Stevens, R. H., Galloway, T. L., Amazeen, P. G., & Likens, A. D. (2016). Cross-level effects between neurophysiology and communication during team training. *Human Factors, 58,* 181-199. doi: 10.1177/0018720815602575

Grand, J. A., Braun, M. T., Kuljanin, G., Kozlowski, S. W., & Chao, G. T. (2016). The dynamics of team cognition: A process-oriented theory of knowledge emergence in teams. *Journal of Applied Psychology, 101*(10), 1353.

Grimm, D., Gorman, J. C., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2017). Demonstration of a method for real-time detection of anomalies in team communication. *Proceedings of the Human Factors and Ergonomics Society 61st Annual Meeting.*

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2015, June). An integrated emotion-aware framework for intelligent tutoring systems. In International Conference on Artificial Intelligence in Education (pp. 616-619). Springer International Publishing.

Husebo, S. E., Dieckman, P., Rystedt, H., Soreide, E., & Friberg, F. (2013). The relationship between facilitators' questions and the level of reflection in postsimulation debriefing. *Simul. Healthc. 8,* 135-141.

Johnson, J. (1990). Modes in non-computer devices. International Journal of Man-Machine Studies, 32, 423–438.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284. doi: 10.1080/01638539809545028

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In Authoring tools for advanced technology learning environments (pp. 491-544). Springer Netherlands.
Radtke, P., Johnston, J., Biddle, E., & Carolan, T. (2007). Integrating and presenting performance information in Simulation Based Warfare Scenarios. In the Proceedings of I/ITSEC.
Salas, E., Benishek, L., Coultas, C., Dietz, A., Grossman, R., Lazzara, E., & Oglesby, J. (2015). *Team training essentials: A research-based guide*. New York, NY: Routledge.
Salas, E., Diaz-Granados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors*, *50*(6), 903-933.
Sawyer, T. L., & Deering, S. (2013). Adaptation of the US Army's after-action review for simulation debriefing in healthcare. *Simul in Healthc*, *8*, 388-397.
Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction impacts on team mental models, processes, and effectiveness. Small Group Research, 39(3), 303-327.
Sottilare, R.A., Burke, C. Shawn, Salas, E., Sinatra, A., Johnston, J. & Gilbert, S. Designing adaptive instruction for teams: A meta-analysis. In: *International Journal of Artificial Intelligence in Education* (2017): 1-40.
Stevens, R., Galloway, T., Gorman, J., Willemsen-Dunlap, A., Halpin, D. (2016, July). *Toward objective measures of team dynamics during healthcare simulation training.* The International Symposium on Human Factors and Ergonomics in Health Care, Orlando, FL. doi: 10.1177/2327857916051010
Stevens, R.H. & Galloway, T. (2014). Toward a quantitative description of the neurodynamic organizations of teams. *Social Neuroscience 9(2),* 160-173.
Stevens, R. H., & Galloway, T. (2015). Modeling the neurodynamic organizations and interactions of teams. *Social Neuroscience 11*, 123-139. doi: 10: 1080/17470919.2015.1056883.
Stevens, R. H., & Galloway, T. (2017). Are neurodynamic organizations a fundamental property of teamwork? *Frontiers in Psychology*, May, 2017. doi: 10.3389/fpsyg.2017.00644.
Stevens, R. H., Galloway, T. & Willemsen-Dunlap, A. (2017). A team's neurodynamic organization is more than the sum of its members. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Vol 61, Issue 1, pp. 2010 - 2014
Stevens, R., Galloway, T., Lamb, J., Steed, R., & Lamb, C. (2017). Linking team neurodynamic organizations with observational ratings of team performance. In *Innovative Assessment of Collaboration* (pp. 315-330). A. A. Von Davier, P. C. Kyllonen, & M. Zhu eds. Springer International Publishing, Cham, Switzerland.
Stevens, R. H., Galloway, T., Halpin, D., & Willemsen-Dunlap, A. (2016). Healthcare teams neurodyamically reorganize when resolving uncertainty. *Entropy*, *18*: 427, doi: 10.3390/e18120427
Stevens, R., Willemsen-Dunlap, A., Gorman, J., Galloway, T., Halpin, D., & Grimm, D. (2018). Neurodynamic and communication analysis of healthcare teams during simulation debriefings. *Proceedings of the 2018 International Symposium on Human Factors and Ergonomics in Healthcare, Boston, MA.*
VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46(4)*, 197-221.
Wood, H., & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers & Education, 33*, 153-169.
Wenger, E. (1987). Artificial Intelligence and Tutoring Systems. Los Altos, CA: Morgan Kaufmann.
Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognizing and responding to student affect. International Journal of Learning Technology, 4(3-4), 129-164.
Zachary, W., Rosoff, A., Miller, L., and Read, S. (2013). Context as Cognitive Process: An integrative Framework for Supporting Decision Making. In Laskey, K., Emmons, I., and Costa, P. (Eds). *Proceedings of 2013 Semantic Technologies in Intelligence, Defense and Security.*CEUR Conf Proceeding Vol-1097. Pp48-55 http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-1097/STIDS2013_T07_ZacharyEtAl.pdf
Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., & Burns, J. (1999). The Advanced Embedded Training System (AETS): An intelligent embedded tutoring system for tactical team training. *International Journal of Artificial Intelligence in Education, 10*, 257-277.

Chapter 9 – Assessment of Team Performance in Psychomotor Domains

Anne M. Sinatra[1], Jong W. Kim[1,2], Joan H. Johnston[1], Robert A. Sottilare[1]
US Army Research Laboratory[1]; Oak Ridge Associated Universities[2]

Introduction

Implementing team tutoring functionality in a domain-independent intelligent tutoring system framework is a difficult challenge. There needs to be considerations that go into keeping the authoring tools and features that are available as flexible as possible because they will be used in many different domains or topic areas. Some domains may have more traditional computer based instruction. However, there are many others that require physical motions or manipulation of objects. Therefore, there needs to be planning to have sensors and mechanisms for measuring the actions that are taken by an individual in a psychomotor task in order to be able to tutor them. However, input and measurements are not the only considerations, there also has to be a plan in place about how information will be provided to the individual (e.g., through a mobile device, through audio, through a computer present in the room with them). While some of the details and considerations of an individual psychomotor task has been previously discussed (Kim, Sottilare, Goodwin & Hu, 2017), the current chapter goes one step further into the area of *team psychomotor tutoring*. Every challenge that existed with individual psychomotor tutoring not only still exists in the team version, but there are now many more. There are many types of teams, skills, roles, and configurations that need to be taken into account. The division of teamwork and taskwork needs to be thought through, and assessment and feedback now is at two levels: the individual and the team. Additionally, team feedback is more than just adding together individual feedback. The approach to grading team performance is going to be highly dependent on the characteristics of the domain, and the configurations of the teams. Further, there are technological challenges that involve not only having multiple computers communicate with each other to gauge what the team is doing, but also dealing with the coordination of team sensors and the real-time processing of data from multiple learners at the same time.

The current chapter discusses the challenges and goals associated with developing psychomotor tutors for teams. The discussion is framed in terms of the development of the flexible domain-independent intelligent tutoring system framework, the Generalized Intelligent Framework for Tutoring (GIFT). GIFT has been used to create tutors in individual psychomotor domains such as marksmanship (Goldberg & Amburn, 2015; Goldberg, Amburn, Ragusa & Chen, 2017) and golfing (Kim, Dancy, Goldberg & Sottilare, 2017). Further, team tutoring has been demonstrated in GIFT in a computer-based game task (Bonner et al., 2016; Gilbert et al., 2017). Many considerations go into the design process to ensure that GIFT can support not only team tutoring, but team tutoring in different domains, configurations, and types of tasks (e.g., computer-based, psychomotor). In this chapter we discuss the measures needed to assess performance in team tasks, methods associated with psychomotor team assessment, and ideas that should be taken into consideration as team tutoring elements continue to be incorporated into GIFT. In order to highlight the complexity that is involved in individual domains, we discuss the psychomotor task of room clearing and the requirements of a tutor for this domain.

Team Tutoring in GIFT

GIFT is a domain-independent framework for creating intelligent tutoring systems (Sottilare, Brawner, Sinatra & Johnston, 2017). It includes tools that allow individuals without a background in computer science to create adaptive tutoring courses. A subject matter expert (SME) can bring his or her own existing instructional materials and develop tutors and assessments. Traditional intelligent tutoring systems have a learner module, pedagogical module, domain module, and tutor-user interface (Sottilare, Graesser, Hu & Holden, 2013). Traditionally, information about the learner and his or her performance is in the learner model and module. The pedagogical model and module stores strategies that can be implemented based on performance. The domain model and module stores information that is specific to the domain that is being instructed, whereas the tutor-user interface allows the learner to interact with the system. GIFT contains these components, but it additionally has a sensor module and a gateway module. The sensor module allows for sensor data to be processed and to influence the performance of the learner in the system. The gateway module allows the GIFT software to communicate with external programs or simulations, such as Virtual Battlespace 3 or PowerPoint. Since GIFT is domain independent, the only domain dependent content that exists is within the domain module. This allows for reuse of the other modules and configurations.

In the current state, GIFT is configured to create tutors for individual learners. One of the ultimate goals of GIFT is to be able to author and provide team tutoring. Initial theoretical work has been completed to move GIFT toward this goal (Sottilare, Burke, Salas, Sinatra, Johnston & Gilbert, 2017). Team tutoring scenarios have initially been implemented in GIFT (Gilbert et al., 2017), and work continues to be completed in order to further expand the team functionality (McCormack, Kilcullen, Sinatra, Brown, & Beaubien, in press), and ultimately the GIFT architecture. While computer-based team tutoring is a goal in the evolution of GIFT, it is also important to be able to provide a tutor for psychomotor tasks that require movement in a real world environment.

Psychomotor Domain: Room Clearing Task

A team's execution of coordinated movement is critical to completing complex tasks in such high performing environments as military missions, team sports, commercial aviation, and medical care. Individual psychomotor skills are critical to performing these coordinated moves, but excellent individual performance may not mitigate the poor performance of an entire team. Conversely, an entire team can perform very well even with a couple of low performers. This is because successful team performance is contingent on team members having a shared understanding of the team task. Team members should understand their roles in performing the task, how individual poor performance affects the entire team's performance, and how good team coordination can compensate for an individual team member's poor performance (Salas, Benishek, Coultas, et al., 2015).

A guiding principle of effective team training is to ensure individual skills are developed first, then train and develop *team* task work skills before developing teamwork skills (Salas et al., 2015). Team members need to understand and know how to perform expected task-based behaviors before they can understand how teamwork skills are critical to performing the task.

Conducting effective team training involves planning, training development and implementation, and scenario adaptation. The planning phase first involves identifying skills requirements using an organization's bona fide mission task list. Team skills requirements are used to identify learning requirements, training objectives, and simulation training strategies and technologies.

The next step involves conducting an Event Based Approach to Training (EBAT) (Salas et al., 2015). The EBAT method develops simulation scenarios with critical events that focus on developing the identified skills in the learning objectives. The approach is to first create a scenario set to develop single team skill areas. The next step is establishing team skill proficiency levels and using this information to create assessment tools, such as a behavioral checklists, for the team process behaviors (physical actions and communications) that represent the skills that are expected to occur during the pre-identified critical scenario events.

During the training phase, observers/instructors assess the team's performance on the skills throughout the simulation scenario. Then this information is processed and used during the after action review (AAR). The AAR is conducted with a form of tutorial dialog called "team self-correction" in which instructors facilitate members in discussing the impact of their process behaviors (actions) on mission success.

During the adaptation phase, team outcome measures are developed and used to set goals for the next training session. Teams may be required to repeat the same scenario if they do not meet the minimal proficiency requirements. When teams meet the minimum proficiency, then scenarios are modified to train new skill areas, or are made more complex to focus on skills integration. Training scenarios are strategically designed to enable teams to systematically build capacity to simultaneously execute the skills together. This can involve gradually increasing typical scenario task stressors (e.g., workload and time pressure), which provides the opportunity for teams to learn how to effectively employ their complete skillset. The AAR is then used to review and improve on integrating the multiple skill areas.

In this chapter we employ a use case analysis to illustrate how to identify psychomotor skills requirements at the team level and recommend team performance assessments. We then discuss the implications of these methods to provide recommendations on how the GIFT environment can support adaptive team training.

Military tasks involve performing specific Tactics, Techniques and Procedures (TTPs) through physical maneuvers and communications that enable effective tactics execution. They are comprised of four main competency areas: tactical maneuver, decision making, teamwork, and resilience. A good example of a common military task that relies heavily on team member psychomotor skills is "room clearing" (U.S. Department of the Army, 10 June 2011, Army Tactics, Techniques and Procedures – ATTP 3-06.11). It involves seizing control of a room and its inhabitants (both hostile and other) by rapidly and methodically eliminating the enemy, dominating the room, and controlling the situation. The five task segments are: 1) Prepare to Enter, 2) Enter the Room, 3) Clear the Room, 4) Secure the Room, and 5) Completion. The ATTP 3-06.11 indicates that for the mission to be effective, teams must effectively perform coordinated maneuvers and team communications in order to enable rapid identification and engagement of threats according to the rules of engagement, and to provide a safe haven for other non-threat personnel.

To illustrate, we focus on the "Clear the Room" task segment which involves a dismounted squad's fire team entering a room quickly and moving immediately to a point of domination (POD). While individual performance is critical to completing this task, task success – room domination – can only be achieved by the team. It is a continuous, coordinated, and efficient movement of all four members; they do not stop until reaching their respective POD. The two types of room clearing techniques are the "opposing corner" technique (used when Soldiers are experienced and the team has worked together) and the "strong wall technique" (used when Soldiers are inexperienced, integrating new team members, or when working with developing foreign forces) (see Figure 1, steps 1 through 8).

For this illustration, we elaborate on the "strong wall technique." Figure 1, steps 1 through 8, shows the action sequence with arrows. One at a time, each Soldier enters the room and then turns, moves and scans their weapon in their designated sector, and moves to their respective POD with their back against the wall through which they had entered, and continuing to scan their sector.

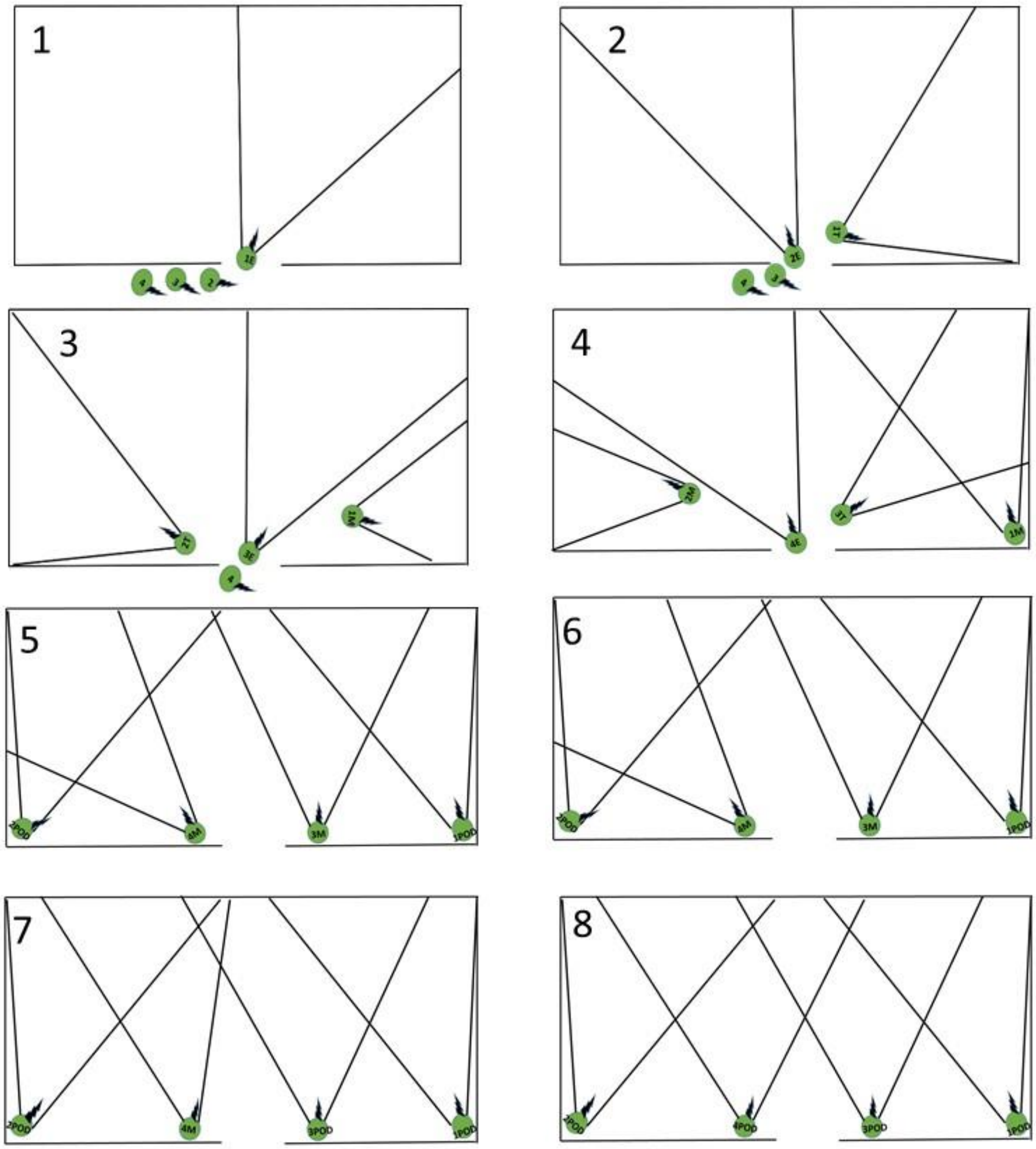


Figure 1. Eight Steps Illustrating the Strong Wall Technique for Room Clearing.

Step 1: Soldier 1 enters the room.

Step 2: Soldier 1 turns to the right, scanning her sector, while Soldier 2 enters the room.

Step 3: Soldier 1 moves to her POD in the lower right corner of the room, Soldier 2 has turned to the left and scans his sector, and Soldier 3 has entered the room.

Step 4: Soldier 1 has moved to her POD location and continues to scan her sector, Soldier 2 has moved toward his POD, Soldier 3 has turned right and is scanning his sector, and Soldier 4 has entered the room.

Step 5: Soldier 1 is established at the POD with her back to the wall and continues scanning. Soldier 2 has reached his POD location, Soldier 3 has moved toward the POD location and Soldier 4 has moved toward his POD location.

Step 6: Soldier 2 is established at his POD location. Soldier 3 has almost reached completion. Soldier 4 has reached his POD location and continues scanning.

Step 7: All Soldiers are positioned at their PODs as they continue to scan their sectors.

Step 8: The Soldiers have completed the Strong Wall configuration and continue scanning their sectors.

The development of a scoring rubric for a team task is essential, and needs to consider both the individual performance, and the performance of the team. The actions taken to enter the room by different Soldiers can be scored to assess performance. Table 1 is an example scoring sheet that summarizes these actions as good team performance in room clearing, for the four Soldiers (each row represents a different soldier) and for the different parts of the task (each column is a different step in the task). Soldier scores are in the far right column whereas Squad scores for each move in steps 1 through 8 are listed along the bottom row. While Table 1 portrays the ideal expert team performance for execution of room clearing, Table 2 illustrates poor execution and performance. The green fill in Tables 1 and 2 indicates that the Soldiers and Squads completed their actions correctly. However, in Table 2 the percentage of correct turns quickly degrades showing how poor Soldier performance (shown in yellow) at the beginning of the task can quickly propagate through the entire squad.

Table 1: Good Team Performance

Soldier	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Soldier Scores**
1	Enters	Turns	Moves	Moves/ Scans	POD	POD	POD	POD/ Scans	100
2		Enters	Turns	Moves	Moves/ Scans	POD	POD	POD/ Scans	100
3			Enters	Turns	Moves	Moves/ Scans	POD	POD/ Scans	100
4				Enters	Turns	Moves	Moves/Scans	POD/ Scans	100
Squad Scores	100	100	100	100	100	100	100	100	

Table 2: Poor Team Performance

Soldier	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Soldier Scores**
1	Enters	Turns	Moves	Moves/ Scans	POD	POD	POD	POD/ Scans	60
2		Enters	Moves	Moves	Moves	Turns	Moves	Stops	20
3			Enters	Moves	Turns	Moves	Moves	Stops	20
4				Enters	Turns	Moves	Moves/Scans	POD/ Scans	20
Squad Scores	100	100	66	25	0	0	0	0	

Developing a scoring rubric such as the ones in Tables 1 and 2 requires a full understanding of the task, and will require training individuals who will observe the behavior of the team and grade them in real time. Quantifying these actions and getting multiple observers to agree on ratings is difficult, but eventually they are able to learn and understand the task. When using computer-based training there are additional challenges. While a human observer can automatically see what is going on and make judgements, in a computer there needs to be sensors that tell the system what the team members are doing. These sensors need to be carefully selected to ensure that the information that is needed to assess performance is being collected. Assessment rules need to be written for each of the actions that need to be taken, and real-time processing needs to occur by the system to determine if the team members are performing as expected. In addition to writing the rules for the individuals and the teams, sensors often output large amounts of data that need to be processed in real-time. There are also additional technological challenges that relate to ensuring that the data between team members is synchronized and can be assessed at the same time.

Initial Measures and Sensors Required to Assess a Room Clearing Task

In order for an intelligent tutor to assess performance during a room clearing scenario, there will need to be a mechanism for tracking the physical location and movements of the team members, and to see the result of their actions. A cellular phone that includes GPS and an accelerometer may be useful as not only a sensor input to the system, but also as a way to provide individualized feedback to the team based on their behaviors. If there is a message that needs to go out to the team member it can be received on their specific phone. Further, even if performance is only being tracked on the overall level, an AAR can be provided through this means. In order to assess performance in the specific task of room clearing, it will be necessary to know the location and orientation of each team member, and the actions that he or she are taking. If the training was occurring in a computer based simulation this information would be captured through actions on a mouse or key pushes. In a live psychomotor situation, this information will need to occur through sensor inputs, and the use of a glasses based eye tracker may be advantageous. If a sensor could additionally be placed on the glasses to determine which way the individuals' gaze is aimed, it will also provide information that can be used for assessment. The communications that the team members engage in during the session can be used for assessment, and the recording/process of it can be facilitated by having the team members wear headsets. This will allow for the information to be recorded and either processed for content in real time using natural language processing, or after the fact. After all of the sensors, tools, and assessment methods are determined for a tutor, work will need to be completed to make sure that the intelligent tutoring system is able to process the data appropriately and provide relevant feedback and assessments. In general, there are a number of different types of team performance assessments that can occur during a tutoring session. The next section provides initial thoughts about methods for assessing team performance in a psychomotor tutor.

Methods for Assessing Team Task Performance in Psychomotor Domains

Methods of team performance assessment can be generally viewed as qualitative/quantitative, natural/simulated, and summative/formative. Such methods could include using behavioral markers, self-assessment, an event-based (or a scenario-based) assessment, or observations by an expert. An expert instructor would observe a team engaging in taskwork, and make qualitative notes or likert-type ratings on a score card like the one in Tables 1 and 2. A scenario-based assessment would occur in a natural environment, which is an actual real place, such as conducting the room clearing training in an actual building with multiple rooms. This mockup training can be also executed in a simulated environment, which either replicates a real setting or an approximation of one. This methodology can be viewed from the perspectives of a summative and formative assessment. In a summative assessment, team performance is measured after the team finishes a particular unit of the training course. In a formative assessment, performance is measured while the trainees are being instructed and their skill development progress is being evaluated.

In the case of room clearing, it could occur in a natural environment, such as a real room, and require that the learners engaging in the training are instrumented with different sensors. Sensors can include accelerometers and GPS, such that information about the movements of the learners can be tracked. The behaviors and performance of the individuals can be tracked, and then used to determine if performance changes over time for the team members. Additionally, the score cards in Tables 1 and 2 would need to be operationalized within the tutoring system, such that performance could be assessed.

An overarching goal in team performance assessment is to evaluate *performance change* by team members. Performance change, in general, can be assessed by comparing scores from the pre- and post-test in a summative or formative assessment. For instance, in the domain model, all the content and information needed for the tutor can be defined, and the learner model can include measures of the relevant learner characteristics for the specific domain. If the number of knowledge components in the domain model is considerably large, measuring performance by assessment from each knowledge component (and from each team member) would be a challenge because it may not be feasible to isolate the effect of each knowledge component (Martin, Mitrovic, Koedinger, Mathan, 2011).

An alternative way is to assess performance change in a time series manner; a learning curve can be used to assess performance. In Industry, the observed performance change in productivity has been treated as a consequence of growing storage of knowledge, and it is referred to as a learning curve. In Psychology, there has been a considerable amount of research on investigating learning and retention to identify empirical regularities in behavior and to test a theory with that empirical data (Anderson, Fincham, & Douglass, 1999; Card, English & Burr, 1978; Heathcote, Brown & Mewhort, 2000; Newell & Rosenbloom, 1981; Seibel, 1963;). The learning curve can be mathematically expressed as in equation 1:

$$y = ax^{-b} \qquad (1)$$

In the mathematical form of equations 1 and 2, a indiciates a constant for the range of learning, and b indicates the rates of learning (ranging from 0 to 1). If it is closer to 1, it means the learning is fast or very adaptive. The term, x, refers to the number of practice trials. It is a log linear model. It can be considered as the most common and simplest mathematical formula. This equation can be estimated using the ordinary least squares method, whereas the logarithmic transformation of the equation allows regression analysis using the least-squares criterion as seen in equation 2:

$$\ln(y) = \ln(a) - bln(x) + e \qquad (2)$$

It would be worthwhile applying this assessment approach (e.g., the learning curve) for individual assessment to an assessment approach for team members. The time to complete a task, the amount of errors during the performance, and the number of times that the learner uses hints while answering a question can all be the parameters for the equation. Using such parameters, we can plot performance change over the practice trials (by seconds, minutes, days, or months).

In the room clearing example, a team is comprised of a number of individuals. Each individual team member has a different role to take during a sequence of training scenarios. A correct and precise assessment for the team psychomotor performance necessitates an understanding of physical states, physiological states, and cognitive states. These factors would pose significant challenges in collecting behavioral data, assessing the alignment of those behaviors with a model of expert/individual and team behavior, and visualizing performance change for improved learning analytics. Ultimately, the goal is to assess the rate of progress toward the learning and performance objectives from the team perspective.

Performance Assessment Methods from an Individual to a Team

In a precision-required psychomotor task (e.g., rifle marksmanship), a breath control skill (e.g., slow breathing) can be considered as an assessment measure of performance. Other precision-required psychomotor tasks include golf putting, dart throwing, free throw, and archery. It has been reported that breath control skill training would lessen adverse effects caused by stress (Bouchard et al., 2012). A biosensor (such as BioHarness™) can be used to measure the differential size of the expansion and contraction of the thoracic cavity, and to assess a specific breath control technique during the task (Goldberg et al., 2017). A physio-cognitive model in ACT-R/Φ has been proposed to provide a theoretical account of the relationship among physical, physiological, and cognitive factors, and to predict performance. The assumption is that a breath control skill (e.g., slow breathing) can delink memory from physiological arousal and affect physical and cognitive processes in the stages of learning (Kim, Dancy, & Sottilare, in press; Dancy & Kim, in press).

For the room clearing example, a BioHarness™ can provide information about whether the team members are using the appropriate technique to maximize their performance during the task. Also, several in-built smartphone sensors are useful to measure motions (e.g., acceleration by x, y, z axes). Other sensors (e.g., barometer, gyroscope, GPS) can be also used to evaluate performance (e.g., move, walk, crawl). Static and dynamic acceleration values can be used to classify and cluster animal behaviors, i.e., walking, sitting, preying, resting, and so on (Fehlmann et al., 2017). The output from all of these sensors needs to be synchronized into a timestamped data set for an improved learning analytics and adaptive instructions (e.g., Ocumpaugh et al., 2015).

It is now necessary to consider assessment for a squad so there are multiple users with multiple sensors. The individual performance assessment can be applied to an assessment of team performance in the aforementioned room clearing case. When a dismounted squad team enters a room quickly, moves to a point of domination (PoD), and performs a given mission, the team's physiological states (i.e., heart rate and respiratory rate) would affect the cognitive and physical performance.

Team tactical breathing can be taught in a GIFT course and performance can be assessed by sensor integration. The team members can be equipped with a BioHarness™ as a respiratory measure and other sensors as a cognitive or physical measure. Tactical breathing is useful to improve precision and accuracy in a psychomotor task (e.g., rifle marksmanship). These measures are useful to determine the state of the individual learners and of the team members. It is important to understand how team members develop psychomotor skills from a novice to an expert. An unobtrusive assessment is important to assess the rate of progress toward the learning goal and the development of psychomotor skills. It is also necessary to understand how individual accomplishments can support team objectives. The Domain Module in GIFT has a configuration that is called Domain Knowledge File (DKF). For assessment of a team, the GIFT architecture can allow multiple DKFs simultaneously to assess the team performance (for more information, see the chapter by Brawner, Sinatra, & Gilbert, in press).

Challenges in Applying Methods to Assess Team Task Performance in Psychomotor Domains

Providing a construct of team performance measurement is essential to better understand a team. It would be a challenge since assessment can be context specific (e.g., a resuscitation task by a medical team in an emergency room), and therefore, the basic elements for assessment including team communication, collaboration within a team, roles & responsibilities, team conflicts management, would be necessary for such construct (Marriage & Kinnear, 2016). It is important to consider how measures and methods can be administered effectively, and to provide a rationale for the instantiation of training principles in adaptive training systems.

One way to handle the varying domains and tasks is to create a formal generalized model which is a theory-based assumption and hypothesis describing team performance. The models should be precisely developed and altered to the dynamic demands of a given task and situation (Pew & Mavor, 1998). Thus, the methodology for the team performance assessment should include: (a) the model that can computationally and predictably describe the rich spectrum of team learning and performance, and (b) the empirical data that is collected from the field and laboratory, and (c) the training application (e.g., an adaptive instructional system) to which researchers apply scientific theories for further test and evaluation. A more collective computational model that can represent team performance and teamwork is worth exploring, which can advance the tools for assessment.

Current State of Team Tutoring in GIFT and Conderations for Tutoring Teams in Psychomotor Domains

In the current version of GIFT, team tutoring has been demonstrated from a technological perspective by creating a tutor that involves two teammates simultaneously engaging with a surveillance task (Gilbert et al., 2017). The surveillance tutor was further scaled up to include a total of three team members, with the third member performing a different role than the other two. Current work is being conducted in GIFT to expand the size of the team that is being tutored, and creating a more indepth task that elicits team performance measures (McCormack et al., in press). While the foundation for team tutoring has begun to be implemented in GIFT, all of the attempts to date have been in computer-based simulated environments. In GIFT, psychomotor tutoring has been demonstrated at the individual level for the domain of marksmanship (Goldberg et al., 2017). However, team psychomotor tasks have not yet been achieved. Establishing ways to assess team state and team performance have initially been used through leveraging the same tools that exist for individual performance and combining all of the actions of the teammates for a single assessment that is the "team". However, this may not ultimately be the ideal approach to measuring team performance or providing feedback based on it. As GIFT continues to develop, and team models become implemented in GIFT it is important to consider how inputs such as sensor and team communications will be dealt with in real time. For many psychomotor tasks being able to understand what the team member is physically doing and how he or she is oriented will be very important. Additionally, the content and frequency of team communication will be vital in a natural environment.

With regard to individual tutoring of a psychomotor task, it has been acknowledged that the tutoring capability should address the needs beyond the desktop environment. Similarly, with regard to team tutoring for a psychomotor task, it is necessary to consider the process of learning from the perspective of assessment. The fundamental areas of GIFT for team tutoring would include: (a) a capability to author instructional contents for the team, (b) a capability to evaluate the team performance, and (c) a capability to administer team learning analytics for adaptive instructions and feedback. At the same time, if we have a formal model of a team (e.g., team communication) that directly interacts with GIFT, it will provide opportunities to assess team performance by testing theories against team performance data.

Conclusions

In order to conduct team tutoring of a psychomotor domain there are a number of challenges. First, an intelligent tutoring system framework such as GIFT must be able to support assessments from multiple individuals simultaneously. Additionally, there will need to be a way to process and retrieve vital information that is happening in the real world environment, and a way to provide information back to the learner. Ideally, all of this will happen in real time. As demonstrated in this chapter, a team task such as room clearing, while fairly straightforward, results in a need to know the location and movements of each individual on the team and to assess if they are all providing the necessary steps for the task at hand. A different psychomotor task will likely have different requirements, and result in different inputs and assessments. Overall, as GIFT continues to develop in order to support both teams and psychomotor tasks, the authoring tools will need to remain flexible, and allow for course authors to implement various measures of evaluation and feedback to both individuals and entire teams.

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1120.

Best, B. J., & Lebiere, C. (2003). Teamwork, communication, and planning in ACT-R agents engaging in urban combat in virtual environments. In *Proceedings of the 2003 IJCAI workshop on cognitive modeling of agents and multi-agent interactions* (pp. 64-72).

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., ... & Holub, J. (2016, September). The Challenges of Building Intelligent Tutoring Systems for Teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 60, No. 1, pp. 1981-1985). Sage CA: Los Angeles, CA: SAGE Publications.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics*, *21*(8), 601-613.

Dancy, C. L., & Kim, J. W. (accepted). Towards a physio-cognitive model of slow-breathing. In *Proceedings of the 40th Annual Conference of Cognitive Science Society*. Wisconsin, USA: Cognitive Science Society.

Fehlmann, G., O'Riain, M. J., Hopkins, P. W., O'Sullivan, J., Holton, M. D., Shepard, E. L., & King, A. J. (2017). Identification of behaviours from accelerometer data in a wild social primate. *Animal Biotelemetry*, *5*(1), 6.

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

Goldberg, B., & Amburn, C. (2015, August). The application of GIFT in a psychomotor domain of instruction: a marksmanship use case. In *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)* (pp. 115-124).

Goldberg, B., Amburn, C., Ragusa, C., & Chen, D. W. (2017). Modeling expert behavior in support of an adaptive psychomotor training environment: A marksmanship use case. *International Journal of Artificial Intelligence in Education*, 1-31.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic bulletin & review*, *7*(2), 185-207.

Kang, M. (2001). Team-Soar: A computational model for multilevel decision making. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *31*(6), 708-714.

Kim, J. W., Dancy, C., Goldberg, B., & Sottilare, R. (2017, July). A Cognitive Modeling Approach-Does Tactical Breathing in a Psychomotor Task Influence Skill Development during Adaptive Instruction?. In *International Conference on Augmented Cognition* (pp. 162-174). Springer, Cham.

Kim, J. W., Dancy, C., & Sottilare, R. A. (in press). Towards using a physio-cognitive model in tutoring for psychomotor tasks. In *Proceedings of the AIED Workshop on Authoring and Tutoring for Psychomotor, Mobile, and Medical Domains*. England, UK.

Kim, J. W., Sottilare, R. A., Goodwin, G., & Hu, X. (2017). Assessment of Individual Learner Performance in Psychomotor Domains. *Design Recommendations for Intelligent Tutoring System-Volume 5: Assessment Methods*, *5*, 319.

Marriage, B., & Kinnear, J. (2016). Assessing team performance–markers and methods. *Trends in Anaesthesia and Critical Care*, *7*, 11-16.

Martin, B., Mitrovic, A., Koedinger, K. R., & Mathan, S. (2011). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction*, *21*(3), 249-283.

McCormack, R., Kilcullen, T., Sinatra, A.M., Brown, T., & Beaubian, J. (in press). Scenarios for Training Teamwork Skills in Virtual Environments with GIFT. In Proceedings of the 6th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym6).

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, *1*(1981), 1-55.

Ocumpaugh, J., Baker, R. S., Rodrigo, M. M., Salvi, A., Van Velsen, M., Aghababyan, A., & Martin, T. (2015, July). HART: The human affect recording tool. In *Proceedings of the 33rd Annual International Conference on the Design of Communication* (p. 24-29). Limerick, Ireland: ACM.

Pew, R.W. & Mavor, A.S. (1998). *Modeling Human and Organizational Behavior: Applications to Military Simulations*. Washington, DC: National Academy Press.

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, *66*(3), 215.

Salas, E., Benishek, L., Coultas, C., Dietz, A., Grossman, R., Lazzara, E., & Oglesby, J. (2015). *Team training essentials: A research-based guide*. New York, NY: Routledge.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT). *GIFTtutoring. org*.

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

Sottilare, R. A., Graesser, A., Hu, X., & Holden, H. (Eds.). (2013). Preface. *Design recommendations for intelligent tutoring systems: Volume 1-learner modeling* (Vol. 1). US Army Research Laboratory.

U.S. Department of the Army (10 June 2011). Combined Arms Operations in Urban Terrain (ATTP 3-06.11).Washington, DC: USDA Headquarters

Chapter 10 – Team Assessment and Pedagogy as Informed by Sports Coaching and Assessment

Benjamin Goldberg[1], Benjamin Nye[2], H. Chad Lane[3], Mark Guadagnoli[4]
U.S. Army Research Laboratory[1], Institute for Creative Technologies[2], University of Illinois at Urbana-Champaign[3], University of Nevada Las Vegas[4]

Introduction

Sports psychology and team training literature offers a wealth of examples on the development of expert teams, where individual high-performing players are trained to succeed as a group. Compared to many problems where expert teams operate, sports very often have the advantage of clear criteria for success and an increasing amount of data collection on both team and individual performance. As such, effective training techniques for sports teams may offer insights that inform general team training pedagogy, particularly for psychomotor domains such as military operations. Our investigation into this area is exploratory. Here, we examine a few well-documented cases of how assessment and feedback are utilized in the context of sports team training. Given this focus, our work cannot be considered representative of training in any specific sport. However, by comparing between sports, we hope to identify certain qualitative differences in the types of training that are effective for different psychomotor domains that incorporate elements of team dynamics.

Training differs substantially between sports, not just in terms of the physical tasks but also in terms of the coordinated team activities that occur. For example, while only some baseball fielding events involve dependencies between players (e.g., double plays), nearly all plays in football require multi-faceted coordination (e.g., blocking by the offensive line to support a running or passing play, with specific routes run by receivers, etc.). These issues will be considered in the context of three sports: baseball, basketball, and football. These sports differ in terms of key game characteristics such as simultaneous coordination, micro-games within the sport (e.g., at bats, tip-offs), and team roles (Elverdam & Aarseth, 2007; Mueller, Gibbs, & Vetere, 2008; Ward, Farrow, Harris, Williams, Eccles, & Ericsson, 2008).

In this chapter, we consider pedagogical insights offered by three different sources of information from sports coaching and assessment: published reports of sports training, first-hand accounts of team training, and a review of assessment approaches for measuring team performance. These issues are considered in the context of an integrated taxonomy of feedback that considers when feedback was given, who it was given to (e.g., individual vs. team), the type of feedback (e.g., positive vs. negative), and the specificity of feedback (e.g., detailed issues vs. brief note). The goal of this work is to consider how these patterns might generalize to a wider range of learning tasks, to improve both learning and assessment of team performance.

We are particularly interested in how coaches communicate feedback that is directly aligned with training activities (i.e., information conveyed during training, or after training but based on those events). This focus is not because we believe that this is the most important role of a coach: there are many types of coaches with different specialties and there are many roles that feedback can play in different contexts, such as during reflection, mentoring, off-the-field, and more. However, this exploration is focused on how insights from team coaching relate to what is known about feedback in educational settings, and how it might enrich the automated delivery of feedback through intelligent tutoring systems (ITS) or other learning technologies..We are focused on this subset of coaching feedback and activities.

Decades of research has focused on the delivery and effects of feedback on performance (Kluger & Denisi, 1996), learning (Shute, 2008), academic emotions (Pekrun, Cusack, Murayama, Elliot & Thomas, 2014), self-regulation (Baumeister, et al., 2006), and more. In the subset of this literature that investigatedfeedback delivered in ITS platforms, multiple publications have been produced that systematically breakdown what effective expert tutors do from an instructional strategy standpoint, so as to enable an automated system to enact those interaction types (Durlach & Spain, 2012; Shute, 2008). Similarly, it is believed that team-based applications should be based on what strategies and tactics effective coaches apply, along with identifying the conditions and variables that dictate what strategy to apply for what individual team-member, and when. While this chapter only scratches the surface of this research challenge, we aim to establish a theoretically derived taxonomy to guide future studies that aim to code expert coaching practices through observational methods.

Roles for team training: types of teams

Fundamentally, one question for all team training research is what constitutes a team, such as how to distinguish between a "team of experts" versus an "expert team" (Salas, Cannon-Bowers, & Johnston, 1997). For example, Olympic teams for sports such as boxing or downhill skiing often train together but do not necessarily compete simultaneously or in coordination with each other. Even when such teams compete in the same event (e.g., relay races), some teams are closer to a "team of experts" rather than an "expert team." This distinction has often been raised with examples such as the United States Olympics men's basketball team. Historically, these squads have been viewed as including elite players but having minimal practice as a team, and which has been cited as one cause for their losses in the 2004 Summer Olympics (Leopold & Teitelbaum, 2016). With that said, the United States squad won prior Olympics competitions despite limited practice together.

This raises the issue that while an expert team provides a competitive edge, the advantage for well-practiced coordination likely depends greatly on the task or sport. For example, new members of a professional football team coordinate based on extensive playbooks that can take weeks to learn. On the converse, baseball players are often traded and start on their new team the next day. Even within a sport, team training can vary: pitchers and catchers report a week or more earlier to baseball spring training than position players (for 6+ weeks rather than 5 weeks). Despite the variance across these examples, a desired end-state is operationally defining what characteristics and behaviors are congruent with “expert teams”, based on context, and then defining what coaching tactics can be applied to accelerate the acquisition of those defined attributes. While there are recent contributions to the literature that define the framework and behavioral markers that make up a good team (Sottilare et al., 2017), there is little written on how to use those measures to drive pedagogical decisions at the team level.

Accordingly, at the professional or higher amateur level, coaches and managers serve a key role for building an expert team. Part of their contributions are certainly to assemble the initial team of experts (e.g., recruiting players). However, on an ongoing basis coaches must set the tone for the team culture, the systems and strategies that the team trains to master, and deliver the feedback to improve how the team executes these strategies (Lyle, 2002; Ericsson, 2003). This touches multiple areas of how coaches support athletes:

1) Training: Communicating knowledge to the team to help build skills and develop expertise through deliberate practice methods (Ericsson, 2003)

2) Motivation: Reinforcement and punishments that encourage improvement and “buy-in”

3) Culture: Establishing the team's motivation, goals, and self-regulated strategies for training and enforcing norms (e.g., how a team supports and polices itself)

4) Leadership Structure: How a coach determines and trains leaders within a team

As noted earlier, this coaching behavior occurs in many contexts- not just during practice, but also in the locker room, at team dinners, when meeting the families of athletes, and in other venues (Lane, 2004). From the standpoint of this research, we are focused on the types of interventions that coaches perform during or adjacent to training opportunities (e.g., post-game, reviewing video tape, etc.), since these are likely to be the ones that produce the most direct learning impact. This is because coaches can ground their statements in specific practice or game plays that have occurred, as well as direct players to apply the feedback immediately (e.g., retry a task).

Coding Feedback in Coaching in Team Sports

To consider how and when coaches intervene during or after practice and training, we must first identify the features that distinguish between different types of feedback. For this purpose, we outline a potential typology for coding the types of messages that coaches may deliver to players. This outline of feedback messages is based on speech act coding, which considers the functional purpose of dialog as actions (e.g., "Great job" could be coded as "positive feedback"). This approach has a long history in research on human tutoring and instructional systems, with a variety of taxonomies for speech acts (Cade, Copeland, Person, & D'Mello, 2008; Samei, Rus, Nye, & Morrison, 2014). These taxonomies distinguish between qualitatively different categories (e.g., questions vs. answers) and also between different subtypes (e.g., positive vs. negative feedback).

Many of these taxonomies assume a two-person conversation, however, and also fail to consider factors that may be relevant to the context where feedback is delivered (e.g., the difference between yelling from across a field versus taking someone aside to talk quietly). Table 1 notes features that we think are particularly important for feedback in team sports. In this description, we assume that a team task with multiple players is occurring, such as a training exercise, scrimmage, or game. While these represent only one possible set of facets, we believe that they capture some important differences in feedback that may be relevant to different coaching styles. For example, managers are often dichotomized into player-friendly managers versus more authoritarian archetypes, which imply quite different feedback (Boswell, 1984). The types of feedback that coaches use are also almost certainly affected by the composition of players and the context of the team (e.g., part of the season, recent performance/streaks).

Table 1: Features to Code Types of Feedback

Feature	Description	Options
When Given	The timing of feedback in relation to the ongoing task actions.	- Immediate (talk over task) - Abort (stop task) - After-action (after task finishes)
Direct Targets	Who is the primary target of the feedback?	- Individual - Subgroup of participants - Task participants - Full team
Secondary Observers	Who else receives the feedback? This may impact peer influence or vicarious learning (e.g., from others' mistakes)	- Individual - Subgroup of participants - Task participants - Full team
Valence	What is the overall directed tone of the feedback? e.g., feedback used to reinforce effective behavior (positive), correct or punish incorrect behavior (negative), or does not directly advocate correctness	- Positive - Negative - Neutral
Stress/ Emphasis	The level of emphasis for the feedback, from calming to direct to aggressive	- Calming - Direct - Aggressive
Information	The type of task-relevant or learner-relevant information shared (if any). Classifies if the feedback: (1) directs attention to specific elements that affected the outcome, (2) explains a pattern, (3) explains the whole outcome/status, etc. There are likely to be many other information types that might be relevant	- Cue ("Watch your footwork") - Command ("Now 10 throws") - Explanation ("That's because..") - Outcome ("Sloppy play") - General ("Winners don't quit") - Process ("Listen harder") - Progress ("Three more left") - Motivation ("You'll get this.") (others)
Post-Feedback Task Command	Common activity types applied by coaches following a feedback intervention.	- Continues (task not interrupted) - Resume (go where team left off) - Repeat (start activity from the beginning…"start from the top.") - Interject (start new activity based on feedback or assessment) - End (complete activity and move into next training exercise)

Training Feedback: Examples from Different Sports

To examine how feedback is used for team training across different sports, we consider the differences between three common sports in the United States: baseball, football, and basketball. Table 2 outlines some of the similarities and differences between the team coordination required for each sport. These include the team size, the typical number of players substantially involved in a play, and the typical pacing of tasks in the sport (e.g., mostly discrete versus relatively continuous play). While some numbers shift for different leagues (e.g., college versus professional), the order of magnitudes remain the same. Pacing issues might substantially impact how and when feedback can be delivered during practice. An example of a game task with high coordination is also noted, and contrasted against one with relatively little coordination.

Table 2: Comparison between Sports Team Coordination Tasks

	Baseball	Football	Basketball
Games/Season	162	16	82
# of Players	9 roles (25 active)	11 roles (45 active)	5 roles (13 active)
# of Players in a Typical Play	4 (Pitcher, catcher, two fielders)	11 (all)	5 (all)
Role Switching Points	Offense/Defense switch every half inning (e.g., after 3 outs)	Offense/Defense switch at change of possession (e.g., following score, punt, turnover)	Shifts in offense/defense roles following possession change (e.g., following score, rebound, turnover)
Pacing	Discrete plays, with coordinated fielding	Discrete plays, with multiple coordinated activities	Mostly continuous play, broken up by scoring, out of bounds, etc.
Example of High Coordination	Infield double-plays (infrequent)	Offensive line blocking (frequent)	Pick and roll (frequent)
Example of Low Coordination	Hitting with bases empty (frequent)	Punting (infrequent)	Free throws (frequent)

Baseball: Discrete Plays with Well-Defined Individual Contributions

Of the sports considered in this exploration, baseball contains the strongest individual contributions. Even during fielding, plays are coded with the fielders who touch the ball and players who do not touch the ball. Others involved typically only need to coordinate in real time by ensuring proper backup behavior (e.g., the pitcher covering a base when its fielder needs to pursue the ball). While teams may coordinate on fielding patterns (e.g., shifting to the right or to the left), the optimal positions for these plays tend to be relatively deterministic. Within baseball, the two highest-coordination activities tend to be pitchers and catchers (selecting, throwing, and catching pitches) and double-plays (e.g., which tend to require rapid throws from the shortstop to the second baseman, then a second throw to the first baseman). While at bats are common, double plays are infrequent. On the converse, batting is mostly individual.

Baseball training is structured around a manager (head coach), with assistant coaches that include the bench coach (second in command), batting coach (offensive expert), pitching coach (defensive expert), bullpen coach (works with secondary pitchers for a game), and a pair of on-field coaches at first and third base to help call plays. The on-field and bench coaches are often responsible for training on fielding. Above all of these is the general manager, who does not directly manage the players but who may set guidelines for the manager's decisions. In recent years, this has increasingly involved guidance based on statistical data (Thorn, Palmer, & Reuther, 2015). A number of studies have examined the characteristics of baseball managers (e.g., James, 2014). Boswell (1984) categorized managers into four archetypes:

1) Little Napoleons: Authoritarian and intense/stressful styles that emphasize competition;

2) Uncle Robbies: Player-friendly, often humorous leaders who lead with wisdom;

3) Peerless Leaders: Disciplined and dignified styles who lead by embodying this character;

4) Tall Tacticians: Intellectual, clever leadership based on trust in their judgement;

Koppett (2000) reinforced this work with a study of the "family trees" of manager lineage, which traced three managers who served as archetypes for Boswell (the Little Napoleons, Peerless Leaders, and Tall Tacticians). As such, these different managerial styles appear to be not just by chance but also by training and recruitment. In terms of in-game management, there has also been some discussion that traits from Uncle Robbies may be increasingly common among baseball managers (Diamond, 2016). This is potentially due to both the increase in player salaries, which could make authoritarian approaches unfavorable for recruitment and retainment. It may also be due to the increasing influence of stats-based guidance from general managers, which makes managers the lynchpin for building buy-in so that players support front office decisions (as opposed to only justifying their own coaching decisions). This shift may also be due to increasing initiative by baseball players, who are more active in self-regulating their training than in prior decades. As such, due to shifting power dynamics, earlier stages of baseball might offer better models for feedback on psychomotor tasks, while major league baseball might be a better model for studying how managers build and maintain team cohesion among experts across a long season with continuous games and travel. These distinctions are important, as GIFT supports multiple pedagogical models that enact variations in coaching methodologies. As such, coaching styles can be configured and called upon at run-time based on team characteristics that dictate the most appropriate coaching strategy.

Given that this exploration is primarily focused on team learning, we looked further into the coaching practices at the college baseball level. A review was made of publicly posted videos about college coaching practices. Franco (2018) explains a division 1 schedule where players were responsible for three types of training to prepare for their season: weights training (5h/week), conditioning (3h/week), individual instruction with a coach (3h/week), and about 4 game-length practices each week (12-14h/week). Some notable themes of baseball coaching include center around deliberate realistic practice (Ericsson, 2003), such as "perfect practice makes perfect" and "practice like you play" (e.g., moving away from massed drills). The majority of practice consists of either of individual tasks (batting practice, pitcher-catcher bullpen sessions), simulated at-bats (tee-ball drills, fungo bat fielding), or practice games. After-action practice includes verbal feedback and videotape reviews, which might be used to demonstrate models of good performance, discuss an individual player's performance, or discuss team performance.

Individual mechanics (e.g., batting stance, throwing shoulder) appear to be a primary focus for baseball training, even in the context of pair exercises (throwing practice) or games. Coordination is achieved by players independently recognizing the same game state, then following well-practiced procedures that assume their teammates will be in position (e.g., beginning a throw to first base before the first baseman is set up yet). The secondary focus appears to be conditioning exercises. Notably, in reviewing certain tapes, the primary feedback for conditioning (e.g., endurance) exercises is primarily supportive or progress feedback from other players. Finally, team coordination feedback appears to be the tertiary focus (e.g., adjusting fielding positions relative to others). Coordination is primarily practiced through repeating scenarios (e.g., fielding balls hit to a certain area). To help illustrate this balance, a brief "hot mic" video of the assistant coach for Oregon State was coded using the features from Table 1 (OSU Beaver Athletics, 2011). Across a 2 minutes of edited drill footage, the most common speech acts are noted in Table 3.

Table 3: Coding Speech Acts for Two Minutes of a College Baseball Fielding Practice

Count	Feedback Type	Feedback Example
6	Positive-Outcome	"good"
4	Neutral-Explanation	"shorter arm circle out of here it's here"
4	Neutral-Command	"defense let's go one and around five to forty five"
3	Positive-Specific	"ground up very nice"
2	Negative-Specific	"a little lower long and lower long and lower"
2	Neutral-Cue	"light on feet"
1	Positive-Motivation	"you make that play you're our guy"
1	Neutral-Specific	"this way so our knees a little more"
1	Negative-General	"it still needs a little more work"

At least in this clip, the coach provides feedback nearly continuously (24 statements in 2 minutes), as the mechanics unfold, and supports retries to practice issues that he identifies. While multiple players are involved, they each are the primary focus of coaching feedback at different times. In one example, a player is taken aside for a longer explanation of his specific areas to address. However, in later portions of the tape the coach stops active practice and provides a demonstration and strategic feedback to the current practice squad (e.g., directly addresses the group). While this is only one example, it shows a potential process for examining how coaches work with players under different practice conditions, as well as how and when they provide their feedback. The frequency of feedback and its delivery patterns (throughout practice or across practice sessions during the season) might also provide valuable input in terms of how coaches respond to success or failure of the team as a whole, such as by the intensity or attention of feedback given (e.g., showing that they care about the success of the team, either through more frequent, more positive, or even more negative feedback). This level of analysis can form the basis of feedback and coaching patterns that could be implemented in a pedagogical model for a system such as GIFT, such as for tasks which resemble baseball activities (e.g., strong individual contributions, well-defined team interactions).

Football: Discrete Plays, but Highly Interdependent Contributions

The sport of football is an example of team dynamics that requires the coordination of multiple explicit roles to execute a single discrete play/action at a time. For each designated play, both on the offensive and defensive side of the ball, each individual player is assigned a specific role with conditions and standards that determine their behavior during execution. The nuance here is the interdependency across each role during play execution, and how a shared situational understanding across all interacting parts is required to optimize objective outcomes.

From a pragmatic stance, this requires: (1) knowing what specific role you are responsible for and how to execute those functions consistently, (2) knowing how your role fits within the conceptual context of the team across all potential scenarios, and (3) knowing how to adapt your role and communicate those changes based on tactics and behaviors observed within the operational environment (Baker & Côté, 2003). These categorical distinctions of situational understanding are important as they can be used to define training approaches that target the Knowledge, Skills, and Abilities (KSAs) associated for each designated role and at each designated phase (see Table 4). With this framework in place, targeted skill objectives can be deconstructed to identify specific components that drive the selection of practice activity types. In this instance, one can apply a mastery learning paradigm as it adheres to team situational contexts (i.e., an individual cannot integrate within a team until that individual can consistently perform an assigned role at a high level and understand how that role serves the team as a whole). Training programs should begin to target each KSA and apply pedagogical techniques that promote acquisition and retention. For this purpose, the team feedback taxonomy presented above (see Table 1) can be applied to configure feedback and coaching features as they adhere to who and what is being trained within each phase of the Team Situational Understanding taxonomy. Similarly, speech act analytic techniques, like the one described above, can be applied to generate initial feedback policies that associate with coach-athlete dynamics at both the individual and team collective level.

In the domain of football, the type of KSAs trained within each phase will dictate the feedback and motivation approach applied at both the individual and team-context level. Similarly to baseball. Coaches are designated across distinct levels of team composition (e.g., head coach, offensive/defensive coordinators, position coaches, and conditioning coaches) and serve a distinct purpose during training, practice and game execution. Accordingly, each coach applies variations in training strategy based on the situational understanding phase they are responsible, as each phase is directed towards disparate, but complementary skill sets.

Just as a member of the team must maintain situational understanding appropriate to their role and assignment, a coach must also apply situational leadership, where they adapt styles based on environmental indicators and individual differences (Hersey & Blanchard, 1969). This nuance can also be seen above in the baseball example, where the coach varies their coaching technique based on the goal of the interaction and their desired audience. When directing subordinates, coaches often exhibit both relationship and task dimensions as a means to influence future task execution, with most common leadership styles being a mix between *Autocratic, Democratic, Positive Feedback, Social Support,* and *Training and Instruction* leadership philosophies (Turman, 2001). From an ITS standpoint, pedagogical approaches can be designed for each phase that adhere to different coaching philosophies and style; however, it is important to note that these pedagogical determinations are bound by the assessments captured within each specific activity or exercise. In the space of interdependent team interaction, a major focus is on scenario-based interventions that challenge skill application at both the individual and collective level, with situational understanding serving as a guiding framework to ground all training functions. For this purpose, ITS applications designed to train and build relevant team-oriented skill sets are recommended to adhere to structured pedagogical formalizations that associate with phases like the ones described above. To support a domain-agnostic representation, the framework must also extend across variations in team composition and role interdependencies, as many team contexts operate outside of discrete events, where play is much more continuous and dynamic.

Table 4. Situational Understanding Framework Applied for Team Skill Training

Training Situational Understanding (Phase)	Training Strategy	Example	Assessment and Feedback Factors
Know and execute assigned role (1)	*Individual:* Focused interventions that target fundamental application of skill. Deliberate Practice (Ericsson, 2003) *Team:* Working in subteams for practice and peer-learning opportunities	*Individual*: Wide Receiver executing catching drills across multiple flight trajectories and speeds. *Team*: Wide Receiver executing catching drills with cornerback providing coverage.	*Individual*: - Technique and Execution - Mechanics - Reinforcement *Team*: - Challenge and Motivation - Competition
Know how an assigned role fits within the context of the team (2)	*Individual*: Understand role for each individual play, with ability to recall all decision points *Team*: Drill-repeat exercises to promote deep understanding and consistent application	*Individual*: Wide Receiver studies playbook and knows all routes and blocking assignments. *Team*: Offense-only drills with wide-receiver applying proper technique based on given play calls.	*Individual*: - Declarative & Procedural knowledge about context of plays *Team*: - Role Interdependency - Shared Mental Models - Trust
Know how to observe, communicate, & adapt (3)	*Individual*: Study film and prior use cases/examples. *Team*: Subject team to variations in context that require role execution. Scenario-based practice events designed to challenge communication and adaptation across team-members and roles.	*Individual*: Case-based exercises that test ability to predict play outcomes based on pre-play observations. *Team*: Multiple scenarios that require communication between wide receiver and quarterback (e.g., identify single coverage and communicate a play audible to exploit recognized weakness)	*Individual*: - Conceptual Understanding of Opponent Behaviors - Counter-measure tactics based on opponent observations. *Team*: - Communication - Context Analysis - Coordination - Leadership

Basketball: Continuous Play with a Mix of Individual and Team Contributions

Basketball is another excellent domain to guide pedagogical considerations for Team ITS development as it introduces new elements not captured in the examples provided above. In this arena, basketball teams are composed of highly interdependent positional players with roles that are more loosely defined; meaning that roles are transferable at any moment. They are transferable in the sense that each teammate has both offensive and defensive responsibilities, as well as transferable in the sense that each teammate might switch roles within a given play to assist in meeting team objectives (e.g., a forward taking over defensive responsibilities on an opposing guard due to a screen). Three common characteristics of this type of team interaction is high task interdependence (Landi, 2001), high interaction, and uncertainty, which is due to the consequence of interacting team members and an evolving context (Wall, Cordery & Clegg, 2002; Ramos-Villagrasa, Navarro & Garcia-Izquierdo, 2012). In these instances, game flow is much more continuous, where team dynamics are highly reactive and often spontaneous. These dynamic characteristics are seen in many other relevant team compositions, including hockey, rugby, and even in military teams.

Due to these continual shifts in task context, team situational understanding is critical to success, with communication playing a pivotal role in collaborative coordination and cohesion. To address this shifting context, coaches are responsible for designing practice activities that build and strengthen foundational team behaviors, with direct abstraction from the phases defined in Table 4. This involves individual ability drills to build functional skills (e.g.,, dribbling, shooting, passing, etc.); team-/collective-drills to build flow, cohesion, and coordination (e.g., fast-break drills, play rehearsal, scenario-based walk-through); and scrimmage drills that recreate game situations for dedicated practice under the operational context. Formative assessment and directive feedback are applied across all activity and drill types, with task context (i.e., practice activity objectives and the interacting parts) and coaching philosophy (as based on configurations across Table 1) being the determining factors that drive pedagogical interventions.

In the domain of continuous play environments with shifting contexts, an additional element of situational understanding that plays a critical factor in practice and coaching design is team familiarity (i.e., shared mental models; Fletcher & Sottilare, 2017; DeChurch & Mesmer-Magnus, 2010). In this instance, familiarity relates to all parts that constitute a team and the environment for which they operate in. Each serving member of a team must be aware of the strengths and weaknesses not only of their contributions to the team objectives, but across all potential interacting parts, including opposing elements. Knowing what each interacting individual provides to the task environment enables effective decision making for the purpose of exploiting specific strengths and weaknesses for the good of the team (e.g., calling plays to exploit situational weakness, such as taking advantage of an undersized defender). In these scenarios shared mental understanding is critical, where decision cycle times are reduced and automated based on shared understanding of task interdependencies and location and role of each interacting element. It is through initial well-designed practice activities that familiarization is acquired, with advanced application leading to individualized and often creative communication techniques (i.e., using gestures and subtle signals to communicate intent and coordination). In these instances, while the coach emphasizes the importance of teaming behaviors, it is often up to the interacting team-members themselves to devise a specific solution to the coaches defined objective.

The activities and deliberate practice techniques will vary across domain, with the defined team characteristics providing guiding principles to start from. For basketball, the activities will vary from football, as the defined characteristics of that team require it. In these instances, scenario-based team exercises will be created that target specific skill sets. In addition, the specific coaching practices seen across collegiate basketball programs may be of extra relevance for the design of a team pedagogical philosophy for military relevant contexts, as the turn-over for NCAA teams is high, where athletes are eligible for the pros after one year of post high school activity. In these instances, it may be relevant to code specific strategies executed by collegiate basketball coaches that address new additions to an already existing team structure. Accelerating cohesion in these instances is critical to team success.

Challenge point framework

When establishing coaching strategies that target accelerated team development, it is also important to consider how GIFT and ITS can manage challenge in a generalized capacity that applies across domains. The Challenge Point Framework is an approach to model these interactions, which is based on the relationship between task difficulty and the learner's ability. Tasks can become more difficulty, for example, with greater accuracy requirements, more pressure on success, or more to the point of this chapter, a decrease in the feedback presented.

Figure 1 shows the relationship between task difficulty and immediate (practice) performance and learning. Immediate performance (e.g., during practicing) is the solid orange line. As task difficulty increases, performance decreases. The dotted, green line represents long-term learning. As task difficulty increases, learning also increases, up to a point. That point is known as the optimal challenge point (OCP) and designates the point where the individual is being optimally challenged for long-term learning. You will notice that at the OCP, practice performance is compromised. In other words, OCP creates short-term struggle for long-term gain. As shown in the graph, the level of difficulty at which learning is optimized is not the same level of difficulty that promotes the best immediate performance.

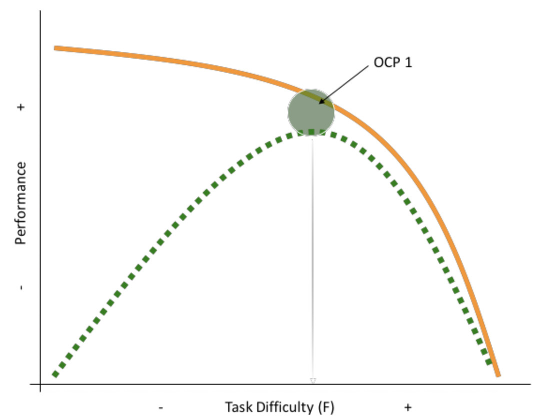


Figure 1. A graphical representation of the optimal challenge point (OCP). Image adapted from Guadagnoli and Lee (2004).

As the learner progresses so does the OCP. Figure 1 models performance/learning for a relatively experienced learner. Figure 2 models performance/learning for both an experiences learner (OCP 1) and a novice learner (OCP 2).

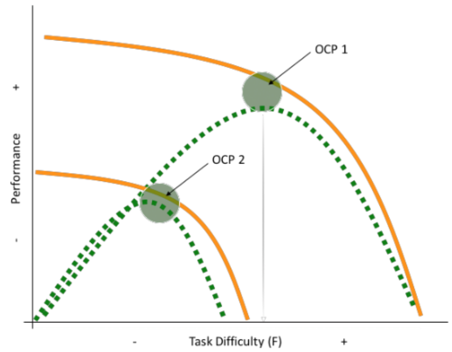


Figure 2. A representation of optimal challenge point for relatively novice (OCP 2) and relatively more experienced learners (OCP 1). Image adapted from Guadagnoli and Lee (2004).

As you see, the same basic relationship between task difficulty, immediate performance, and learning is maintained, but the optimal challenge point (OCP) is at a lower difficulty level for the novice learner than for the experienced learner. As a result, knowing how to increase and decrease challenge for the individual is a key component to optimized learning. As mentioned, feedback is one way to change the level of challenge. Decreases feedback provides less guidance and therefore requires the individual to rely on intrinsic feedback mechanisms. This increases the task difficulty for the individuals, and for more experienced individuals decreased feedback degrades immediate performance but enhances long-term learning.

Based on our understanding of the Challenge Point Framework, it would be much more effective for the coach to deliver less frequent feedback to this level of skilled performers. This would likely yield more struggle during practice but greater performance in the game (as a result of greater learning in practice).

Conclusions and Potential Opportunities

In the design of team-based ITSs, sound pedagogical practice is critical. In this chapter, we argue for the use of sports psychology and sports coaching lessons-learned to guide the design of a generalizable pedagogical model for use in the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012). In its initial implementation, a major requirement is first establishing instructional strategy functions that adhere to the taxonomy presented in Table 1. This involves establishing variables with configurable enumerations in GIFT's team pedagogical model that will inform a set of production rules or agent policies used at run-time. This requires identifying all team-level strategies (i.e., a GIFT strategy requires domain-generalizability) and establishing tactic level representations and their associated dependencies (i.e., how feedback is presented and how adaptations are managed).

With a set of supported pedagogical functions that can be enacted by GIFT, the next requirement is generating a set of policies and/or rule sets that are used at runtime. An overarching objective is to establish empirically informed policies that determine how best to coach a team based on the composition of that team (i.e., team structure, roles, and interdependencies), the personalities that make up that team (i.e., individual differences), and the domain for which they execute as a team (i.e., context to guide strategy configurations). Before a coaching philosophy can be applied, as defined through dedicated policies, there are certain tenets that drive the design of team training events when a specific domain is specified. As seen in the Situation Understanding Model defined in Table 4, a major component of team effectiveness is applying deliberate practice techniques that target specific skill sets and objectives at both the individual and team context. This highlights a need for pedagogical consideration at two levels of interaction: (1) what each individual role of a team experiences before, during, and after a specified exercise (e.g., macro-adaptation), and (2) what assessment and feedback is applied at each managed interaction before, during, and after a specified exercise (micro-adaptation). The goal is to establish generalized rules based on the feedback taxonomy in Table 1, where configurations can be parameterized based on loggable performance and behavioral data sources. As an example, pedagogical model policies will have to differentiate and resolve conflicts across feedback and drill-repeat strategy types (e.g., when to yell feedback vs. when to interrupt an event and instruct the team to start from the beginning).

In a traditional ITS, these policies are derived from interactions observed across expert tutors, where their actions and strategies were coded for strategy analyses to determine what common practices effective tutors consistently utilized. The resulting pedagogical 'best-practices' are then translated and programmed as algorithms that serve as the pedagogical logic that guides feedback delivery and adaptation practices. In the case of team-based ITSs, similar methods are recommended. An approach is to design an observational study that incorporates multiple coaches across multiple disciplines for the purpose of deriving effective strategies applied across the team context in multiple instances. This would require establishing coding schemas that adhere to the taxonomy elements in Table 1, where representations are required that link observable performance/behavior indices and their association with coaching related tactics. These assessment indices to coaching tactic relationships can be configured across distinct coaching philosophies (e.g., *Autocratic, Democratic, Positive Feedback, Social Support*, and *Training and Instruction)*, where a training developer can customize the coach type that drives feedback and strategy interventions.

At a fundamental level, when compared to one-on-one tutoring strategies, the strategies and tactics used by team coaches are only anecdotally understood: while coaches' small decisions in a game are picked apart by sportscaster talking heads for days, how much time is spent rigorously analyzing the content and patterns of how they interact with their players on a day-to-day basis? From a research standpoint, there could be great value in extensive data collection across a number of expert coaches with the goal of building generalizable policy sets for feedback and motivational interventions for a team. However, this would require a major study with deep data collection. An alternative approach that can be applied in the near-term would be to build pedagogical coaching models that are based on specific individuals and/or philosophies that are well documented in the literature (i.e., building models from documented theory, rather than training from data). An example would be explicitly studying one well-represented expert coach (e.g., Duke Basketball coach Mike Krzyzewski) and building policy sets on that individual as represented through books, interviews, video observations, and peer assessments. The results of this approach would probably be too coarse-grained to infer the specific tactical philosophy of that coach regarding their expertise within a domain (e.g., how they provide feedback). Instead, the goal is to extract coaching strategies and methods that focus on managing team behavior and development. Based on these insights, GIFT could build a set of alternate "coaching" pedagogical models based on team training experts. These models could then be analyzed and compared for different kinds of teams and learners, using data from these studies to improve our understanding of how virtual coaches can improve team training.

Acknowledgements

Part of the effort described here is sponsored by the U.S. Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

Baker, J. & Côté, J. (2003). Sport-specific practice and the development of expert decision-making in team ball sports. Journal of Applied Sport Psychology, 15, 12-25.

Baumeister, R. F., Gailliot, M., DeWall, C. N., & Oaten, M. (2006). Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. Journal of personality, 74:6, 1773-1802.

Boswell, T. (1984). From Little Napoleons to Tall Tacticians. In *Why Time Begins on Opening Day*. (pp. 83-104). New York, NY: Penguin.

Bjork, R. A. (2009). Structuring the conditions of training to achieve elite performance: Reflections on elite training programs and related themes in Chapters 10–13. In E. Erikson (Ed.) *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments*. (pp. 312-332). New York, NY: Cambridge University Press.

Cade, W., Copeland, J., Person, N., & D'Mello, S. (2008). Dialogue modes in expert tutoring. In *International Conference on Intelligent tutoring systems (ITS).* (pp. 470-479). Springer Berlin/Heidelberg.

Collins, A., Brown, J. S., & Newman, D. (1989). Cognitive Apprenticeship: Teaching the Craft of Reading, Writing, and Mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494): Lawrence Erlbaum Associates.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. Journal of Applied Psychology, 95:1, 32–53.

Diamonfd, J. (June 7, 2016). What Are Baseball Managers Actually Doing? *The Wall Street Journal.* Retrieved from: www.wsj.com/articles/what-are-baseball-managers-actually-doing- 1465336684 on Jan 15, 2018.

Durlach, P. J., & Spain, R. D. (2012). Framework for instructional technology. In V. G. Duffy (Ed.), *Advances in applied human modeling and simulation* (pp. 222-231). Boca Raton, FL: CRC Press.

Elverdam, C., & Aarseth, E. (2007). Game classification and game design: Construction through critical analysis. Games and Culture, 2:1, 3-22.

Ericsson KA. (2003). The development of elite performance and deliberate practice: an update from the perspective of the expert-performance approach. In Starkes J, Ericsson KA (Eds.) *Expert Performance in Sport: Recent Advances in Research on Sport Expertise,* (pp. 49-81). Champaign, IL.

Fletcher, J., & Sottilare, R. A. (2017). Shared Mental Models in Support of Adaptive Instruction for Teams Using the GIFT Tutoring Architecture. International Journal of Artificial Intelligence in Education, 1-21.

Franco, D. (2018). *Normal Day for D1 Baseball Player.* Retrieved from: www.youtube.com/watch?v=Eanjm6_MQfo on Jan 15, 2018.

Guadagnoli, M. A., Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. Journal of motor behavior, 36:2, 212-24.

Hersey, P. & Blanchard, K. H. (1969). Life cycle theory of leadership. Training and Development Journal, 23:5, 26-35.

Guadagnoli, M. A., Dornier, L. A., Tandy, R. D. (1996). Optimal Length for Summary Knowledge of Results: the Influence of Task-Related Experience and Complexity. Research quarterly for exercise and sport, 67:2, 239-248.

James, B. (2014). *The Bill James Guide to Baseball Managers.* Diversion Books.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological bulletin, 119:2, 254.

Koppett, L. (2000). *The Man in the Dugout: Baseball's Top Managers and How They Got That Way*. Temple University Press.

Landis, R.S. (2001). A note on the stability of team performance. Journal of Applied Psychology, 86, 446-450.

Lane H.C. (2012) Coaching and Mentoring. In Seel N.M. (Eds.) *Encyclopedia of the Sciences of Learning*. Springer, Boston, MA

Leopold B. & Teitelbaum, B. (2016). Red, White and Bronze: The death and rebirth of USA Basketball. Retrieved from: http://www.nbcolympics.com/news/red-white-and-bronze-2004- death-and-rebirth-usa-basketball on January 5, 2018.

Lyle, J. (2002). Sports coaching concepts: A framework for coaches' behavior. Psychology Press.

Mueller, F. F., Gibbs, M. R., & Vetere, F. (2008). Taxonomy of exertion games. In Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat (pp. 263-266). ACM.

OSU Beaver Athletics (2018) 2011 *Oregon State Baseball Webisode #1 - Marty Lees mic'd up for infield drills*. Retrieved from: https://www.youtube.com/watch?v=9ZXkJ2Jv660 on Jan 15, 2018.

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, 29, 115-124.

Ramos-Villagrasa, P. J., Navarro, J., & García-Izquierdo, A. L. (2012). Chaotic dynamics and team effectiveness: Evidence from professional basketball. European Journal of Work and Organizational Psychology, 21:5, 778-802.
Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team?: Emerging training strategies. Naturalistic decision making, 1, 359-370.
Samei, B., Rus, V., Nye, B.D., & Morrison, D. M. (2014). Hierarchical Dialogue Act Classification in Online Tutoring Sessions. In Educational Data Mining (EDM) 2014. (pp. 600-601).
Schmidt & Richard A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. Tutorials in motor neuroscience, 59-75.
Shute, V. J. (2008). Focus on formative feedback. Review of educational research, 78:1, 153-189.
Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence in Education*, 1-40.
Thorn, J., Palmer, P., & Reuther, D. (2015). The hidden game of baseball: A revolutionary approach to baseball and its statistics. University of Chicago Press.
Turman, P. D. (2001). Situational coaching styles: The impact of success and athlete maturity level on coaches' leadership styles over time. *Small group research*, 32(5), 576-594.
Wall, T.D., Cordery, J.L., & Clegg, C.W. (2002). Empowerment, performance and operational uncertainty: a theoretical integration. *Applied Psychology: An International Review*, 51, 146-169.
Ward, P., Farrow, D., Harris, K. R., Williams, A. M., Eccles, D. W., & Ericsson, K. A. (2008). Training perceptual-cognitive skills: Can sport psychology research inform military decision training?. *Military Psychology*, 20(S1), S71.

Chapter 11 – The Role of Context in Team Performance and Team Training

Wayne W. Zachary [1], Benjamin Goldberg [2], Andrew J. Hampton [3]
Starship Health Technologies, LLC [1], Army Research Laboratory [2], University of Memphis [3]

Introduction

Context is a construct through which humans integrate information on the social, physical, and experiential environment and represent it in a way that renders it relevant to current cognitive and behavioral challenges. Context representation and understanding are inextricably bound to expertise in the tasks and domains that make up those cognitive and behavioral challenges. When people engage in activities that require coordination, collaboration, and cooperation (i.e., teamwork), the team itself becomes part of the context. In an expert team[1], the team members have knowledge *about* the boundaries of each other's knowledge, experience, and/or skills that becomes part of each team member's context understanding, representation, and reasoning. This knowledge about the team allows each team member to interpret the actions of teammates and to act adaptively, through processes such as error identification, anticipation of needs for support, or as a basis for inferring teammate's intentions.

In this chapter, we examine the theory of context understanding as well as several examples of teamwork, and use these to consider how context contributes to team performance, and how it can be used in team assessment. We begin by defining context in terms of individual cognition, and proceed to discuss how context as a cognitive construct underlies cooperation and teamwork how affects teamwork performance. This analysis highlights how explicit representations of context information can (and arguably, must) be used in intelligent automated team training and tutoring, especially as could be done using an extended Generalized Intelligent Framework for Tutoring or GIFT (Sottilare, Brawner, Sinatra, & Johnston, 2017).

What Do We Mean by Context?

Consistent with the prior work of Zachary and collegues (particularly Zachary et al. 2013), we define context as a *cognitive process that is representation-centric, constructive, pervasive, and strongly interconnected with domain expertise*. This definition can be unpacked in several stages. At the most fundamental level, when we say context is a cognitive construct we mean that it concerns the representation and processing of symbolic information in a natural or computational system. With regard to human beings, context is internal to the human mind and cannot be directly observed or studied. Rather, it can only be studied indirectly, in terms of, for example, verbal accounts of a person's understanding of momentary context , or through examining accounts of how a person reports using context to reason and make decisions in a dynamic situation. These two examples are chosen to illuminate the representation-centric and constructive process aspects of context. This part of the definition also points out that, as an internal cognitive construct, context is distinct from the philosophical notion of empirical Ground Truth (GT). Some ways in which context and GT may differ are further considered below

In verbal accounts (both experimental and naturalistic), people commonly refer to context as a definitive 'thing', referring to is as 'the' context (not as 'a' context) which implicitly exists in the present, and which refers to the person's understanding of the external situation in the physical and social environment. Yet

---

[1] Which is not the same as a team of experts, as discussed below.

people also readily agree that context changes over time and is thus dynamic. This leads to the view expressed here that the ‘thing’ to which people refer as context is an internal representation that captures and integrates multiple aspects of the external (and internal) environment that are salient to the person at that moment. The representation is also constantly undergoing revision to reflect changing information in the environment, and changing cognitive states in the person as well. The dyanmic nature of this context representation suggests that context is a constructive process with one or many distinct lower level functions involved in building, maintaining, and (when necessary) deconstructing pieces of the context representation. However, what people are conscious of is current content of this representation, making context representation-centric.

The fact that people are able to talk about context also indicates that it is (at least to some degree) metacognitively accessible to introspection, although it appears that this introspectability extends only to the contents of the momentary representation, giving rise to momentary context awareness. This metacognitive accessabiltiy does not appear to extend to the dynamics of the process or to prior states of the representation. Thus, people are consciously aware of the current contents of the representation but not of the underlying processes that maintain it.

The context process is also pervasive, in that people always have some degree of context awareness. Context is not something that is relevant only to work, or to play, or to any other broad differentiator. Rather, it is omni-present, and people feel that they are always “in” some contex. At the same time, though, the degree of context awarness and the depth of context understanding varies a great deal across persons and even across domains. In particular, individuals with a great degree of knowledge about a domain usually have an awarness of context features that are particulalry salient in that domain, in the sense that those features can help define and restrict the range of appropriate decisions, planning, and actions in the domain. Chess masters, for example, represent abstract elements of context in a (mid-game) chess board that novices cannot and can use those abstractions to identify opportunities for application of a tactic many moves in advance. Similarly, air traffic controllers represent context of an air space with constructs that identify potential future conflicts while novices cannot, even though in both examples the novices and experts have access to the same GT stimuli.

The aspect suggests two key ways in which context, as understood by a person, can differ from GT. First, the propositions within a context representation that are verifiable empirically can be inaccurate with regard to GT (for whatever reason). An aircraft is empirically be flying at 5000 feet and on a heading of due north, but can be represented in the context understanding as at 10000 feet and heading north by north west. Second, and in contrast, there are abstracted and inferred components of a context understanding that are not part of the simple empirical GT, and that add insight that is present in the empirical GT The foreseen opportunity by the chess master is an example of this.

Extending this insight into the role of expertise in context understanding has been the work on recognition-primed decision making (RPD; e.g., Klein, 1993, 2008). This body of research points out that a critical feature of domain expertise is the ability to recognize specific features, or patterns of features, of a problem instance (i.e., those elements of context) and directly derive a decision/action from them. Put differently, RPD suggests that experts are aware of different and more pragmatically useful elements of context, but only as it applies to their domain of expertise. To the degree that rich representation of problem instances is part of context, we thus argue that context awareness is intertwined with development of domain expertise.

This last aspect of the definition of context has direct implications for individual and team training. It suggests that increasing domain expertise in individuals— training— must necessarily involve developing an enhanced ability to perceive, understand, and reason about context. This has long been demonstrated in research into the development of individual cognitive skills. In a summative reviews, VanLehn (1996) and

Zachary and Ryder (1997) argue that increases in a skill involve changes in the way the problem and environment are internally represented. This mechanism for this is reducing the procedural knowledge needed by, in essence, trading it off for more useful representation of the problem space, developed through experience in that problem space. We argue here that context representation therefore constitutes a critical enabling foundation for expertise, and that domain expertise cannot be trained without consideration of context, including team context. But what are the key contextual variables in team-based skills, and is context perception, understanding and reasoning fundamentally different under these conditions?

Context in Team Performance and Team Training

Research first began to indicate the importance of context in teamwork and collaboration in the late 1980s. Suchman (1987, 1990) worked from the perspective of understanding how machines and people could collaborate to solve everyday problems (specifically, copying documents). She demonstrated that the problems and frustrations that people experienced in interacting with an artificially intelligent machine arose from the fact that the people shared an implicit understanding of the current task context and tacitly used that understanding in their linguistic interactions (i.e., dialogs), while the smart copier with which they were working had a wildly different representation of the task and task context.

At the same time, researchers in team performance (particulalry Salas, Canon-Bowers, Fiore, and their colleagues) were developing data showing that effective team performance required "team members [to] hold common or overlapping cognitive representation of task requirements, procedures, and role responsibilities" (Cannon-bowers, Salas, & Converse, 1993, p. 22; see also Fiore, Wiltshire, Oglesby, O'Keefe & Salas, 2014). The early work into the role of context in teamwork provided two foundational insights. The first was that the task and role relationships among the people that comprise the cooperating team became part of the context representation for people working in those settings. The second insight, which impudently echoed the work of Suchman, was that the context understanding, including these team/collaboration aspects, had to be shared across team members for the intra-team communication and coordination to be effective. The following two examples demonstrate these two aspects of context and team processes.

1. Dismounted infantry. The location of individual service members in relation to terrain features forms a core element of operational context. It is of key importance to operational tasks such as organizing rendezvous. The Air Force Research Laboratory conducted a study examining the ability of dyads to rendezvous in unfamiliar environments (Hampton, Shalin, Robinson, Simpson, Finomore, Cowgill, Moore, Rapoch, & Gilkej, 2012). The simulation-based experiment tested the change in strategy and performance in standard two-way communication versus "spatialized audio" (wherein participants heard their partner's voice as if it came from that partner's relative physical bearing). This had the effect of shifting the more abstract task analysis levels of comprehension and projection to a relatively simple perceptual task. Participants no longer had to speculate on a partner's location and movement, adjusting narrative explanations for procedural communications (e.g., matching relayed descriptions of road layouts to supposed map quadrants). Likewise, descriptions of local terrain and compass bearing gave way to any discussion at all that would present a steady perceptual signal (including plans for the weekend, classroom assignments, and favorite video games). Unsurprisingly, these shifts in strategy coincided with significantly improved rendezvous times. A separate manipulation of the presence of distinct landmarks (a common navigational strategy) showed little impact in performance metric for the standard audio condition, though they factored prominently into participant discussion. However, in the spatialized audio condition, participants paid little attention to them at all, except as confirmation. At one point, a participant explicitly stated that a landmark did not matter with the auditory channel conveying position. The context of the task, in this case the position of the partner–target, was made immediately clear via perceptual channels.

2. Naval Fleet Air Defense. A US Naval destroyer focuses on the problem of command and control of multiple assets to provide continuous defense of own-ship and the whole surface combatant group from hostile attack form the air. For the critical dimension of airborne threats the function is termed Naval Fleet Air Defense C2. The Air Defense team in the combat-information center (CIC) aboard a destroyer can vary in size from six to eight members (with potentially more outside the CIC) (see Zachary et al, 1998 for a cognitive task analysis of this team). . The roles in the Air Defense team vary to some degree according to the mission and organizational decisions by the ship commander. The broad aid defense function is to detect, identify, monitor, and if necessary engage air vehicles that could pose a threat to own-ship and/or defended assets, particularly an aircraft carrier. The problem is particularly fast paced and complex in airspaces that involve combination of civilian air traffic and military air traffic that can be a mix of own, allied, neutral, and hostile. The extended air defense team includes airborne assets (electronics aircraft, patrol aircraft and fighter aircraft) for which CIC-based team provides control and situational information. Understanding the current mission context is key to both individual and team performance of the Air Defense team. Decisions and actions by all members of the Air defense team, including its commander (the on-duty Tactical Action Officer), depend on each member understanding enough of the context to choose her or his own actions for that context, and the overall strategy taken by the commander depends on that individual having the best and most accurate understanding of the mission context.

The two insights discussed above had immediate and lasting impact on team training, which was framed by the seminal study of team decision-making on Tactical Decision-Making Under Stress (TADMUS). The seven year study, reported in Cannon-Bowers and Salas (1998), initially was “conceived as a response to an incident involving the USS Vincennes, a warship that mistakenly shot down an Iranian airbus” (Cannon-Bowers & Salas, 1998, xxi). It was revolutionary, as the research was conducted in a naturalistic setting, and not in a laboratory. The TADMUS researchers studied how tactical decision-making teams on Naval warship actually worked and made decisions, and how those teams were trained. This naturalistic approach pointed out that while real-world teams may initially train together (as in pre-commission training for a new warship), in actual operation there is a constant rotation of new members into (and existing members out-of) these teams. As a result there may almost never be a situation where each team-member was truly expert in his/her task. Yet, the analyses showed that an “expert team” (i.e., a team that performed as a unit at a high level) was usually able to perform at a high level because it was able to adapt to the differing skill levels of its individual members.

The results indicated that portions of the context understanding that related to team roles and responsibilities, and to how those role and responsibilities were being carried out in the current problem instance, were central to team performance. It also helped explain an often-observed phenomenon, in which “a team of experts is not an expert team” (Salas, Canon-Bowers, & Johnston, 1998). From the perspective of team training, it pointed out the importance of understanding *what* was in these (domain-specific) shared models of team performance and *how* individual trainees and teams could most effectively learn them and apply them. In general, various researchers found that these context models of expert teams involved:

- An ability to form expectations and projections about the behavior of other team members,
- A motivation to proactively support less expert team members in achieving those expectations, and in reacting correctively when those expectations were not met; and
- A set of structured communications skills that served to activate those abilities and motivations across the team.

From a technology perspective, these results led to development of team training systems that involved practice in teamwork tasks (often including elements of computer simulation), and active guidance and feedback in developing the specific abilities, motivations, and communication skills listed above. A detailed review of that literature is beyond the scope of this chapter, but see Freeman and Zachary (2018) for a more detailed review of how team training research evolved forward from the TADMUS program.

The training of such abilities, motivations and skills, however, created new and interesting training problems, particularly regarding assessment and diagnosis. This is because they are based on the unobservable construct of shared context understanding. While behaviors such as compensatory correction of teammate errors or issuing proactive communications to prevent teammate errors can be observed, their absence is not. Moreover, simply observing a communication behavior provides insufficient information to diagnose either the correctness of its timing or correctness of its content. Those key aspects of diagnosis require an an understanding of the underlying context.

This last aspect of determining whether the context understanding is appropriately shared across the team and whether it is correct with response to the underlying environment is particularly difficult. An interesting example of this can be taken from the incident (mentioned above) in which the USS Vincennes shot down an Iranian airliner mistakenly thinking that it was disguised military aircraft preparing to attack the ship[2]. While the team in the ship's Combat Information Center that had the responsibility for air defense had very good information on air situation, the specific sequence of events that were observed were highly anomalous and did not fit previously-observed or trained examples. The tactical lead and team worked to explore narratives that could explain all the observed data and that could be used to form a decision strategy. The team became focused on only one specific narrative, in which the airliner was part of a deception masking an attack-the-ship intention. The then-current high level of military and geo-political tensions in the Persian Gulf provided elements of context for this narrative interpretation. Other recent events in military history, such as the sinking of the HMS Sheffield only six years earlier by a single air-surface missile strike, were also part of the shared context understanding that support this interpretation. This focus on that single narrative interpretation ultimately led the decision to attack the aircraft before it could reach position from which it could launch an air-surface missile. In this example, the context understanding was appropriately shared and appropriately acted-on—it just turned out to be an inaccurate interpretation of the GT situation.

This points to a deeper issue with shared context and distributed action, that of indeterminism (Freeman & Zachary, 2018). Even when a team has a highly overlapping understanding of the current context, any individual team member can take an action that can change the trajectory of the problem evolution for the team (and potentially for its adversary). From that point forward, the actions and decisions of the team and of each member can only be assessed in the light of that changed problem context as it was understood by the team.

The 1977 nuclear reactor accident at Three Mile Island (Cummings, 1980; Rouse et al., 1992) is an example of this. One operator missed a warning light and interpreted the reactor core state as the opposite of what it was. Acting accordingly, based on this perceived context, lead to a cascading set of team actions that were correct given that initial interpretation error and the direct action based on it. This issue is equally present when, as in the Vincennes case, the team has arrived at a context representation that is plausible but in the end factually divergent from the GT situation. Thus, the assessment of overall team and individual team-member actions and decisions can be *indeterminate* with regard to the GT of the situation, and can only be fully evaluated in terms of their (shared) context understanding. Moreover, these situations create meta-problems of requiring actions and communications to repair the correct understanding of the context, which are crucial to team resilience. However, detecting the existence of such meta-problems as well as assessing and training for them, cannot be undertaken without an explicit model of the context understanding of the

[2] This description is simplified here for brevity and classification purposes

team as well as a model of what an accurate context representation would be, given the conditions of the operational environment and the domain expertise of the team.

Managing the Role of Context in Team Training

In many modern team training systems or exercises, trainees interact through simulated work environments. During these interactions, teammates communicate with each other using verbal and computer-mediated methods such as email and text messages. While the trainees develop their individual and team-shared context understanding, human instructors typically function as observers and build their own representation of the team-shared context understanding. These instructors then provide assessments and feedback to the team through the lens of their own internal context representation. This process (which is also used in live training exercises) works, because the instructors/observers apply their domain knowledge and human context-processing ability, often combined with ground-truth situational data, to create a gold-standard context understanding that can be used to assess and instruct the learners. However, in building fully-automated team training tutors using future extensions of GIFT, this function will have to be automated itself. Is this possible?

Can We Build Context-Awareness into Team Trainers?

Our definition of context raises the question of whether context can be framed and modeled as an explicit computational process to be carried out by computational devices. This would be of fundamental importance for automatically generating both team-level and individual context awarness into Intelligent Team Training Systems (ITTS) developed from extensions of the current GIFT (Gilbert et al., 2017). The computational representation of context and the development of computable models of context is current research interest in computer science in artificial intelligence (Lawless & Mittu, 2017; Lawless et al., 2018). Zachary and colleagues (Zachary, Rosoff, Miller, & Read, 2013; Zachary et al., 2015) have developed a cognitive computational theory of context understanding and awareness called the Narratively-Integrated Multi-level (NIM) context framework that has also been successfully used to generate computer-models of context awareness (Zachary et al 2017; Zachary & Carpenter, in press). This computational context theory is outlined below as a potential way forward for context-inclusion in ITTS.

NIM combines cognitive theories of situational awareness and narrative reasoning while providing links to decision-making and planning via theories of RPD and of cognitive skills as discussed above. NIM defines a general structure for context as a computational process that includes three fundamental aspects:

- a *declarative representation* of momentary context awareness;
- a set of *constructive knowledge sources* that constantly build, abstract, modify, and deconstruct the declarative representation as on-going maintenance of the context awareness; and
- a set of *translation knowledge sources* that link patterns of information in the declarative context representation to decisions and actions that are stimulated by or must be adapted to the momentary context.

Research on the structure of mental models across domains also found that there is a consistent structure to (expert) mental models. Under the rubric of situation awarness, Endsley and colleagues (Endsley 1995, Endsley & Garland, 2000) identified three increasingly abstract levels that, across domains, were used by experts to provide context information. These are:

1) *Perception*, in which the person perceives the status, attributes, and dynamics of relevant elements in the situation and their current states,

2) *Comprehension*, in which the person understands how the perceived elements can impact his/her situational goals;

3) *Projection*, in which the person estimates the future potential actions of elements in the environment in near-time, and

4) *Expectations*, in which the person posits specific future events[3] involving objects or relationship in the situation that would or could be diagnostic for specific courses of action of decisions.

Inability to deal with situations involving adversarial interactions constitutes a key limitation of Endsley's situational awareness framework. Other research, however, focused on the high-level narrative or story-based representations that people use to structure, reason about, and plan for sequences of interactions (e.g., Bruner, 1991; Mataes & Sanger, 1999). Integration of these two approaches was hampered in large part because of the significant differences in their formal and computational foundations, until the NIM framework resolved the conflict by adding an additional level of abstraction above the three situational levels. This higher level of context links situations to story-oriented representation of context, via representational levels for:

5) *Action Units*, in which the person identifies specific sets of events, that can be viewed as instances of an elementary unit of some general;

6) *Story Units*, in which the person anticipates or recognizes representing specific sets of actions units that can be treated as instances of a building block of a general narrative. Called Story Units (see Miller et al., 2006) these are building blocks of narratives that the person is aware could be unfolding in the current situation; and

7) *Plausible Narratives*, in which the person identifies and tracks a general narrative against the set of Action and Story units that are expected and/or recognized. In some cases, people are able to identify and track multiple alterative plausible narratives, though is appears generally difficult for people to do this. On the other hand, identifying multiple alternative narratives is often an explicit goal of individuals or teams in certain domains, such as intelligence analysis, criminal investigation, or market analysis.

These higher levels of the declarative representation create distinct levels of understanding of the interaction involving human motivations and intentions that give meaning to the situational information in the lower levels. The general features of our NIM working context theory are pictured in Figure 1 below.

[3] An Event is a concept that interconnects specific agents or objects, relationships to other agents or objects, and specific temporal-spatial information. Its cognitive foundations are discussed in Barwise and Perry, 1983; Zacks et al. 2007; Radvansky and Zacks 2011).

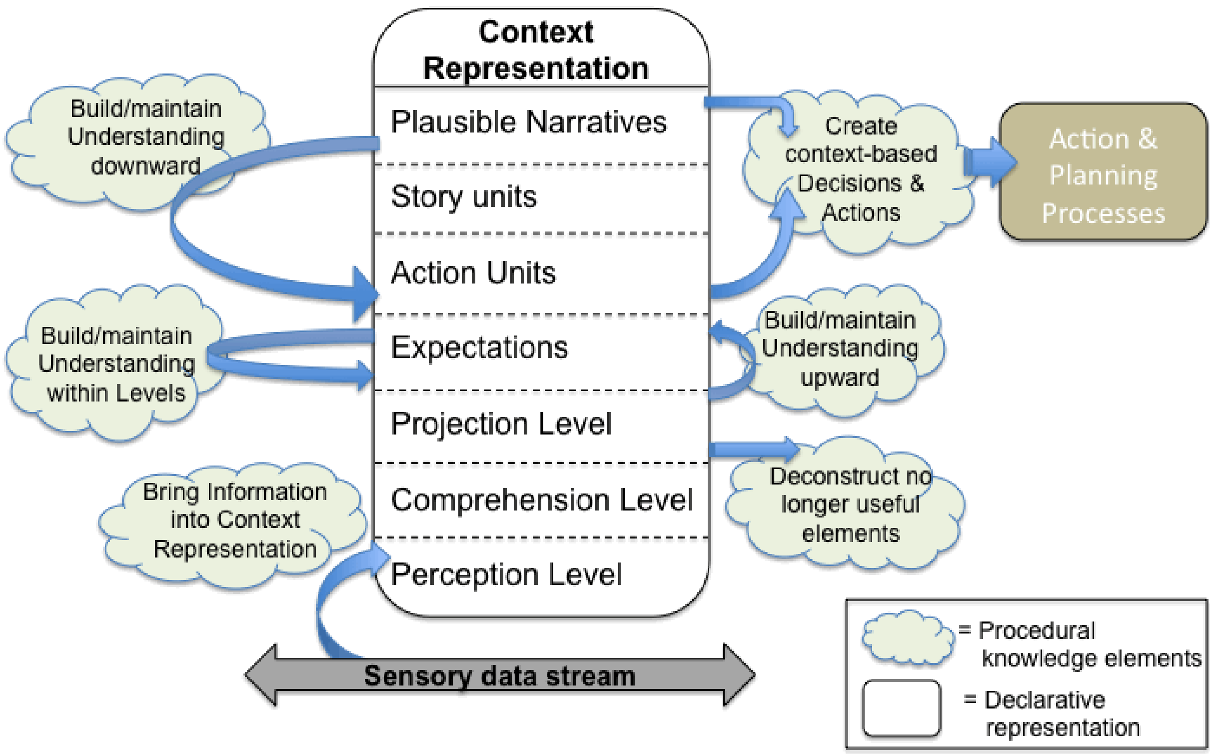


Figure 1. Context as cognitive computational process

Recommendations for GIFT

When considering an ITTS, context will play a critical role in managing the assessment space of a given scenario based on all of the interacting team-parts. The assessments will be used to infer skill and competency, and will ultimately be used to drive pedagogical decisions that aim at improving task performance and team behaviors. When considering the implementation of context oriented assessments within GIFT, there are multiple perspectives that must be addressed. In its current state, GIFT's assessment logic for practice-based activities is configured within a Domain Knowledge File (DKF). It is within the DKF where a developer establishes a set of tasks that have context dependent conditions that drives real-time performance tracking for managing the selection of instructional strategies. For each represented task in the DKF, there is a start- and end-trigger that determine when certain conditions are active and dormant based on the environmental conditions in relation to the scenario GT and direct behaviors observed by the trainee. In GIFT's current state, the DKF authoring schemas are relatively stable for individualized tasks, though this process can be quite complex. Introducing team elements can increase the complexity of this task substantially, though redundancies in the context representation across team role could be exploited to manage this complexity. .

From an architectural standpoint, GIFT needs to support multiple users interacting within a single domain space, where models are configured to track relevant states across both individual and team-oriented objective. Previous projects exploring this challenge have applied a hierarchical representation of DKFs (e.g., Gilbert et al., 2017). In this case, each individual team member has a DKF specifying their individual tasks within the scenario as a whole. Above that is a team-based DKF that tracks states relevant at the composite level for each specified group (e.g., team communication, cohesion, etc.). The performance and behavior tracked at the individual level is used at the team level to monitor team objectives across a number of competing objectives. The current issue with this approach is that each DKF operates independently and there is not a bi-directional component in terms of performance state reporting. In this instance, the performance state of an individual, as measured through their DKF, can influence assessments at the team-level, but that activity cannot re-orient the context of another team-member. It is in this fashion that GIFT's architecture must be modified to support context modeling in a bi-directional capacity that allows the context of individual tasks to be directly influenced by actions taken by another team-member. Naturally, these construct representations need to be added within GIFT's domain module where context models are applied across all trainee and team formalizations.

A rudimentary form of context model can be represented in GIFT's current state through well-established start- and end-trigger specifications for each task and objective being tracked. However this is far from ideal and probably far from the minimal required to address in the discussions above of the specific case examples. What is missing are context relevant constructs at the individual level that are used to infer team states, and mechanisms to capture and propagate persistent or transitory interdependencies between and among teammates. For example, the effects of one teammate's actions on a problem space may directly impact the context of another teammates objectives, with a communication requirement to dictate that change. In this instance, one teammate must perform specific duties and communicate critical information across team channels so the outcomes of tasks at one level can be accounted for at another level. For this purpose, GIFT needs to track successful communication practices across team members, which is a relevant team construct that needs to be represented in GIFT. However as noted earlier, determining whether the content of a communications is correct and situationally appropriate, or whether a needed communication did not occur, requires information only in the individual and team context representations. As a result, all the prior communications become (at least potentially) context for the assessment of the present and future communication. Importantly, these distinctions imply architectural relationships with the problem/mission simulation model. The context models must capture specific kinds of data to build and track the individual-role and team context representations. This process of context modeling could plausibly be implemented using a NIM context model framework.

With the architectural components in place, the next perspective that must be addressed is the authoring workflows to establish this context oriented construct as it relates to specific scenarios and embedded assessments. The content required to construct the context model must be established in a way that is flexible and agnostic to the domain or task within which it is being applied. An author needs to establish individual and team context representations in addition to individual procedural knowledge. An important aspect that must be addressed in establishing dependent and interdependent relationships as it pertains across each individual team member and their associated task responsibilities and objectives. This also requires establishing interaction models that dictate how context is shared and updated across all individual members of a team. This in essence establishes context update triggers that are used to modify context oriented content across all interdependent teammates.

With context established within domain level assessments, pedagogical considerations must be addressed next. With new assessment methods in place, new pedagogical approaches and scaffolding models must be explored that examine the incorporation of context relevant data and indeterminacy. This requires embedding scenario relevant information in feedback scripts through automated mechanisms. A potential approach is designing feedback templates that can incorporate data inputs as they are derived from a training environment in real-time. This enables contextualized feedback as it relates to the specific interacting elements within the environment, where captured data within log files can be parsed for specific variable information that relates to the assessment that triggered the remediation materials. In addition, team-level pedagogy should include context relevant information as it pertains to other members of the team who are dependent upon or impacted by that individual's interactions. Building out instructional strategies that address second- and third-order effects with context specific data will assist individuals in developing shared mental models that associate their role within a larger team context.

Conclusions

The construct of context representation is a critical component when considering the design of an ITTS capability. Computationally representing the interacting variables that constitute context is required to ensure assessment practices are appropriately linked across team members and account for dynamic changes in the interacting operational environment. In addition, research shows context representation improves with domain expertise, where establishing accurate representations of context should be a training objective that incorporates specific pedagogical interventions that target the identified skill sets. With a computational representation of context, assessment techniques can be investigated that target points of indeterminism, where an individual's perceived context does not match ground truth, where specific impasses or misconceptions in understanding can be better recognized. By incorporating a context layer in an ITTS's domain model, assessments and pedagogical interventions can be better managed based on dynamic changes in the task environment.

References

Barwise, J. & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry 18,* 1, 1-21

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1990). Cognitive psychology and team training: Training shared mental models and complex systems. *Human Factors Society Bulletin, 33*(12), 1-4.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In J. Castellan, Jr. (Ed.). *Individual and group decision making: Current issues* (pp. 221–246). Lawrence Erlbaum Assoc: Hillsdale, NJ.

Cummings, G. E. (1980). Operator/instrumentation interactions during the Three Mile Island incident. *IEEE Transactions on Nuclear Science, 27*(1), 931-934.

Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement*. Boca Raton FL: CRC Press.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors 37*(1),32-64.

Fiore, S. M., Wiltshire, T. J., Oglesby, J. M., O'Keefe, W. S., & Salas, E. (2014). Complex collaborative problem-solving processes in mission control. *Aviation, space, and environmental medicine, 85*(4), 456-461.

Freeman, J., & Zachary, W. (2018). Intelligent Tutoring for Military Team Training: Lessons learned from US Military Research. In Johnson, J., Sottilare, R., Sinatra, A., Burke, S (Eds.). *Building intelligent tutoring systems for teams: What matters*. Bingley, UK: Emerald Publishing.

Gilbert, S. B., Slavina, A., Dorneich, M. C., Sinatra, A. M., Bonner, D., Johnston, J., ... & Winer, E. (2017). Creating a team tutor using GIFT. *International Journal of Artificial Intelligence in Education,* 1-28. doi:10.1007/s40593-017-0151-2

Hampton, A., Shalin, V. L., Robinson, E., Simpson, B., Finomore, V., Cowgill, J., Moore, T., Rapoch, T., & Gilkey, R. (2012). The Impact of spatialized communications on team navigation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 463-467). SAGE Publications.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision Making in Action: Models and Methods, 5*(4), 138-147.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In *Proceedings of the Human Factors and Ergonomics Society annual meeting, (Vol. 30, No. 6,* pp. 576-580). SAGE Publications.

Lawless, W., Mittu, R., Sofge, D., & Morrison, J. (Eds.)(in press). *Computational context: The value, theory and application of context with AI*. Boca Raton, FL:CRC Press

Lawless, W., & Mittu, R. (Eds.). (2017). *Proceedings of the AAAI Symposium: Computational context: Why it's important, what it means, and can it be computed?*. Stanford, CA:American Association for Artificial Intelligence.

Levesque, L. L., Wilson, J. M., & Wholey, D. R. (2001). Cognitive divergence and shared mental models in software development project teams. *Journal of Organizational Behavior, 22*(2), 135-144.

Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior, 27*(4), 403-418.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273.

Mateas, M., & Sanger, P. (1999). Narrative intelligence. In *AAAI Technical Report FS-99-01*. New York: American Association for Artificial Intelligence.

Radvansky, G. A., & Zacks, J. M. (2011). *Event perception. Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 608-620.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics, 22*(6), 1296-1308.

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can turn a team of experts into an expert team? Emerging training strategies. In C. E. Zsambok, G. & Klein (Eds.) *Naturalistic decision making* (pp. 359-370). Mahwah, NJ: LEA.

Smith-Jentsch, K. A., Mathieu, J. E., & Kraiger, K. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology, 90*(3), 523.

Sottilare, R., Brawner, K., Sinatra, A., & Johnston, J. (2017). *An updated concept for a Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors, 41*(1), 61-71.

Suchman, L.A. (1987). *Plans and situated action: The problem of human-machine communication*, New York, NY: Cambridge University Press

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47(1),* 513-539.

Westli, H. K., Johnsen, B. H., Eid, J., Rasten, I., & Brattebø, G. (2010). Teamwork skills, shared mental models, and performance in simulated trauma teams: An independent group design. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 18*(1), 47.

Zachary, W, Carpenter, T., & Santarelli, T. (in press). Using computational context models to generate robot adaptive interactions with humans. In Lawless, W., Mittu, R., Sofge, D, and Morrison, J. (Eds.) *Computational context: The value, theory and application of context with AI. Boca Raton,* FL: CRC Press.

Zachary, W., & Carpenter, T. (2017). Cognitively-inspired computational context. In Lawless, W. and Mittu, R. (Eds). *Proceedings of the AAAI Symposium: Computational context: Why it's important, what it means, and can it be computed?* Stanford, CA: American Association for Artificial Intelligence.

Zachary, W., Rosoff, A., Miller, L., & Read, S. (2013). Context as cognitive process: An integrative framework for supporting decision making. In Laskey, K., Emmons, I., and Costa, P. (Eds). *Proceedings of 2013 semantic technologies in intelligence, defense and security* (pp. 48–55).

Zachary, W., Rosoff, A., Thomas, R., Iordanov, V., Read, S., & Miller, L. (2015). Toward computational, embeddable, cognitive models of context. *Procedia Manufacturing, 3*, 5269-5276.

Zachary, W., Ryder, J., & Hicinbothom, H. (2000). Building Cognitive Task Analyses and Models of a Decision-Making Team in a Complex Real-Time Environment. In Schraagen, J.M.C., Chipman, S.F., & Shalin, V.L. (Eds.). *Cognitive Task Analysis* (pp. 365-383). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Zacks J.M., Speer N.K., Swallow K.M., Braver T.S., & Reynolds J.R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin 133*,273–293.

Chapter 12 – Constructing Individual Conversation Characteristics Curves (ICCC) for Interactive Intelligent Tutoring Environments (IITE)

Xiangen Hu[1,2], Nia Dowell[3], Zhiqiang Cai[1], Arthur C. Graesser[1], Genghu Shi[1], Jody L. Cockroft[1], & Paul Shorter[4]
The University of Memphis[1], Central China Normal University[2], University of Michigan[3],
U.S. Army Research Laboratory[4]

Introduction

Most of existing intelligent tutoring systems (ITS) applications involve one or at most two interactive conversational avatars (CAs) and one student (Graesser et al., 2012; Hu, Morrison, & Cai, 2013). It is reasonable to consider a case where using multiple (two or more) avatars and multiple human learners are co-present in the same interactive intelligent tutoring environment (IITE). This chapter explores a computational model that assesses the quality of contributions of avatars and the verbal behavior of human learners during a sequence of turns in natural language interactions, where a "turn" in a conversation is the time one of the participants contribute. The total number of turns is the total number of contributions. This approach is a further development of an earlier chapter (Hu, Morrison, & CAI, 2013) in volume I of this series.

The collaborative exchanges we consider in this chapter follow some specific assumptions in the application scenarios, as specified below. However, we consider more general situations in the discussion section at the end. We consider a special IITE[4] where conversations between avatars and human learners are focused on a theme topic, with the following assumptions:

A1: Avatars and human learners are aware (have memories) of recent contributions of other avatars and human learners.

A2: Avatars and human learners are equal. For avatars, they are either independently controlled by ITS engines or collectively controlled by a single ITS engine. Here ITS engine refers to something similar to the AutoTutor Conversation Engine (ACE, Nye, Graesser, & Hu, 2014).

A3: Contributions from each of the conversation participants (human or avatar) are in the form of natural language (NL) and in written form with a sequential order of the contributions. Time intervals between conversations are not considered in the current model.

A4: There are no simultaneous contributions from two or more participants.

A5: There is a semantic encoding method that is capable of computing the semantic similarity between any collections of NL contributions. Such semantic similarity measure can be used to create a semantic classifier that can group contributions into semantically equivalent groups.

[4] Although IILE can be more general, in this paper, we specifically consider learning environments where conversations between avatars and human learners are focused on a theme topic

Relevant research questions

There is a growing literature that investigates learning and performance processes in groups, as in the case of collaborative learning, collaborative problem solving, and collaborative work (Fiore et al., 2010; Graesser et al., 2018; Salas, Cooke, & Rosen, 2008; Shaffer, 2017; Stahl, Roschmann, & Suthers, 2006; Stahl & Rosé, 2013; Suthers, Dwyer, Medina, & Satrap, 2010). This literature emphasizes sociocognitive group processes, such as coordination, common ground, elaboration, regulation and integration of ideas. In this chapter, we apply these well-developed theoretical lenses in the context of the special IITEs to explore a computational model that characterizes avatars and human learner's behavior.

Our approach builds on *Group Communication Analysis* (GCA), a methodology for quantifying and characterizing the discourse dynamics between learners in online multi-party interactions (Dowell, 2017; Dowell, Nixon, & Graesser, 2018 under-review). GCA applies automated computational linguistic analysis to the *sequential* interactions of participants in online group communication. GCA both captures the structure of the group discussion and quantifies the complex semantic cohesion relationships between learners' contributions over time, revealing intra- and interpersonal processes in group communication. In doing so, this methodology goes beyond previous models for automated group communication, which often rely on counting the number of utterances exchanged between learners.

The GCA framework incorporates definitions and theoretical constructs that are based on research and best practices from several areas where sociocognitive processes, group interaction, and collaborative skills have been assessed. These include areas such as computer-supported collaborative learning (CSCL), computer-supported cooperative work (CSCW), teams, organizational psychology, assessment in work contexts and the PISA 2015 collaborative problem-solving framework. Despite differences in orientation between the disciplines where these frameworks have originated, the conversational behaviors that have been identified as valuable are quite similar, and the GCA provides six learning-relevant interaction measures (summarized in Table 1), which are briefly reviewed below (Dowell et al., 2018).

Posting a message on the forums is often operationalized by researchers and instructors as participation (Hrastinski, 2008) and considered a requirement for any online learning group interaction. It signifies a willingness and readiness of learners to externalize and share information and thoughts (Hesse, Care, Buder, Sassenberg, & Griffin, 2015). Participation, has been shown to have a beneficial influence on various learning outcomes, including retention rates, learner satisfaction, and social capital (Hrastinski, 2008). GCA approaches participation as a necessary, but not sufficient component for characterizing the interactions between MOOC learners.

Internal cohesion is a sociocognitive measure that can serve as a proxy for individual self-monitoring and reflection processes during peer interactions. That is, successful collaboration requires that each individual monitor and reflect on their own knowledge and contributions to the group (Barron, 2000; OECD, 2013); a behavior explained within self-regulation theory (Chan, 2012; Malmberg, Järvelä, & Järvenoja, 2017; Zimmerman, 2001). Consequently, during peer-learning individuals need to appropriately build on and integrate their own views with those of the group (Kreijns, Kirschner, & Jochems, 2003). Given that a participant's current and previous contributions should be, to some extent, semantically related to each other, a measure of internal cohesion can indicate the extent to which they have monitored and reflected on their previous discourse (i.e. self-regulation). Overly high levels of internal cohesion might suggest that a participant is not evolving their thoughts, but rather reiterating the same static view. Conversely, low levels of internal cohesion might indicate that a participant has no consistent perspective to offer the conversation, and is echoing the views of others, or is only engaging at a surface level within discussion thread topics.

Learners must also monitor and build on the perspectives of their collaborative partners to achieve and maintain a shared understanding of the task and its solutions (Dillenbourg & Traum, 2006; Graesser, Dowell, & Clewley, 2017; Hmelo-Silver & Barrows, 2008; Stahl & Rosé, 2013). In the CSCL literature this shared understanding has been referred to as knowledge convergence, or common ground (Clark & Brennan, 1991; Roschelle & Teasley, 1995). It is achieved through communication and interaction, such as building a shared representation of the meaning of the goal, coordinating efforts, and viewpoints of group members, and mutual monitoring of progress towards the solution. Responsivity is a sociocognitive GCA measure, which captures monitoring and regulatory processes externalized during communication with peers. This measure reflects the extent to which an individual monitors and incorporates the information provided by the peers in their new contributions. The measure is implemented by examining the semantic relatedness between the individual's current contribution and the previous contributions of their collaborative partners. For example, if an individual's contributions are, on average, only minimally related to those of their peers, it would the individual exhibits low responsivity.

The GCA's social impact measure captures the extent to which a learners' contributions are seen as meaningful, or worthy of further discussion (i.e. uptake), by their peers. Social impact is measured through the analysis of the semantic relatedness between the learner's current contribution and those that follow from their collaborative partners. Individual messages that are more semantically related to the subsequent contributions indicate a high social impact of their authors on the unfolding group discourse.

Peer interactions provide the opportunity to expand the pool of available information, thereby enabling groups to reach higher quality solutions than could be reached by any one individual. However, despite the intuitive importance of (new) information sharing, a consistent finding from research is that groups predominantly discuss information that has been already shared (known to all participants) at the expense of information that has not been shared (known to a single member) (see Mesmer-Magnus & Dechurch, 2009 for a review). The distinction between given (old) information versus new information in discourse is a foundational distinction in theories of discourse processing (Price, 1981). Given information includes words, concepts, and ideas that have already been disclosed in the discourse; new information involves words, concepts and ideas that have not yet been mentioned, and builds on the given information or launches a new thread of ideas. The GCA captures the extent to which learners provide new information, compared to referring to previously shared information, with a measure called newness.

The team performance literature also advocates for concise communication between group members (Gorman, Cooke, & Kiekel, 2004). An example of this can be seen in formal teams, like military units, which typically adopt conventionalized terminology and standardized patterns of communication. It is suggested that this concise communication is possible when there is more common ground within the team and the presence of shared mental models of the task and team interaction (Klein, Feltovich, Bradshaw, & Woods, 2005). The GCA's communication density measure was first introduced by Gorman et al.(2003) in team communication analysis to measure the extent to which a team conveys information in a concise manner. Specifically, the rate of meaningful discourse is defined by the ratio of semantic content to number of words used to convey that content.

Using this approach, we would be able to tackle topics and questions to improve collaborative IITE, including:

1. Characterize and quantify participants' sociocognitive processes in IITEs.
2. Discover learner profiles or roles based on regularities in their communication patterns.
3. Understand how the individual sociocognitive measures and roles influence learning in IITEs.
4. Conduct fine-grain, real-time temporal analysis and assess the degree to which learners are occupying more or less productive roles.
5. Understand how role diversity and composition influence learner performance.

6. Create optimal group compositions in IITEs using multiple avatars and human learners.

Data Structure under Consideration

There are two ways that the sequential interactions of participant's data can be organized. One fits the traditional method such that the data is organized in the form of tables or a matrix. In this case, the data are a collection of L_{ent} of turn-based NL contributions from each of the N participants, $n=1, ..., N$ are the index of the participants, and $t=1, ...,T$ are the index of the total turn-based NL contributions of the participants. Assumption A4 where no simultaneous contributions from two or more participants indicates that only one none empty L_{ent} for any given t. The data under consideration will be equivalent to a matrix of N x T, where elements are either a turn-based NL contribution or empty. N is the number of participants, and T is the number of conversation turns. In a two-dimensional data matrix[5], we do not have information for the "target" of the speech.

In order to capture all the interactions in the space, we start by use xAPI statements (ADL, http://adlnet.github.io/) To capture all details of the sequential interactions of participants' data, including time of the contribution and the target of the speech, we use xAPI statements with four key components: "actor", "verb", "object", and "target" with a timestamp of the natural order of the contribution of the participants. Specifically, these four components are included in the basic information of an xAPI statement:

- *actor* is a unique identifier for the participant (avatars or humans).
- *verb* is a collection of speech act classifiers such as "answered", "asked", etc. They are a small subset of xAPI vocabulary (http://xapi.vocab.pub/verbs/index.html).
- *object* would be the contribution of the actor, in natural language, and
- *target* of the contribution, which is a subset of all participating avatars and human.

In xAPI statements, a four-dimensional array can be extracted, where the first dimension (*actor*) has possible of N values (N learners or avatars), the second dimension (*verb*) has one of K possible values[6](i.e., a speech act), the third dimension would be the contribution of the actor in NL, and the fourth dimension is *target.* It may have up to 2^N possible values[7]; the fifth dimension would be the timestamp. Consider that there are only a finite number of contributions considered in a given IITE, we assume the fifth variable is discrete and is used only for the index of contribution order. An example of an xAPI statement is attached in the appendix.

Individual Conversation Characteristic Curve (ICCC)

We introduce Individual Conversation Characteristic Curve (ICCC) as mathematical model for each of the participants in an IITE. It can be used to compute each of the GCA measures at any given turn. ICCC for any of the participants is a sequence of T six-dimensional real valued vectors, where the six dimensions correspond to the six GCA measures. To help the rest of the discussions, we introduce the following notations: For each of the xAPI statement, S_t, there are the following critical elements:

- $p_t \in \{p_1,...p_N\}$, participant who made contribution at time t,
- v_t: speech act of p_t ,
- l_t: the NL language contribution made p_t ,

[5] call it matrix, but entries are not numerical values. They are NL contributions.
[6] This would be the list of verbs included in the xAPI vocabulary.
[7] Consider the possible case that user self elaborates so the object would the "self".

- $g_t \subseteq \{p_1,...p_N\}$, subset of participants that are *target* of p_t contribution,
- t: the timestamp of p_t contribution.

Additional values from xAPI statements at or before time t:

- Cumulative contribution for p_n is $L_{ent} = \cup\{l_i \mid p_n = p_i,\ i=1,...,t\}$.
- Cumulative target of p_n is $G_{n,t} = \cup\{g_i \mid p_n = p_i,\ i=1,...,t\}$.
- Cumulative contributions for any participant are $U_t = \cup\{L_{n,t} \mid n=1,...,N\}$.

For any t, the t^{th} vectors is computed from a t^{th} xAPI statement in the learning record store (LRS) and vectors are computed from earlier turns. Each of the six GCA dimensions *A* (*Participation*), *B* (*Responsivity*), *C* (*Internal Cohesion*), *D* (*Social Impact*), *E* (*Newness*), and *F* (*Communication Density*), all are computed numerically and indexed by the contribution order, *t*. Assume a given IITE involves N participants, $p=\{p_1,...p_N\}$, each time *t, one (and only one)* of participants contribute L_t addressing to a *subset* T_t of participants.

Constructing ICCC

With the notations introduced above, we outline algorithms for computing ICCC from xAPI statements. We consider participant p_n ($n=1,...,N$). Let ($A_{n,t}$, $B_{n,t}$, $C_{n,t}$, $D_{n,t}$, $E_{n,t}$, $F_{n,t}$) be the t^{th} vector of ICCC for p_n. With (p_t, v_t, l_t, g_t, t) in each of the LRS statements and derived values L_{ent} and $G_{n,t}$ at or before time t, values of ($A_{n,t}$, $B_{n,t}$, $C_{n,t}$, $D_{n,t}$, $E_{n,t}$, $F_{n,t}$) can be constrained, in the following qualitative fashion. Consider any participant p_n at time *t:*

Table 1: GCA Measures and Computations

Measure	Description	Computation
Participation	Mean participation of any participant above or below what you would expect from equal participation in a group of the size of theirs	$A_{n,t}$ increase in two cases: ● if p_n is the participant at p_t, namely, $A_{n,t} \geq A_{n,(t-1)}$, if $p_t = p_n$. ● if p_n is one of the target of other participant, namely $A_{n,t} \geq A_{n,(t-1)}$ if $p_n \in g_t$ and $p_t \neq p_n$ In addition $A_{n,t}$ may also be a function of semantic density of l_t. which is also used for other GCA measures.
Overall Responsivity	Measure of how responsive a participant's contributions are to all other group members' recent contributions	$B_{n,t}$ is a function of two cases, if p_n is the participant at p_t, ● the semantic overlap between l_t and $U_{(t-1)}$, namely, $B_{n,t}$, is a monotonic function of semantic similarity between $U_{(t-1)}$ and l_t, ● g_t is non-empty. namely, $B_{n,t} \geq B_{n,(t-1)}$ if $p_t = p_n$ and $g_t \neq$ ⍰.
Internal Cohesion	How semantically similar a participant's contributions are with their own recent contributions	$C_{n,t}$ is a monotonic function of semantic similarity between $L_{n,(t-1)}$ and l_t if $p_n \in g_t$. Namely, $C_{n,t} = f(S(L_{n,(t-1)}, l_t))$, if $p_n \in g_t$, where S is the semantic similarity measure between $L_{n,(t-1)}$ and l_t, and f is a monotonic function.
Social Impact	Measure of how contributions initiated by the corresponding participant have triggered follow-up responses	$D_{n,t}$ is related to three factors*:* ● $D_{n,t}$ is a monotonic function of $S(L_{n,(t-1)}, l_t)$ when $p_t \neq p_n$ ● $D_{n,t}$ increases if $p_t \neq p_n$ and $p_n \in g_t$. ● $D_{n,t}$ increases if $Gn_{,(t-1)} \subset G_{n,t}$.
Newness	The amount of new information a participant provides, on average	$E_{n,t}$ is a monotonic decreasing function of $S(L_{n,(t-1)}, l_t)$ if $p_t = p_n$.
Communication Density	The amount of semantically meaningful information	$F_{n,t}$ is a value is computed as a function of semantic density of l_t.

To compute all GCA measures for a participant, $A_{n,t}$ and $B_{n,t}$ depend on the frequency of contributions and frequency of being targeted in other participant's contributions. All six measures either partly or entirely depend on content of the contributions. For example, the values $S(U_{(t-1)}, l_t)$, $S(L_{n,(t-1)}, l_t)$ used for for $B_{n,t}$ $C_{n,t}$ $D_{n,t}$ $E_{n,t}$ are the semantic similarity between $L_{n,(t-1)}$ and l_t. *Semantic density* contributes to the computation of $A_{n,t}$ and $F_{n,t}$. While frequency of contribution and frequency of being targeted in other's contributions requires no further clarification, we observe that semantic analytic methods are essential for the computation of ICCC. We will use the next section to focus on semantic analysis.

Semantic Representation and Analysis (SRA)

The task of semantic analysis for computing ICCC vectors in this chapter is simple. One needs to be able to represent Cumulative contribution for p_n, namely, $L_{n,(t-1)}$ and p_t's current contribution l_t and compute their semantic similarity. This task requires (1) semantic encoding of a collection of sentences in each turn and (2) numerically computing the similarity between any two sets of sentences. There are a number of methods in computational linguistics (Riordan & Jones, 2011), especially corpus linguistics developed for the above task. For example, the well-known Latent Semantic Analysis (LSA, Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Landauer, Foltz, & Laham, 1998) represent semantics of words, sentences, and documents in the form of numerical vectors and semantic similarity between any two items (word, sentence, or document) are computed as the cosine of the corresponding vectors of the two items. Similarly, High dimensional Analogue of Language (HAL, Burgess & Lund, 1995), Latent Dirichlet Allocation (LDA, Blei, Ng, & Jordan, 2002; Steyvers & Griffiths, 2007; Steyvers, Smyth, Rosen-Zvi, & Griffiths, 2004) achieves a similar goal but is based on a different algorithm for extracting semantic representation vectors and different computational algorithms for computing semantic similarity. Instead of using specific semantic encoding method in computing ICCC vectors, we chose to consider a generic semantic analytic framework that are applicable to most of the existing approaches. This framework is called Semantic Representation and Analysis (SRA, Hu et al., 2014). SRA is based on three basic assumptions that are critical to existing vector-based semantic encoding methods such as LSA, HAL, and LDA. The three assumptions are:

1. *Hierarchical*: Semantics of different levels of a language entity may be represented differently.
2. *Algebraic*: The semantics of any level of language entities must be capable of being represented numerically or algebraically.
3. *Computational*: (The semantic representations of a higher-level language entity are computed as a function of semantic representations for its lower-level language entities. At the lowest level of language entities, a numerical semantic comparison measure must exist between any two items (e.g., words).

The above three assumptions, especially the third assumption, make it possible to introduce the concept of *Induced Semantic Structure* (ISS) at the lowest level of the semantic space. Intuitively, ISS can be understood as the *nearest neighbor* (NeNe) that researchers uses to judge appropriateness of a semantic space. Given a semantic space, ISS is an L x L matrix where W is the number of tokens[8] in a given language. The element (*i*,*j*) in such a matrix will be the semantic similarity between token *i* and token *j*. The numerical values for each of the elements (*i*,*j*) serves several roles in addition to being elements in a Matrix. They are used to derive the NeNe for each of the tokens. Given any row *i*, the values are semantic similarities between token *i* and all other tokens. It is straightforward to create order of the numerical values of the semantic similarity to obtain NeNe of token *i* with respect to other tokens. Since the numerical values of semantic similarity are only used to *derive* the ordering in NeNe (See Table 2 as an example of NeNe). NeNe no longer *explicitly* contains detailed semantic processing details, such as vector representation and semantic similarity computations. With this special property, one can ask the following intuitive questions about similarity of any two tokens: what is the overlap between the top *N* NeNe of token *i* and *j*? How does the overlap vary when one changes the value of *N*? These intuitions helped Hu et al. (Hu et al., 2014) introduce alternative semantic similarity measurements within the SRA framework. We propose to use those measures of semantic similarity in computing ICCC, in particular, the *combinatorial semantic similarity* between two tokens (Landauer, Foltz, & Laham, 1998)

[8] Also the smallest language entities in a given language. In the case of English, a typical token is a word, but could also be meaningful symbols.

Table 2: NeNe for "life" from LSA Spaces Generated from TASA Corpus for Three Grades.

6th	9th	12th
life	life	life
reincarnation	contemplated	death
premiums	reincarnation	lifetime
policyholder	sai	hamlin
premium	pipal	pipal
sai	nirvana	nirvana
cycles	lifetime	zarathustra
holdover	death	ahuramazda
condemning	hinduism	ahriman
chekhov	afterlife	policyholder
captial	excerpted	romantics
pipal	ribman	essayists
nirvana	reaffirm	sai
hinduism	militarily	beaumarchais
span	kindless	1658
priori	condemning	pseudonym
maturity	premiums	poquelin
immoral	premium	ribman
humane	policyholder	kindless

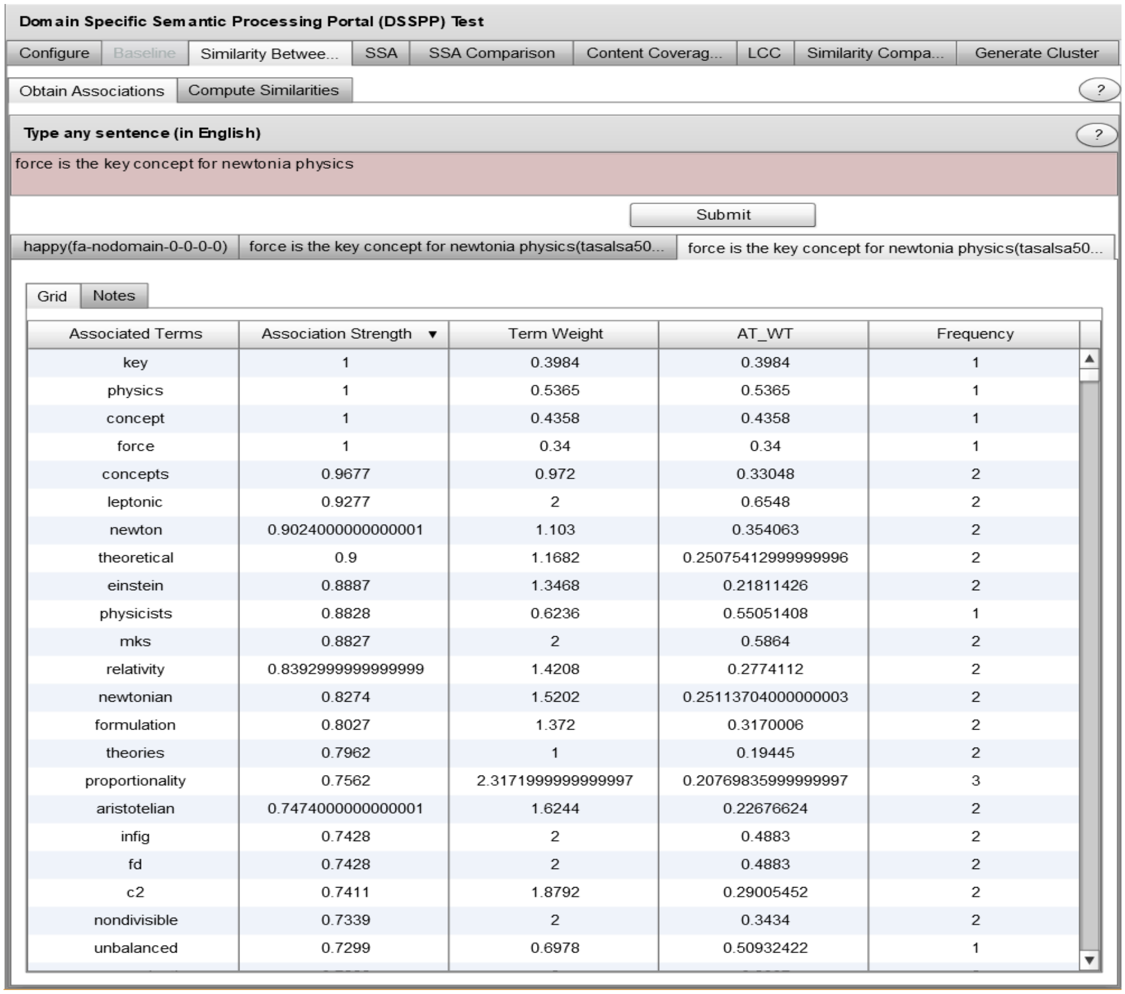
Combinatorial Semantic Similarity

Combinatorial semantic similarity (CSS) is one of three similarity measures introduced in SRA framework (Hu et al., 2014). CSS between any two tokens are computed in the following steps:

- Obtain NeNe from each of the tokens. It is possible that the NeNe are derived from two completely different semantic encoding algorithms.
- Choose a large integer *N* (can be any positive integer). Denote C_1={x| x is in the top N NeNe of Token 1}, C_2={x| x is in the top N NeNe of Token 2}.

The Combinatorial semantic similarity with top-NeNe is computed as simple set operation of C_1 and C_2: and is computed as the number of tokens in the *join* of C_1 and C_2 divided by the number of tokens in the *union* of C_1 and C_2.

CSS can be used to compute $S(L_{n,(t-1)}, l_t)$ for a participant's *Internal Cohesion* and *Social Impact* at time *t*. With the SRA framework within the ISS, for any collection of tokens (such as a sentence), additional measures can be obtained. For example, Figure 1 shows NeNe of a sentence "Force is a key concept in Newtonian physics". Not only does it list the NeNe, but it also has the term weights of the all the associated tokens. The values can be used to measure *Communication Density*. A simple measure would be to use the following algorithm: If l_t contains K tokens, then the sum of the top c*K NeNe term weights would be used to compute of $F_{n,t}$, namely, $F_{n,t}$ is monotonically related to the term weight of top c*K NeNe, where c can be an integer greater than 1.



Associated Terms	Association Strength	Term Weight	AT_WT	Frequency
key	1	0.3984	0.3984	1
physics	1	0.5365	0.5365	1
concept	1	0.4358	0.4358	1
force	1	0.34	0.34	1
concepts	0.9677	0.972	0.33048	2
leptonic	0.9277	2	0.6548	2
newton	0.9024000000000001	1.103	0.354063	2
theoretical	0.9	1.1682	0.25075412999999996	2
einstein	0.8887	1.3468	0.21811426	2
physicists	0.8828	0.6236	0.55051408	1
mks	0.8827	2	0.5864	2
relativity	0.8392999999999999	1.4208	0.2774112	2
newtonian	0.8274	1.5202	0.25113704000000003	2
formulation	0.8027	1.372	0.3170006	2
theories	0.7962	1	0.19445	2
proportionality	0.7562	2.3171999999999997	0.20769835999999997	3
aristotelian	0.7474000000000001	1.6244	0.22676624	2
infig	0.7428	2	0.4883	2
fd	0.7428	2	0.4883	2
c2	0.7411	1.8792	0.29005452	2
nondivisible	0.7339	2	0.3434	2
unbalanced	0.7299	0.6978	0.50932422	1

Figure 1. Demonstration of SRA. The NeNe of a Sentence. The Third Column Lists the Term Weight of the Associated Tokens.

Conclusions

In this chapter, we considered basic data structure from an IITE where multiple participants contribute in a group discussion. We propose to use xAPI statements to capture the dynamic nature of the interaction and provide basic algorithms for computing individual GCA measurements for each of the participants as function of the contribution order (time). The change of GCA values from time to time characterize each individual's interaction with the team. Previous studies use GCA for group as post-hoc evaluation of the team and individuals. With the ICCC proposed in this Chapter, we are able to evaluate team and individuals in real time. With real time evaluation, it is possible to monitor and regulate IITE to maximize the effectiveness of group discussion in learning/training. For computational purposes, we have also recommended a general semantic analytical framework for computing semantic similarity between the contributions of the participants.

Earlier of this chapter, we have outlined six relevant research topics and questions for this work. To tackle these topics and questions we would suggest using the proposed GCA-based ICCC approach to characterize and quantify learners sociocognitive processes in real-time. Such information can then be fed back into the

IITE and provide support in cultivating learners with self-regulatory and collaborative skills. This feedback could be presented as the individual GCA measures or temporal cluster analysis of the GCA based ICCCs could be used to depict a more holistic learner engagement profiles. Both of these feedback options would allow learners to assess the degree to which they are occupy more or less productive roles within the conversation, and adjust their contributions accordingly. Further, real-time ICCC information can be used to create optimal group compositions and participant diversity, by altering the CA's sociocognitive interaction approach through the GCA measures.

Recommendations for Future Research

Below we have provided a list of research and development questions for the future, especially for GIFT-like framework for team learning environments.

1. With ICCC, namely, GCA measurement for each individual immediately after his/her/its contribution, we may be able to ask theoretical questions that the previous application of GCA (mostly post-hoc) could not. For example, how likely is an individual (human) sensitive to the change of their GCA measurements during group interaction? How effective are the GCA-based feedback contributions? How can we maximize team learning with balanced GCA distributions for each individual participant?
2. In this Chapter, we outline our algorithms in qualitative fashion. What are the best computational formulae that are consistent with the qualitative constraints?
3. We have used SRA as generic framework for semantic similarity computation. In real applications, we need to either use one such as LSA or a combination of several computations. How can we select the best semantic engine to perform the semantic similarity computation?
4. We have proposed to use xAPI statements to capture the dynamic group interaction in team-based learning environments. However, in the existing xAPI vocabulary, there are only limited verbs that are not covering variety of speech acts in group discussions. Should we consider a standard set of vocabularies for xAPI statements for group learning/training?
5. In this chapter, we generically consider that members of IITE are participants in the conversation (human or avatar). In the case of intelligent tutoring systems (ITS), it is conceivable that some of the participants may not be a machine rather than a human. The question is how to maximize the role of ITS participants when they are blended with a group of human learners?

References

ADL. (n.d.). adlnet.github.io. Retrieved May 30, 2018, from http://adlnet.github.io/

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, *9*, 403–436.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2002). Latent dirichlet allocation. In *Advances in neural information processing systems* (pp. 601–608).

Burgess, C., & Lund, K. (1995). Hyperspace analogue to language (HAL): A general model of semantic memory. In *annual meeting of the Psychonomic Society, Los Angeles*.

Chan, C. K. K. (2012). Co-regulation of learning in computer-supported collaborative learning environments: a discussion. *Metacognition and Learning*, *7*, 63–73.

Clark, H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*, 391.

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *Journal of the Learning Sciences*, *15*, 121–151.

Dowell, N. M. (2017). A computational linguistic analysis of learners' discourse in computer-mediated group learning environments. *Dissertation*, *University of Memphis*, 1–195.

Dowell, N. M., Nixon, T., & Graesser, A. C. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multi-party interactions. *[revise and Resubmit] Behavior Research Methods*. Retrieved from https://arxiv.org/abs/1801.03563

Fiore, S. M., Rosen, M. A., Smith-Jentsch, K. A., Salas, E., Letsky, M., & Warner, N. (2010). Toward an Understanding of Macrocognition in Teams: Predicting Processes in Complex Collaborative Contexts. *Human Factors*, *52*, 203–224.

Gorman, J. C., Cooke, N. J., & Kiekel, P. A. (2004). Dynamical perspectives on team cognition. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *48*, 673–677.

Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). Evaluation of latent semantic analysis-based measures of team communications content. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *47*, 424–428.

Graesser, A. C., D'Mello, S., Hu, X., Cai, Z., Olney, A., & Morgan, B. (2012). AutoTutor. In &. C. B. P. M. McCarthy (Ed.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 169–187). IGI Global.

Graesser, A. C., Dowell, N. M., & Clewley, D. (2017). Assessing Collaborative Problem Solving Through Conversational Agents. In A. A. Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative Assessment of Collaboration* (pp. 65–80). Springer International Publishing.

Graesser, A. C., Dowell, N. M., Hampton, A., Lippert, A. M., Li, H., & Shaffer, D. W. (2018). Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 Program for International Student Assessment. In J. J. Johnston, R. Sottilare, A. Sinatra, & C. S. Burke (Eds.), *RMGT 19: Building intelligent tutoring systems for teams: What matters.* West Yorkshire, UK: Emerald Publishing.

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Springer Netherlands.

Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and Instruction*, *26*, 48–94.

Hrastinski, S. (2008). What is online learner participation? A literature review. *Computers & Education*, *51*, 1755–1765.

Hu, X., Morrison, D. M., & Cai, Z. (2013). Conversation-based Intelligent Tutoring System. In R. Sottilare, A.C. Graesser, X. Hu., & H. Holden (Ed.), *Design Recommendations for Intelligent Tutoring Systems: Learner Modeling* (Vol. 1, pp. 97–110). Orlando: U.S. Army Research Laboratory.

Hu, X., Nye, B. D., Gao, C., Huang, X., Xie, J., & Shubeck, K. (2014). Semantic Representation Analysis: A General Framework for Individualized, Domain-Specific and Context-Sensitive Semantic Processing. In *Foundations of Augmented Cognition. Advancing Human Performance and Decision-Making through Adaptive Systems* (pp. 35–46). Springer, Cham.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground coordination in joint activity. In *Organizational Simulation* (pp. 139–184). John Wiley & Sons, Inc.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*, 335–353.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, *49*, 160–174.

Mesmer-Magnus, J. R., & Dechurch, L. A. (2009). Information sharing and team performance: a meta-analysis. *The Journal of Applied Psychology*, *94*, 535–546.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, *24*, 427–469.

OECD. (2013). *PISA 2015 collaborative problem solving framework*. Oxford, U.K: OECD Publishing.

Price, E. F. (1981). Toward a taxonomy of given/new information. In P. Cole (Ed.), *Radical Pragmatics*. New York, NY: Academic Press.

Riordan, B., & Jones, M. N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*, 303–345.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem- solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 67–97). Berlin: Springer-Verlag.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: discoveries and developments. *Human Factors*, *50*, 540–547.

Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.

Stahl, G., Roschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge: Cambridge University Press.

Stahl, G., & Rosé, C. P. (2013). Theories of Team Cognition: Cross-Disciplinary Perspectives. In E. Salas, S. M. Fiore, & M. P. Letsky (Eds.), *Theories of team cognition: Cross-disciplinary perspectives* (pp. 111–134). New York, NY: Routledge.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. *Handbook of Latent Semantic Analysis*, *427*, 424–440.

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 306–315). ACM.

Suthers, D. D., Dwyer, N., Medina, R., & Satrap, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, *5*, 5–42.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: theoretical perspectives*. Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Appendix

xAPI statements for *actor, verb, object, context* for generating ICCC.


```
“actor”: {
            "mbox": "mailto:JohnDoe@emailaccount.com",
           "objectType": "Agent",
           "name": "John Doe”
           },
“verb”: {
           “id”: “http://adlnet.gov/expapi/verbs/answered”,
           “display”: {
           “en-US”: “answered”
           }
},
“object”: {
           “objectType”: “Activity”,
           “id”: “http://www.example.com/activities/discussion”,
           “definition”: {
                      "extensions":{
                                 "http://www.example.com/discussion": {
                                            "timeTaken": 675,
                                            "timeStart": "2018-01-16T16:18:57.581Z",
                                            "contribution": "Making very good sense!"
                                 }
                                 "http://www.example.com/answerto": {
                                            "member1": {
                                                       "mbox": "mailto:team_member1@emailaccount.com",
                                                       "objectType": "Agent",
                                                       "name": "Team Member 1”
                                            }
                                 }
                                 "http://www.example.com/target": {
                                            "member1": {
                                                       "mbox": "mailto:team_member1@emailaccount.com",
                                                       "objectType": "Agent",
                                                       "name": "Team Member 1”
                                            },
                                            "member2": {
                                                       "mbox": "mailto:team_member2@emailaccount.com",
                                                       "objectType": "Agent",
                                                       "name": "Team Member 2”
                                            }
                                 }

                      }
           },
}
```

Section III - Socio-Cultural Applications

Dr. Anne M. Sinatra, Ed.

CHAPTER 13 – SOCIO-CULTURAL APPLICATIONS

Anne M. Sinatra
US Army Research Laboratory

Core Ideas

The chapters in this section focus on the ***socio-cultural*** aspects of team tutoring in intelligent tutoring systems (ITSs). Team interactions are inherently social in nature. Whether teams are communicating with each other through typed messages, verbally, or using hand signals, communication itself is vital to being able to perform a team task. The amount, type of, and focus of team communications may differ based on the domain or tasks that are being engaged in. However, it is important for any ITS for teams to have approaches to deal with team communications and to account for the subtleties of social interactions that occur between teammates.

While team communication is particularly difficult to deal with in real-time for ITSs, it is important to put considerations into place about how a team may communicate and interact with each other. Even if an ITS cannot semantically process the meaning of typed communication between teammates in real-time, the communication may be vital for them to complete their team task. Therefore, it may be important to allow communication to happen and for socio-cultural interactions to occur, even if it cannot be used for real-time assessment, but instead to facilitate the task performance. As demonstrated in the chapters within this section, an understanding of the social interactions that occur between teammates and the change in dynamics when approaching a novel task can be important to assess even after the fact, and can lead to improvements in the way that researchers and educators structure their ITSs for teams. Since the Generalized Intelligent Framework for Tutoring (GIFT) is a domain-independent framework, the approaches that are put in place for authoring team tutors should account for and allow for varying socio-cultural interactions, including those discussed in the chapters in this section. Namely, the types of teammates should be flexible (e.g., all human, mix of human and computer-based), the types of communication should be flexible, consideration should be given to how to semantically process communication, and there should be an understanding that team tasks may change dynamically such that the people within the team may shift tasks/responsibilities as needed.

The chapters in this section collect a number of different issues and considerations that are important to the socio-cultural aspects of team tutoring. As demonstrated in the chapters in this section, while a team will need to communicate with each other, the composition of the types of team members may differ greatly. There may be teams that are comprised of humans, or a mix of human and computer-based teammates. Some of the issues that arise with a mix of types of teammates are the feelings that the human teammates have toward non-human teammates. This issue has been captured in the following chapters through the discussions of sentiment towards non-human teammates and through the use of trialogues. Additionally, based on the task that the team is engaging with, the team itself, or the demands of the task may change abruptly, and the team will be required to adapt accordingly in dynamic situations. Further, as social interactions in a computer-based team tutoring situation are mainly based on verbal and typed communication, it is important to have approaches to interpret and assess this information in real-time as well as near-real time. All of the chapters in this section approach different aspects of team socio-cultural interactions, while providing suggestions and improvements on how the issues can be dealt with in an ITS for teams.

Individual Chapters

The chapter by ***Hao, Zapata-Rivera, Graesser, Cai, Hu and Goldberg***, "Towards an Intelligent Tutor for Teamwork: Responding to Human Sentiments" describes a collaborative problem solving activity between two individuals who work together in a simulation task and interact with virtual agents. For their research, they collected data from a large number of teams using Amazon Mechanical Turk and examined the conversations that occurred between the team members. The sentiment associated with the conversations were coded. Several causes of negative sentiment were determined, and the authors provide suggestions on how to reduce negative sentiments toward virtual agents in future ITSs. The authors point out that team tutoring provides an opportunity for researchers to collect conversational information to determine what the individuals think of the agents that they are interacting with, and that this information can be leveraged to improve implementations of agents in tutors.

The chapter by ***Gorman, D'Mello, Stevens, and Burke***, "Characteristics and Mechanisms of Team Effectiveness in Dynamic Environments" discusses the common team dynamics that are present in diverse dynamic environments. There is special emphasis put on the fact that real-world task environments can frequently change and that novel situations may be encountered. The authors provide a description of effective team strategies, and how a team may adjust their communication and behaviors based on novel events that occur, as well as changes that may occur in team membership. Further, the characteristics of effective teams in dynamic environments are discussed.

The chapter by ***Cai, Hampton, Graesser, Hu, Cockroft, Shaffer, and Dorneich***, "Roles of Talking Agents in Online Collaborative Learning Environments" discusses the benefits of including conversational agents within collaborative learning environments. The authors provide examples and summaries of different types of environments and talking agents that have been created and used with ITSs through the years. The benefits of using these agents in ITSs are discussed, and recommendations are provided on how they could be implemented in GIFT.

The chapter by ***Foltz***, "Automating the Assessment of Team Collaboration through Communication Analysis" discusses approaches that have been used for analyzing team communication that happens during team tasks. A communication analysis pipeline is described, and the process of speech recognition and content modeling is explained. The book chapter describes the different domains that the approaches have been used in, and how it can be leveraged to predict performance. The chapter discusses the benefits of having near-real time analysis and modeling of communication in teams.

Chapter 14 – Towards an Intelligent Tutor for Teamwork: Responding to Human Sentiments

Jiangang Hao[1], Diego Zapata-Rivera[1], Arthur C. Graesser[2], Zhiqiang Cai[2], Xiangen Hu[2], & Benjamin Goldberg[3]
Educational Testing Service[1], University of Memphis[2], U.S. Army Research Laboratory[3]

Introduction

An intelligent agent can play a significant role in interactive learning, assessment, and teamwork (Baylor, 2011; Johnson, Phillips, & Chase, 2009; Chou, Chan, & Lin, 2003; Johnson & Lester, 2016; Kumar, Ai, Beuth, & Rosé, 2010; Moreno, Mayer, Spires & Lester., 2001; Schroeder, Adesope, & Gilbert, 2013). Like Barrón-Estrade, Zatarain-Cabada, Oramas-Bustillos and González-Hernández (2017) and Louwerse, Graesser, Lu, and Mitchell(2005), we believe that a good intelligent agent system should not only provide reasonable content responses to humans but also appropriately react to human opinions (e.g., sentiments, attitudes) about the virtual agent's responses. Therefore, we think that measuring and using human opinions, particularly negative ones, about the virtual agent's responses in real time can play an important role in developing a better responding mechanism.

It is challenging to obtain human opinions about the virtual agents in real time for the intelligent tutoring system (ITS) with a single human participant. Post-task surveys are generally used to evaluate how satisfied humans are with the virtual agents at the end of the task. Think-aloud protocols are used to probe how humans interact with the virtual agent in real time. However, the post-task survey approach cannot capture real-time opinions of humans toward the virtual agents, whereas the concurrent think-aloud approach is often too intrusive and may get in the way of eliciting the real opinions from the human participants. As such, there is not much opportunity to capture the evidence of human opinions against the virtual agent in real time.

However, when we extend the ITS to support multiple human participants, e.g., team work, it becomes possible to capture human participants' opinions towards the virtual agents in real time because these opinions may be embedded in the communications among the human participants. Understanding what causes negative sentiments and what causes positive sentiments against the virtual agents allows researchers to improve their design of how the virtual agents should interact with the team members.

In this chapter, we show how we measured human opinions about a virtual agent's responses in real time and explored the possible causes within a problem-solving task. We then identified agent-created situations that are likely to induce negative sentiments and developed a virtual agent responding mechanism in order to achieve a better ITS for teamwork. In our collaborative problem-solving task, two humans collaborate via text chat to work on a simulation task about volcano science, in which they interact with two conversational virtual agents (Hao, Liu, von Davier & Kyllonen, 2015; Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015). We chose this task as an illustration of our exploratory work because the human-human conversations contain rich information that can be mined regarding their sentiments toward the virtual agents.

METHODS

Simulation Task

The simulation task used in this study is part of the ETS Collaborative Science Assessment Prototype (ECSAP, Hao, Liu, von Davier, & Kyllonen, 2017b) and has been modified from a single-user simulation, the Volcano Trialogue. The Volcano Trialogue simulation was implemented using authoring tools such as conversation-space diagrams (Zapata-Rivera, Jackson, & Katz, 2015) and a version of the AutoTutor Script Authoring Tool (Cai, Graesser, & Hu, 2015) that focuses on assessment, called ASATA. In the Volcano Trialogue simulation, a human participant interacts with two virtual agents (see Graesser, Li, & Forsyth, 2014), one as a peer student (Art) and another as a professor (Dr. Garcia), to complete a task on volcano science (Zapata-Rivera et al., 2014). In the collaborative version of the simulation, Tetralogue (Hao et al., 2015; Liu et al., 2015), there is one additional human participant being added to the task. The two human participants collaborate (via text-mediated chats) with each other to interact with the two virtual agents to complete the volcano science task. We show a screenshot of the Tetralogue simulation in Figure 1.

Figure 1. Tetralogue Simulation

All of the turn-by-turn conversations and timestamped responses to the questions were recorded in a carefully designed log file (Hao, Smith, Mislevy, von Davier & Bauer, 2016). The conversations were used to measure collaborative problem solving (CPS) skills (Liu et al., 2015), while the responses to the in-simulation science items were used to measure science inquiry skills. Among the conversational turns between the human participants, there are many that express their opinions towards the virtual agents. These opinion-bearing communication acts provide an opportunity for us to understand what opinions human display under what circumstance. The general procedure to leverage this is straightforward. We first develop an automated sentiment detector that can detect the sentiments among the human conversations related to the virtual agents. Then we identify the situations that cause negative sentiments towards the virtual agents. Based on this, we can develop a responding mechanism that specifies how the virtual agent should react to particular situations in order to maintain an appropriate environment for productive collaboration.

Data Collection and Annotation

We collected data from 483 dyadic teams recruited from Amazon Mechanical Turk. Among them, we found that 227 teams explicitly mentioned the names of the virtual agents (Art or Dr. Garcia) in their conversations. A statement can be also classified into either a factual or opinion statement. For opinion statements, there are two key components, the target and sentiment (Liu, 2012). Since our interest is in the human opinion toward the virtual agents, the proxies of the opinion targets of interest should be the names of the virtual agents. Sentiment is defined as the underlying feeling, attitude, evaluation, or emotion associated with an opinion. It is represented as a triple (y, o, i), where y is the type of the sentiment, o is the orientation of the sentiment and i is the intensity of the sentiment (Liu, 2012). The orientation and intensity of sentiment can theoretically be on different scales. In our work, we consider the simple scale such as positive (+1), negative (-1) and neutral (0).

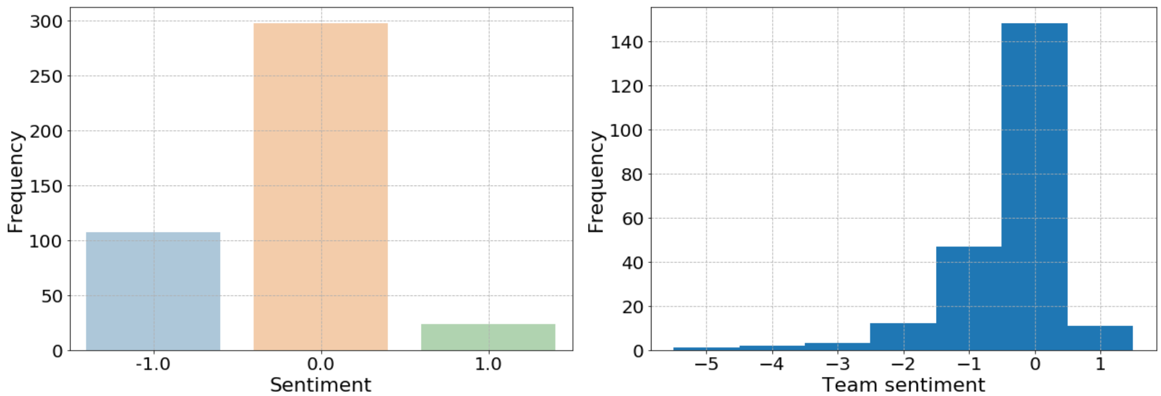


Figure 2. The left panel shows the overall distribution of the sentiment per each turn of conversations. The negative (-1), neutral (0) and positive (1) sentiments are denoted in different colors. The right panel shows the distribution of the total sentiment observations of each team.

Two human coders assigned sentiments to a total of 428 turns of conversations that contain the virtual agent’s names, and achieved an agreement of 0.83 when we computed Cohen’s unweighted Kappa. Each team could have several turns of conversations and each of them has an assigned sentiment, so we computed a team sentiment score as the sum of the sentiments of each turn of the conversations in a given team. In Figure 2, the left panel shows the distribution of the sentiment scores from all conversations of all teams. The right panel shows the distribution of team sentiment scores. It is apparent that the majority of the teams show a neutral team sentiment (0) but there are more teams showing negative (-1) than positive (+1) sentiments. To give readers a sense about these sentiment-bearing conversations, we listed some typical conversations from three teams in the Table 1.

Table 1. Some typical conversations belong to different sentiment categories. We highlighted the virtual agents' names in red color. The A and B denote the two team members.

Team ID	Conversations	Sentiment
tetralogue1485_tetralogue1403	A: mine i don't know what this Art guy is looking at.	negative
	B: and Art is creepy.	negative
	A: Because Art is a thief and a fool.	negative
	A: yes so this Art guy obviously slept through his degree.	negative
	A: ok let's say yes but if Art tricks us I swear....	negative
tetralogue2032_tetralogue1358	A: Yes Art, you are smarter than me.	positive
	A: Maybe we can respond by asking Art about the rest of the dates?	neutral
	B: oh Art!	neutral
	A: Yes, Art.	positive
	A: This is all Art.	neutral
	A: I don't agree with Art.	neutral
tetralogue1507_tetralogue1457	A: Hello Dr. Garcia.	neutral
	B: oops, Art's notes says his notes are for seismometers 1,2,3. Whereas mine only indicates seismometer 1.	neutral
	B: I would keep keep Art's note.	neutral
	B: I agree with Art.	neutral
	A: let's agree with Art.	neutral

Potential Causes of the Negative Sentiments

The team sentiment that we defined in the previous paragraph provides a convenient way for us to identify those teams that show strong negative sentiments. This will further allow us to examine the corresponding behaviors and responses of the virtual agents in those teams, based on which we can identify and classify the potential causes of the negative sentiments. In Table 2, we listed the major causes of some typical negative sentiments from teams in which team sentiments were -2 or lower (18 teams out of 66). The main causes for negative statements were multifaceted. The first could be attributed to using text from the user's note on Art's note (19%). The version of the system for this study used text from one of the user's notes to create a note for Art that used a different number of seismometers; this gave the user an opportunity to compare the quality of two similar notes. However, some users did not like seeing similar text in Art's note. A second main cause could be attributed to lacking trust in the virtual agents' capabilities (human vs. machine) (24%); some users did not trust the virtual agent's capabilities to "do" the assigned work because they were merely virtual agents. A third cause is attributed to the virtual agents' appearance (17%); although virtual agents for this version of the system were adult characters (see Figure 1), some users did not like the appearance of the virtual agents. A fourth was a combination of causes a-c (12%). A fifth was attributed to default responses to "Other" statements (6%); when the system does not know how to respond to a particular statement, the statement is classified as "Other"; some users did not like the default response to "Other" statements.

Table 2. Distribution of major causes of negative sentiments for teams with total negative statements of -2 or less.

Cause	Count	Proportion
Using user text on Art's note	19	41%
Lack of trust in virtual agents' capabilities (Human vs. Machine)	11	24%
Virtual agents' appearance	8	17%
Combination of 2 or 3 of the above	6	12%
Default response to 'Other'	2	6%

A similar distribution was observed for teams whose total negative sentiments was -1 (48 teams out of 66; see Table 3). The main causes for negative statements were (a) using text from the user's note on Art's note (19%), (b) statements demonstrating lack of trust in virtual agents' capabilities (44%), and (c) virtual agents' appearance (17%).

Table 3. Distribution of major causes of negative sentiments for teams with total negative statements of -1.

Cause	Count	Proportion
Using user text on Art's note	9	19%
Lack of trust in virtual agents' capabilities (Human vs. Machine)	21	44%
Virtual agents' appearance	8	17%
Default response to 'Other'	4	8%
Other cases	3	6%
Technical issue	3	6%

Response Mechanism

In most ITS for individual learners, the virtual agent's response mechanism primarily focuses on providing useful content information to the human participants. However, when we extend the ITS to teamwork, we will have the additional opportunity to detect human participants' sentiments towards the virtual agents. Based on this information, we should expand the traditional virtual agent response mechanism by including appropriate responses to mitigate the negative sentiments. In general, participants should be informed about the capabilities of the system in order to avoid unreasonable expectations at the beginning of the task. Then, in the process, the ITS should implement a detector to identify negative sentiments and a classifier to classify the detected negative sentiments. Based on the classification, some scripts can be designed to mitigate the escalation of these situations. In Table 2 and 3, we listed some causes of the negative sentiments. However, some of the causes, such as the "Using user text on Art's note" are task specific and can be fixed by revising the task design. But some causes, such as "Lack of trust in virtual agents' capabilities" could be pretty generic across a wide range of tasks and cannot be easily removed by redesigning the task. In Table 4, we listed some sample responses from the virtual agent after detecting certain categories of causes of negative sentiments that are generic across different tasks.

Table 4. Sample responses from the virtual agent after detecting a negative sentiment category

Cause	Sample responses to mitigate the negative sentiments
Lack of trust in virtual agents' capabilities (Human vs. Machine)	Virtual agent: Although I am a virtual character, I have been programmed by experts to help you through this task. Virtual agent: I have been helping many students on this topic for a long time. Virtual agent: I may have misunderstood what you said, could you please provide more information about it?
Virtual agents' appearance	Virtual agent: Don't judge a book by its cover. Virtual agent: My appearance may have surprised you but I am perfectly capable of helping you. Virtual agent: Moving on to important matters...

Discussion

Sentiment analysis can provide useful information on how users perceive the virtual agents. This information can be used to improve the appearance of virtual agents as well as their interactions with users. Such information has already guided us to change the task design of our simulation-based tasks and these changes have contributed to a positive user experience. For example, regarding the users' perceptions of virtual agents, we are exploring how the virtual peer's prior knowledge on the topic of the simulation (High vs. Low) and type of question (Compare alternatives vs. Agree/disagree with the peer's conclusion) influence the quality of evidence collected (Sparks, Andrews, Zapata-Rivera, Lehman, & James, 2016) and the types of emotions users experience in these types of assessment systems (Lehman & Zapata-Rivera, 2017).

The current study described a general scheme to leverage the sentiment information from an ITS for teamwork and identified two important types of potential causes for negative sentiments based on our Tetralogue simulation. These two types of causes correspond to the situations where humans made comments questioning the capabilities of the virtual agents (human vs. machine) and the situations where humans overreacted to the appearance of the virtual agents. More research on how to properly react to these situations will be important for improving the user experience of ITSs for teamwork. Moreover, these teamwork-centered improvements using agents need to be incorporated in an ITS framework such as the Generalized Intelligent Framework for Tutoring (GIFT, Sottilare, Brawner, Goldberg, & Holden, 2012).

Conclusions and Recommendations for Future Research

In this chapter, we illustrated how human opinions about a virtual agent's responses can be measured in a teamwork task and used to improve the responding mechanism of ITS for teamwork. Incorporating a teamwork-centered layer to the GIFT architecture could provide an opportunity to make adjustments based on data captured in real time. Because GIFT routinely incorporates production rules to respond to a variety of learner knowledge states and history of activities in the learning record store, it is feasible that this teamwork-centered component could therefore be readily migrated into GIFT. Nonetheless, it does introduce a possible complication, namely tracking multiple team members and learners instead of one at a time.

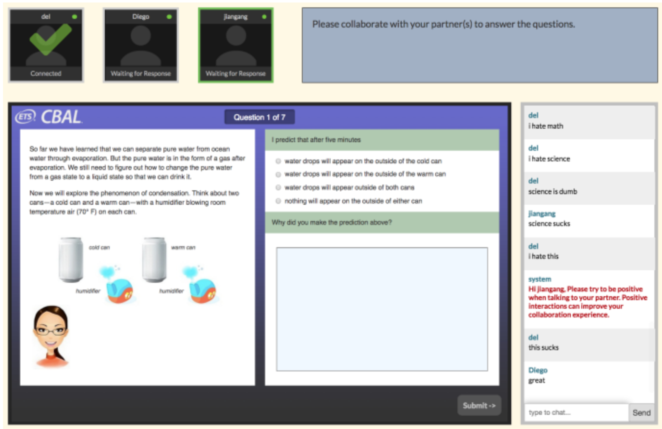


Figure 3. Intelligent Facilitator for teamwork based on the EPCAL platform.

Generally speaking, developing an ITS for teamwork is much more challenging than developing an ITS for individual learners because the system will need to respond to each team member individually as well as the interactions among the team members. However, if we reduce the role of the virtual agent from a tutor to a facilitator, e.g., an intelligent facilitator (Hao, Chen, Flor, Liu, & von Davier, 2017a; Hao et al., 2017c), the system becomes more feasible. Based on a comprehensive literature review of facilitation mechanism of teamwork (Hebert & Zapata-Rivera, 2018), we are implementing an intelligent facilitation system (IFS) based on the ETS Platform for Collaborative Assessment and Learning (EPCAL, Hao et al., 2017c). The EPCAL platform features a modularized design with full capability to manage team formation, task progress, and receive external feedback. Figure 3 shows a screenshot of the user interface of EPCAL. Empirical data will help us to understand the actual challenge of interacting with teams and pave the road toward an IFS for teamwork.

References

Barrón-Estrada, M. L., Zatarain-Cabada, R., Oramas-Bustillos, R., & González-Hernández, F. (2017, July). Sentiment Analysis in an Affective Intelligent Tutoring System. In Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on (pp. 394-397). IEEE.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, *59*(2), 291-300.

Cai, Z., Graesser, A.C., & Hu, X. (2015). ASAT: AutoTutor script authoring tool. In. R. Sottilare, A.C. Graesser, X. Hu, & K. Brawner (Eds.), *Design Recommendations for Intelligent Tutoring Systems: Authoring Tools* (Vol.3)(pp.199-210). Orlando, FL: Army Research Laboratory

Chou, C. Y., Chan, T. W., & Lin, C. J. (2003). Redefining the learning companion: the past, present, and future of educational agents. *Computers & Education*, 40(3), 255-269.

Graesser, A. C., Li, H., & Forsyth, C. (2014). Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science*, *23*(5), 374-380.

Hao, J., Chen, L., Flor, M., Liu, L., & von Davier, A.A., (2017a), CPS-rater: automated sequential annotation for conversations in collaborative problem-solving activities, ETS Research Report No. RR-17-58, Princeton, NJ: Educational Testing Service

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017b). Initial steps towards a standardized assessment for CPS: Practical challenges and strategies. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative Assessment of Collaboration*. New York: Springer.

Hao, J., Liu, L., von Davier, A.A., Lederer, N., Zapata-Rivera, D, Jkal, P., & Bakkenson, M., (2017c), EP-CAL: ETS Platform for Collaborative Assessment and Learning, *ETS Research Report No*. RR-17-49. Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12181

Hao, J., Smith L., Mislevy, R., von Davier, A., & Bauer, M., (2016). Taming log files from the game and simulation-based assessment: Data model and data analysis tool. *ETS Research Report* RR-16-11. Princeton, NJ: Educational Testing Service.

Hao, J., Liu, L., von Davier, A., & Kyllonen, P. (2015), Assessing collaborative problem solving with simulation based tasks, proceeding of 11th international conference on computer supported collaborative learning, Gothenburg, Sweden

Hebert, D., & Zapata-Rivera, D. (2018). Intelligent Support for Collaboration: A Literature Review and Proposed Framework. Paper accepted for presentation at the annual meeting of the American Educational Research Association. New York, NY.

Johnson, B. G., Phillips, F., & Chase, L. G. (2009). An intelligent tutoring system for the accounting cycle: Enhancing textbook homework with artificial intelligence. *Journal of Accounting Education*, *27*(1), 30-39.

Johnson, W.L., & Lester, J.C. (2016). Face-to-face interaction with pedagogical agents, Twenty years later. *International Journal of Artificial Intelligence in Education*. *26* (1), 25–36

Kumar, R., Ai, H., Beuth, J. L., & Rosé, C. P. (2010, June). Socially capable conversational tutors can be effective in collaborative learning situations. In *International Conference on Intelligent Tutoring Systems* (pp. 156-164). Springer, Berlin, Heidelberg.

Lehman, B., & Zapata-Rivera, D. (2017). Student emotions during conversation-based assessments. Paper presented at the annual meeting of the AERA, San Antonio, TX.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, *5*(1), 1- 167.

Liu, L., Hao, J., von Davier, A., Kyllonen, P., & Zapata-Rivera, D. (2015). A tough nut to crack: Measuring collaborative problem solving. Y. Rosen, S. Ferrara, & M. Mosharraf (Eds). *Handbook of Research on Computational Tools for Real-World Skill Development*. Hershey, PA: IGI-Global.

Louwerse, M.M., Graesser, A.C., Lu, S., & Mitchell, H.H. (2005). Social cues in animated conversational agents. *Applied Cognitive Psychology, 19*, 693-704.

Moreno, R., Mayer, R.E., Spires, H.A., & Lester J.C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*(2) 177-213.

Schroeder, N., Adesope, O., & Gilbert, R (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research 49*(1), 1-39.

Sparks, J.R., Andrews, J.J., Zapata-Rivera, D., Lehman, B., & James, K. (2017). Students' perceptions of collaborator expertise in science inquiry tasks using simulated conversations. Paper presented at the annual meeting of the AERA, San Antonio, TX.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED).

Zapata-Rivera, D., Jackson, T., & Katz, I.R. (2015) Authoring Conversation-based Assessment Scenarios. In R. A. Sottilare, A. C. Graesser, X. Hu, and K. Brawner (Eds.), Design Recommendations for Intelligent Tutoring Systems Volume 3: Authoring Tools and Expert Modeling Techniques. (pp. 169–178). U.S. Army Research Laboratory.

Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., & Katz, I. R., (2014) Science Inquiry Skills using Trialogues. *12th International conference on Intelligent Tutoring Systems*, 625-626.

Chapter 15 – Characteristics and Mechanisms of Team Effectiveness in Dynamic Environments

Jamie C. Gorman[1], Sidney K. D'Mello[2], Ronald H. Stevens[3,4] & C. Shawn Burke[5]
Georgia Institute of Technology[1], University of Colorado-Boulder[2], University of California-Los Angeles[3], The Learning Chameleon, Inc.[4], University of Central Florida[5]

Introduction

How teams achieve and maintain effectiveness in uncertain and unpredictable environments is a central issue in team research. Real-world task environments are dynamic, such that even a highly trained team can encounter novel situations for which they have not been trained. Nevertheless, effective teams should be able to match their dynamics to those of the task environment. Following Thelen (2000), we identify four characteristics of effective team dynamics that are expressed through team variations and adaptations as a team functions in a dynamic environment. Next, we explore the underpinnings of effective team dynamics in terms of basic mechanisms and processes, including: (1) distributed cognition and multisensory perception, (2) team neurodynamics, (3) interpersonal processes and team coordination dynamics, and (4) team member knowledge, skills, and attitudes. Implications for assessment and training using team intelligent tutoring systems (team ITSs; Sottilare et al., 2017) are considered.

Teams as Dynamic Entities

Teams, defined as two or more people who work interdependently toward a common goal (Salas, Dickinson, Converse, & Tannenbaum, 1992), can be viewed as inherently dynamic entities. For example, a basketball team has an identity that is preserved under changes in team membership, changes in team roles, changes in team leadership, and is (ideally) robust to changes in rules, styles of play, and other changing conditions of its surrounding environment. Viewed as dynamical entities, there are different variations and adaptations that allow a team to remain effective (Gorman, Grimm, & Dunbar, in press). These include processes such as leadership emergence, team member turnover, and dynamic role restructuring. Although these processes, for example, leadership emergence, have typically been studied in terms of individual traits and abilities (e.g., cognitive ability and personality traits; Kickul & Neuman, 2000; Luria & Berson, 2013), we suggest that they should be viewed dynamically, as a real-time response to the changing demands put on a team by its surrounding environment. That is, in a teleological sense, a team has a function or goal that can only be reached through real-time variations and adaptations such as those already mentioned. In this light, what is team effectiveness, and how should it be trained? What are the underlying mechanisms and processes that allow teams to maintain their effectiveness in dynamic, uncertain environments?

Characteristics of Effective Teams

Assuming teams are dynamic entities, what are the general characteristics of effective teams? We extend Thelen's (2000) propositions for what it means to be skilled to the context of maintaining team effectiveness (Gorman et al., in press):

Four General Characteristics of Team Effectiveness in Dynamic Environments:

1. Effective teams are adept at assessing the current situation and noticing the relevant stimuli in the physical and social environment for accomplishing the shared goal.

2. Effective teams are adept at behaving in similar, but not identical ways, in choosing an action that fits the current situation. This requires a flexible and generative response to the changing environment, perhaps similar to, but not identical with, responses used in the past.

3. Effective teams have consistent behavior in similar (routine) situations but are good at changing their actions rapidly and appropriately as the situation requires (i.e., in novel, non-routine situations).

4. Effective teams have a repertoire of adaptations (the sources of variation and adaptation mentioned earlier) through which Characteristics 1-3 are expressed.

These characteristics suggest that for sources of team variation and adaptation to be successful, they must account for changing task demands. In the following sections, we highlight several sources of variation and adaptation through which teams coordinate their behaviors to achieve their common goals. These can also be thought of as processes or mechanisms that are critical to our understanding of how to train and assess team effectiveness in dynamic environments using team ITSs.

Processes and Mechanisms through which Teams Maintain Effectiveness

The distinction between process and mechanism can be debated. In this chapter, we distinguish between dynamic processes through which teams vary their behaviors and the underlying mechanisms that cause those processes. For example, the process through which different teams evolve different customary ways of referring to the same task-relevant concepts—lexical entrainment (Brennan, 1996)—can be associated with a mechanism (long-memory; Gorman, Dunbar, Grimm, & Gipson, 2017). We argue this distinction is important because, in many cases, mechanisms must be trained and their resulting processes must be assessed.

(1) Distributed Cognition and Multisensory Perception

Distributed cognition pertains to cognition that extends beyond the individual to collections of interacting individuals and their environment. For example, in the classic case of flight navigation (Hutchins, 1995), the pilot and co-pilot share a mental representation constructed from the instrument panels and visual cues in their immediate environments. They also coordinate actions, which are reflected in changes in the environment itself, leading to an updated shared mental representation, and so on. Thus, the pilot, co-pilot, and the environment form a distributed system, which compensates for the limitations of the individual components (Giere & Moffatt, 2003). Distributed cognition is distinct from traditional cognition in that the unit of analysis is a *system* not an individual. It is also different from collective cognition by virtue of its inherent inclusion of the environment as part of the distributed system (Giere, 2007).

Distributed cognition is multifaceted and multisensory in complex environments. It extends beyond cold cognition into the realm of emotion. Considerable research has indicated that emotions play a critical role in modulating cognitive processes within an individual (Clore & Huntsinger, 2007). Similar effects are expected to occur across interacting individuals, the most well-known example being mood contagion (Barsade, 2002). Distributed cognition is also fundamentally social in nature, implicating multimodal interpersonal coordination processes, such as linguistic coordination (Fusaroli & Tylén, 2012), mood contagion (Barsade, 2002), joint attention (Richardson, Dale, & Kirkham, 2007), and physiological synchronization (Konvalinka et al., 2011).

Thus, distributed cognition is an umbrella term for socio-cognitive-affective processes that are multimodal, interact over multiple spatial and temporal scales, and are embedded in a constantly-changing environment.

How can interacting individuals cope with this level of complexity? By working as a team, of course. Specifically, by self-organizing into interpersonal synergies (Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli, Rączaszek-Leonardi, & Tylén, 2014), interacting individuals can function as a single unit, reducing the number of degrees of freedom (dimensional compression) and allowing individual components to react to changes in others (reciprocal compensation) (Riley, Richardson, Shockley, & Ramenzoni, 2011). Hence, dimensional compression and reciprocal compensation can be thought of as mechanisms of change through which individuals gel into a cohesive, effective team.

(2) Team Neurodynamics

Before teams physically organize themselves, they mentally organize themselves by matching within and across-brain efforts to task and inter-personnel demands. Effective teams balance these momentary relationships until the environment becomes uncertain, when they must reorganize their thinking, roles, and configurations into more appropriate corrective structures. The neurodynamic processes associated with these re-organizations are poorly understood.

One approach for understanding the neurodynamic organizations that occur during team uncertainty begins with capturing and modeling the electroencephalographic (EEG) signals of team members during required training (Stevens, Galloway, Wang, Berka et al, 2012; Stevens, Gorman, Amazeen, Likens & Galloway, 2013). EEG is the recording of electrical activity of the brain at different regions along the scalp. The rhythmic patterns in the electrical oscillations from different brain regions contain signals representing complex facets of brain activity that can be modeled at both the individual and team level by constructing symbolic temporal and spatial neurodynamic histories of team performance (Stevens & Galloway, 2017).

These models have shown the tendency of team members to enter into prolonged (up to minutes) metastable neurodynamic relationships with each other as they encounter and resolve disturbances to their rhythms. As data from submarine navigation and healthcare teams have accumulated, a trend has emerged showing that less experienced teams have larger and more persistent neurodynamic organizations than more experienced teams (Stevens, Galloway, Lamb, Steed & Lamb, 2017).

In other words, when seeking new or different ways to balance the demands of the changing environment, teams adopt a more organized configuration, neurodynamically speaking, but when these challenges and uncertainties are resolved, teams, and especially efficient teams, return to a less structured neurodynamic organization. The length of these neurodynamic fluctuations can be seconds or much longer depending on the nature of the uncertainty, the experience of the team, and the amount of new information that has to be acquired, synthesized, and exchanged before the team can return to an efficient operating mode.

More recently, research has shown that the neurodynamic organizational modeling can be extended to individual team members, which enables direct quantitative comparison of the organizational contributions of each team member to the overall team neurodynamic organization (Stevens & Galloway, 2017). To the extent that increased neurodynamic organization of individuals occurs during periods of uncertainty, stress, and moments of increased attention (Stevens, Galloway, Halpin, & Willemsen-Dunlap, 2016), the individual neurodynamics may identify opportunities to insert scaffolding or feedback triggers for individual team members, or the team as a whole.

(3) Interpersonal Processes and Team Coordination Dynamics

Compelling team members to interact in novel ways (rather than in a rote, fixed manner) exercises the interpersonal coordination processes teams need in order to maintain effectiveness in dynamic environments (Gorman, 2014). This idea has been demonstrated by mixing team member affiliations while maintaining team member roles (Gorman, Amazeen, & Cooke, 2010) and through "perturbation training", which forces teams to organize new solutions to coordination problems (Gorman, Cooke, & Amazeen, 2010). These intervention and training approaches are based on the principle that team effectiveness happens during the process of team interaction and is not inherent in static properties or traits of team members. In terms of assessment, the flexibility and stability of team interaction (e.g., communication) patterns provide measures that are correlated with the characteristics of team effectiveness described earlier. For instance, more flexible team coordination dynamics are correlated with overcoming more team situation awareness "roadblocks", which are novel events that interfere with achieving a shared goal unless properly acted upon (Gorman et al., 2010a). Team mixing and perturbation training are two mechanisms that lead to flexible team coordination dynamics.

Long-term memory is a component of peoples' information processing architecture, in which knowledge persists on longer timescales than the information currently being processed in short-term, or working, memory. Parallel to these memory concepts, long-memory (Beran, 1994) is a mechanism through which teams hold persistent communication and coordination patterns on longer timescales than the information currently being processed by the team (Gorman et al., 2017). The significance of long-memory is that it provides a mechanism through which team variations and adaptations are consolidated into the history of a team. In this way, more permanent features of team interaction are conserved as new team members come and go, team member roles change, and environmental conditions change, all the while informing the ongoing coordination processes of this dynamic entity known as a team. Long-memory effects can be observed as the presence of power-law scaling of team communication, and other interaction patterns, wherein correlations between communication or interaction events persist over relatively long timescales of task performance (e.g., days; Gorman, 2005; or hours; Gorman et al., 2017).

(4) Knowledge, Skills, and Attitudes

In addition to the above lines of work, much work has been conducted to uncover the knowledge, skills, and attitudes that facilitate team interaction processes. In doing so, research has noted the dynamic, cyclical nature of team interaction and argued that in working towards a goal/objective, teams progress through a series of performance episodes (i.e., temporal cycles of goal-directed activity, see Marks, Mathieu, & Zaccaro, 2001). Within these performance episodes, team members oscillate between action phases (i.e., periods of time when members are directly engaged in behaviors that support task accomplishment) and transition phases (i.e., periods of time when teams focus on planning/evaluation activities to guide goal accomplishment, Marks et al., 2001). In terms of the behaviors that comprise each of these phases, meta-analytic evidence has shown monitoring progress towards goals, systems monitoring, team monitoring/backup behavior, and coordination as being important within action phases (LePine, Piccolo, Jackson, Mathieu, & Saul, 2008). In contrast, behaviors such as mission analysis and planning, goal specification, and strategy formulation have been found to occur during transition phases and are related to effective team performance (LePine et al., 2008).

Complementing the above work, is work by Salas, Shuffler, Thayer, Bedwell, and Lazzara (2015) who developed a heuristic that can be utilized to understand the factors that should be considered when selecting, developing, and maintaining teams. This heuristic describes six core processes and states that serve to facilitate teamwork. In line with the work mentioned above, Salas et al. note the importance of behaviors such as coordination and communication. In addition to this, they highlight coaching as being a fundamental team process. In this vein, coaching refers to the "enactment of leadership behaviors to establish goals and set direction that leads to the successful accomplishment of these goals" (Salas et al., 2015, p. 603). The importance of team leadership has also been supported through meta-analytic work (e.g., Burke, Stagl, Klein, Goodwin, Salas, & Halpin, 2006; D'Innocenzo, Mathieu, & Kukenberger, 2016; Wang, Waldman,

& Zhang, 2013). The developed heuristic also highlights the importance of emergent states (e.g., cooperation, conflict, cognition). Emergent states have been defined as, "properties of the team that are typically dynamic in nature and vary as a function of team context, inputs, processes, and outcomes" (Marks et al., 2001, p. 357). Emergent states are primarily cognitive and attitudinal in nature. Cognitive states that have been supported as being important to team effectiveness include shared mental models and transactive memory systems. The specific elements of shared cognition which have been shown to be important include knowledge of member roles and responsibilities, team mission objectives and norms, and who has what knowledge, skills, and abilities on the team (Salas et al., 2015; DeChurch & Mesmer-Magnus, 2010; Lewis, 2004). Shared cognition has often been argued to be a primary coordination mechanism, especially with respect to implicit coordination within teams.

In addition to shared cognition, empirical work has also shown several attitudinal states to be important for team effectiveness (e.g., collective efficacy, trust, cohesion, psychological safety). Collective efficacy, or belief in the team's ability to accomplish a given task, has been shown to be related to team effort, increased team performance, and team satisfaction (Stajkovic, Lee, & Nyberg, 2009; Gully, Incalcaterra, Joshi, & Beaubien, 2002). Trust has been found to be related to information sharing in teams and increased team performance (DeJong, Dirks, & Gillespie, 2016; Lee, Gillespie, Mann, & Wearing, 2010). Related to trust, psychological safety (i.e., "shared belief that the team is safe for interpersonal risk taking", Edmondson, 1999, p. 354) has been shown to facilitate team performance through its facilitation of learning behaviors (Edmondson, 1999). Finally, several meta-analytic efforts have shown the relationship between cohesion (i.e., attraction/bonding to the group) and team performance (e.g., Mathieu, Kukenberger, D'Innocenzo, & Reilly, 2015; Beal, Cohen, Burke, & McLendon, 2003). While the above states do not reflect all the attitudinal states that have been shown to be related to team effectiveness, they reflect many of those most commonly examined. Attitudinal states have been argued to be the "motivational drivers of teamwork" (see Salas et al., 2015, p. 603).

Conclusions and Recommendations for FUTURE research

We have offered a view of teams as dynamic entities, or systems, along with several defining characteristics of effective teams from this perspective. Distinguishing between mechanism and process allows one to focus on what needs to be trained and what needs to be assessed in order to enhance team effectiveness in dynamic environments. But how do mechanisms and processes at the dynamic team level, and their training and assessments, fit with team ITSs and the Generalized Intelligent Framework for Tutoring (GIFT)?

Distributed cognition and team coordination dynamics suggest novel mechanisms for enhancing team effectiveness (e.g., compression/reciprocal compensation; perturbation training) that could feasibly be introduced in a team ITS environment. However, the specifics of how to introduce them into team ITS and, moreover, how to assess their effects is an underdeveloped area of research. In the case of team neurodynamics, individual contributions of team members to team dynamics and scaffolding and feedback based on this information presents a similar challenge. Even in the case of knowledge, skills, and attitudes, where much more is known about training team effectiveness, the assessments that go along with the implied team dynamics (e.g., temporal cycles of goal-directed activity) are underdeveloped. Thus, we think that research aimed at understanding how to implement training of dynamic team mechanisms and better assessment of dynamic team processes are critical areas for future research on team ITSs.

GIFT seems ideal for training mechanisms and behaviors at the individual level. However, to implement some of the training, assessments, mechanisms, and processes described in this chapter might require revisions of GIFT or development of new GIFT modules. In particular, the assessment of socio-cognitive-affective processes in distributed cognition, individual contribution to team neurodynamics, and the flexibility of team coordination dynamics might require a real-time analysis engine in order to understand the effects of training interventions on these processes. Moreover, as described in the team neurodynamics section, real-time assessments at times of stress or uncertainty would provide the opportunity to introduce timely perturbations or feedback tailored to a team's ongoing performance and development. Hence, another area for future research and development is real-time analysis methods that would allow for training interventions and assessments at the dynamic team level.

In the section on knowledge, skills, and attitudes, we noted recent research showing the importance of coaching for enhancing team effectiveness. We suggest that the role of coach might be an appropriate metaphor for team ITSs for training effective teams. As a coach, we might imagine team ITSs guiding team members through training scenarios, where the ITS operationalizes the mechanisms described in this chapter, inserts them into the scenario in real time, assesses the team, and provides feedback/guidance based on this information. If the team does not respond effectively to a scenario through appropriate team variation and adaptation, then remediation could include feedback on how the team members interacted or guidance on understanding how to change their team dynamics. To realize this vision, however, we need to better understand how teams as dynamic entities can be incorporated into the team ITS framework.

References

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47*(4), 644-675.

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology, 88*(6), 989-1004.

Beran, J. (1994). *Statistics for long-memory processes* (Vol. 61). New York, NY: Chapman & Hall.

Brennan, S. (1996). Lexical entrainment in spontaneous dialog. Proceedings, 1996 International Symposium on Spoken Dialogue (ISSD-96), 41–44.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly, 17*, 288-307.

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393-399. doi: 10.1016/j.tics.2007.08.005

Dale, R., Fusaroli, R., Duran, N., & Richardson, D. C. (2013). The self-organization of human interaction. *Psychology of Learning and Motivation, 59*, 43-95.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice, 14*(1), 1-14.

D'Innocenzo, L., Mathieu, J. E., & Kukenberger, M. R. (2016). A meta-analysis of different forms of shared leadership-team performance relations. *Journal of Management, 42*(7), 1964-1991.

DeJong, B.A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology, 10*(8), 1134-1150.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*, 350-383.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32*, 147-157.

Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: a dynamical approach. *Interaction Studies, 13*(1), 103-124.

Giere, R. N. (2007). Distributed cognition without distributed knowing. *Social Epistemology*, 21(3), 313-320.

Giere, R. N., & Moffatt, B. (2003). Distributed cognition: Where the cognitive and the social merge. *Social Studies of Science, 33*(2), 301-310.

Gorman, J. C. (2005). The concept of long memory in assessing the global effects of augmented team cognition. *Proceedings of the International Conference on Human-Computer Interaction*, *11*, 22-27.

Gorman, J. C. (2014). Team coordination and dynamics: Two central issues. *Current Directions in Psychological Science*, *23,* 355-360.

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010a). Team Coordination Dynamics. *Nonlinear Dynamics Psychology and Life Sciences, 14,* 265-289.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010b). Training adaptive teams. *Human Factors, 52,* 295-307.

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modeling teams as dynamical systems. *Frontiers in Psychology, 8,* 1053.

Gorman, J. C., Grimm, D. A., & Dunbar, T. A. (in press). Defining and measuring team effectiveness in dynamic environments and implications for team ITS. In Building Intelligent Tutoring Systems for Teams: What Matters, Vol. 19. Emerald Publishing Limited.

Gully, S. M., Incalcaterra, K. A., Joshi, A., & Beaubien, J. M. (2002). A meta-analysis of team efficacy, potency, and performance: Interdependence and level of analysis as moderators of observed relationships. *Journal of Applied Psychology, 87*(5), 819-832.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*(3), 265-288.

Kickul, J. & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAs. *Journal of Business and Psychology, 15*, 27-51.

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., . . . Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences, 108*(20), 8514-8519.

Lee, P., Gillespie, N., Mann, L., & Wearing, A. (2010). Leadership and trust: Their effect on knowledge sharing and team performance. *Management Learning, 41*(4), 473-491.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*, 273-307.

Lewis, K. (2004). Knowledge and performance in knowledge-worker teams: A longitudinal study of transactive memory systems. *Management Science, 50*(11), 1519-1533.

Luria, G. & Berson, Y. (2013). How do leadership motives affect informal and formal leadership emergence? *Journal of Organizational Behavior, 34,* 995-1015.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework of team processes. *The Academy of Management Review, 26*(3), 356-376.

Mathieu, J. M., Kukenberger, M. R., D'Innocenzo, L., & Reilly, G. (2015). Modeling reciprocal team cohesion-performance relationships, as impacted by shared leadership and members' competence. *Journal of Applied Psychology, 100*(3), 713-734.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407-413.

Riley, M. A., Richardson, M. J., Shockley, K., & Ramenzoni, V. C. (2011). Interpersonal synergies. *Frontiers in Psychology, 2*(38).

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management, 54*(4), 599-622.

Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H., & Gilbert, S.B. (2017). Designing adaptive instruction for teams: A meta-analysis. *International Journal of Artificial Intelligence Education,* 1 – 40. https://doi.org/10.1007/s40593-017-0146-z

Stajkovic, A., Lee Dongseop, & Nyberg, A. J. (2009). Collective efficacy, group potency, and group performance: Meta-analyses of their relationships, and a test of a mediation model. *Journal of Applied Psychology, 94*(3), 814-828.

Stevens, R. H., & Galloway, T. (2017). Are neurodynamic organizations a fundamental property of teamwork? *Frontiers in Psychology*. doi: 10.3389/fpsyg.2017.00644.

Stevens, R. H., Galloway, T., Halpin, D., & Willemsen-Dunlap, A. (2016). Healthcare teams neurodyamically reorganize when resolving uncertainty. *Entropy, 18*: 427, doi: 10.3390/e18120427

Stevens, R., Galloway, T., Lamb, J., Steed, R., & Lamb, C. (2017). Linking team neurodynamic organizations with observational ratings of team performance. In *Innovative Assessment of Collaboration* (pp. 315-330). A. A. Von Davier, P. C. Kyllonen, & M. Zhu eds. Springer International Publishing, Cham, Switzerland.

Stevens, R.H., Galloway, T., Wang, P., Berka, C., Tan, V., Wohlgemuth, T., Lamb, J., Buckles, R. (2012). Modeling the neurodynamic complexity of submarine navigation teams. *Computational and Mathematical Organization Theory*, *19(3*), 346-369.

Stevens, R. H., Gorman, J.C., Amazeen, P., Likens, A., and Galloway, T. (2013). The organizational dynamics of teams. *Nonlinear Dynamics, Psychology and Life Sciences 17*(1), 67-86.

Thelen, E. (2000). Grounded in the world: Developmental origins of the embodied mind. *Infancy, 1,* 3-28.

Wang, D., Waldman, D., & Zhang, Z. (2013). A meta-analysis of shared leadership and team effectiveness. *Journal of Applied Psychology, 99*(2), 181-198.

Chapter 16 – Roles of Talking Agents in Online Collaborative Learning Environments

Zhiqiang Cai[1], Andrew J. Hampton[1], Art Graesser[1], Xiangen Hu[1,2], Jody L. Cockroft[1], David W. Shaffer[3], & Michael C. Dorneich[4]
The University of Memphis[1], Central China Normal University[2], University of Wisconsin-Madison[3], Iowa State University[4]

Introduction

Collaborative learning is a process that involves a group of people working together on a shared learning goal. Traditional classroom collaborative learning requires learners to be at the same location. With internet support, team members can now collaborate online. Many online tutoring systems have been developed to allow learners to participate in group collaborations from distant locations. Collaborators bring different ideas and experiences to the group, working online to construct answers to questions and solutions to problems. Talking is an important feature in collaboration. As Golub (1988) pointed out, it is in the talking that much of the learning occurs. Online collaborative learning environments can easily provide utilities for collaborators to talk through chat or messaging interfaces.

Group members in online collaborative learning environments are not necessarily all human. Instead, some can be intelligent computer agents. An intelligent computer agent is a software robot that performs tasks a human user may do on computers. Such agents have "brains" that can find answers to questions or solutions to problems. In addition, they could talk through avatars that can deliver speeches, gesture, and convey emotions. Such talking agents may play different roles, such as a tutor who helps an online student learn specific concepts, a peer student who provides alternative perspectives, a teacher who guides a team through a complete learning path, a facilitator who maintains healthy social interaction among team members, etc. The role a talking agent plays is usually specified by the types of tasks the agent takes, the manner in which the agent perform the tasks, and the protocols that determine how the agent interacts with other collaborators. In this chapter, we present a subset of talking agents we have created in the last decade in intelligent tutoring systems. We describe their roles, how they are specified, and how they are authored. We conclude with suggestions about standardization and integration of role-based talking agents in the Generalized Intelligent Framework for Tutoring (GIFT).

Talking Agents in Intelligent Tutoring Systems

Single Tutor Agent in the Early AutoTutor Systems

The first AutoTutor system was developed in the late 1990s (Graesser, Wiemer-Hastings, Wiemer-Hastings, & Kreuz, 1999). The system was designed to help learners taking an introductory course in computer literacy. It presents a set of problems on the fundamentals of computer hardware, operating systems, and the Internet. A single tutor agent and a human learner collaboratively construct solutions to each problem. Based on the discovery that tutoring by normal, untrained tutors proves (Person, Graesser, Kreuz, & Pomeroy, 2003), the agent was designed to simulate a normal unskilled human tutor who knows the learning content but does not have pedagogical skills in tutoring. That is, the tutor agent simulates a domain expert with appropriate domain knowledge.

The tutor agent presents as an avatar with a low quality voice and some simple hand gestures, limited by the speech synthesis technology at that time. The interaction between the agent and a human learner is through a so called "expectation–misconception tailored" (EMT) conversation (Graesser, 2016). An EMT conversation is started by the tutor agent initially posing a main question about a problem that the human learner tries to answer. The tutor agent evaluates the learner's answer and gives appropriate feedback (positive, neutral, negative, etc.). If the answer does not cover all parts (expectations) of the ideal answer to the main question, the computer agent selects an expectation and asks a hint/prompt question. Then the learner answers, the tutor evaluates, the tutor gives feedback and the tutor selects the next expectation to work on. This process repeats until either all expectations are covered or all prepared questions are exhausted. During the course of the iteration, if a misconception is identified, the computer agent corrects the misconception and the process continues. Examples of such conversations can be found in published articles, for example, Graesser (2016), and Graesser et al. (1999, 2005).

The responsibilities of this type of agent includes: 1) presenting a problem; 2) evaluating human learners' answers; 3) giving immediate feedback; 4) selecting expectations and asking hint/prompt questions; 5) identifying and correcting misconceptions; and 6) presenting the final solution to the problem. In order to do all these tasks, the agent needs a "brain" that stores the problem, the solution, expectations, misconceptions, hint/prompt questions, answers, and feedback speeches. Moreover, its brain needs to be intelligent enough that it can evaluate a learner's language and make decisions on what expectation/misconception to work on and what hint/prompt to ask the learner. The intelligent brain of this agent is equipped with a semantic engine that can classify speech acts and is capable of matching stored answers. The semantic engine uses latent semantic analysis (Landauer, McNamara, Dennis, & Kintsch, 2007) and regular expressions for flexibility and inclusivity, though specifics of their implementation lies beyond the scope of this chapter. In short, this agent "understands" natural language input and knows how to select expectations, misconceptions and questions.

The single agent AutoTutor is relatively easy to construct because of the simplicity of the conversation rules. The process of authoring one problem can be briefly described as follows:

1) Describe the problem and ask a main question;

2) Prepare an ideal answer to the main question;

3) Split the ideal answer into multiple expectations;

4) For each expectation, prepare a set of hint/prompt question-answer pairs.

5) Prepare a list of possible misconceptions and correct responses.

Due to the simplicity of authoring, this type of agent can be easily created for tutoring in other domains. In fact, the second AutoTutor system focuses on Newtonian physics and features improvements to the agent's avatar and speech engine. The development of such single talking agent AutoTutor systems form the essential conversation patterns for AutoTutor systems. Of course, a learning system may have multiple agents, where each agent plays a meaningful role. In the next example, we present a system with four talking agents: a domain tutor agent and three self-regulated learning strategy agents.

Multiple Talking Agents in MetaTutor

MetaTutor utilizes multiple talking agents (Azevedo et al., 2009, 2012) in an effective environment for learning human body systems. This system differs from earlier AutoTutor systems in that, in addition to giving feedback to content learning, the multiple talking agents prompt learners to employ self-regulated learning strategies and give feedback to the learner on the outcome of the strategy. In MetaTutor, there are four talking agents. Gavin, the main agent, provides guidance in navigating content and gives feedback. Three additional satellite agents, Pam, Mary and Sam, provide assistance on different self-regulated learning skills in three learning phases, including planning, monitoring, and applying learning strategies (Graesser & McNamara, 2010). The planning phase requires skills such as setting up a major learning goal and some subgoals. The metacognitive monitoring demands judgments of learning, feeling of knowing, content evaluation, adequacy of a strategy, and progress toward goals. The learning strategies include searching for relevant information, taking notes, drawing tables or diagrams, re-reading, elaborating the material, making inferences, and coordinating information sources.

The learning materials in MetaTutor are presented in 41 pages of text and static diagrams on the human circulatory system. A learner starts by setting up an overarching learning goal and some sub-goals guided by Pam. Gavin then guides the learner to navigate the content and scaffolds content learning. A skill panel on the interface allows a learner to instantiate interaction with an agent. For example, a learner may click a button to indicate that he/she wants to take a note. The learning strategy agent, Sam, will then interact with the learner so that the learner applies this strategy effectively (taking a good note, in this case). Azevedo et al. (2012) showed that providing prompts and feedback to self-regulated learning skills produces significant learning gains.

The responsibilities of Gavin are guiding navigation and leading EMT conversations. The other three agents' responsibilities include: 1) prompting the use of a specific set of self-regulated learning strategies; 2) evaluating the outcomes of the use of suggested strategy; and 3) giving immediate feedback.

The authoring tasks in building pedagogical agents (Pam, Mary, and Sam) are different from building domain expert agents. To suggest an appropriate strategy to a learner at a specific moment of the learning process implies that the agent knows how to map learning states to strategies. Therefore, a learning state to learning strategy mapping needs to be stored in the agents' brain. The agents need to be capable of evaluating the outcomes of strategy use. Endowing the agents with this capability is challenging technically. For example, when a learner uses the strategy "drawing a diagram", evaluating the diagram involves image processing utilities to determine quality. Similarly, evaluating the "taking notes" strategy requires natural language processing utilities.

Interactions in MetaTutor are dialogs between one talking agent and one human learner. Although there are four talking agents in this system, they never appear at the same time. There are no interactions between agents and they do not appear to coordinate with one another. This type of agent can usually be built independently. The learning state and user actions (e.g., click a button) trigger the presence and absence of each agent. Once present, an agent performs its responsibilities independent of other agents.

Of course, talking agents could interact with each other. The question is, what benefits can be obtained from interactions between talking agents? In the next example, we present a system, Operation ARA, with interactions involving three parties: a tutor agent, a peer student agent and a human learner.

Trialogues in Operation ARA

Operation ARA (Acquiring Research Acumen) is a later AutoTutor system with five talking agents. This system helps learners with critical thinking and scientific reasoning (Halpern et al., 2012; Millis et al., 2011). It covers 21 scientific concepts, such as "theories and hypotheses", "science and pseudoscience", "operational definitions", "control groups", "sample size", etc. Examples used in the system were taken from different science domains, including psychology, sociology, biology and chemistry. The system contains five major components: 1) an intriguing story presented across the whole learning process; 2) an interactive eBook that explains the basic concepts and principles of scientific thinking; 3) a set of tutoring conversations that helps learners understand the eBook chapters; 4) a *Jeopardy!*-like game to identify flaws in different cases; and 5) an interrogation game for learners to identify flawed cases by asking questions.

The five talking agents used in Operation ARA were: Dr. Quinn, Scott, Glass, Tracy, and Broth. Dr. Quinn is a tutor agent who presents all the time, just like a teacher in a classroom, and participates in all activities. Glass and Tracy are peer student agents. Glass participates in conversations for reading comprehension, whereas Tracy participates in competition against a human learner in identifying flaws in case studies. Broth is a storyteller who helps advance the storyline, of whom we will not say much. Scott, is an interrogator in the interrogation game.

Operation ARA uses AutoTutor trialogues to interact with human learners (Cai et al., 2011; So, Zapata-Rivera, Cho, Luce, & Battistiui, 2015). An AutoTutor trialogue is usually a conversation among a tutor agent, a peer student agent, and a human learner. The tutor agent presents problems, asks questions, gives hints, makes judgements on students' answers, and gives immediate feedback. The peer student collaborates with the human learner to construct a solution to the presented problem. Several pedagogical modes were used adaptively, based on the human learner's performance. For the low performance learner, the system employs a vicarious mode in which the interaction is mostly among the tutor agent and the peer student agent. The human learner is occasionally asked simple questions to make sure he or she is engaged in learning. A tutoring mode trialogue suits medium performance learners. In this mode, the conversation is mostly between the tutor agent and the human learner. The peer student in this mode plays a role of a sidekick, giving feedback to the human learner. When the human learner's knowledge is high enough, a teachable agent (Leelawong et al., 2003) mode deploys, in which the human learner is asked to help the peer student agent to construct a solution to a problem. This mode gives the human a chance to learn by teaching. In addition to these three modes, another mode, called competition, pits the peer student against the human learner in a competition to answer a main question or hint/prompt questions. Wallace et al. (2009) demonstrated that this mode promotes high learning engagement.

The responsibilities of Dr. Quinn in Operation ARA are similar to the tutor agent Gavin in MetaTutor. She has the knowledge about critical thinking and scientific reasoning. She is responsible for asking questions, evaluating answers, giving immediate feedback, and presenting the final answer.

Glass participates in three modes of trialogues: vicarious learning, tutoring, and teachable agent. Simulating a peer student, Glass's responsibilities include: 1) answering Dr. Quinn's questions (answers could be good, partial or bad); 2) asking human learner questions (teachable agent mode); and 3) giving feedback to human learner.

Tracy participates in competition mode trialogues, appearing as a peer student. Her programming contains a list of flaws in each of a set of cases. When Dr. Quinn presents a case, Tracy competes against the human learner in identifying flaws in the case. Tracy's responsibility is simple: when it is her turn, she selects a true or false item from the flaw list. If Tracy's score is lower than the learner's, she selects a true item. If her score is higher than the learner's she selects a false item.

Scott participates as an expert interrogator in an interrogation game. In the game, a suspect (hidden agent) presents a case, Scott tries to determine whether or not there is a flaw by asking the suspect questions. Scott has a repository of both good and bad questions, but lets the human learner help him construct questions first. Once a human learner provides a question, Scott matches it to a stored question (good or bad) and presents it to the suspect. Good questions reveal the flaw (if any) more quickly, resulting in higher credits. Scott has only one responsibility: presenting a stored question that best matches the human's. Note that Scott needs the capability of performing semantic matching.

Talking agents in Operation ARA form different types of trialogues associated with adaptive pedagogical strategies. This would not be possible without the addition of peer student agents. Although agents in Operation ARA trialogues have different identities and responsibilities, they do not function independently. In each trialogue, two agents are more like one mind with two heads. It evaluates the human's input and makes decisions on what to say and who to say it to (the tutor agent or the peer student agent). Building independent agents that can participate in trialogues with different roles would be more challenging, because each agent needs to be able to understand the other two parties. The interaction protocol will be more complex and less constrained.

One thing that is common in all of the above systems is that talking agents respond to the human learner's language input. However, in more complex learning environments, human learners may be allowed to *do* in addition to *talk*. Human learners' actions, resulting in the form of "world events" may trigger responses from talking agents. The following examples show how agents can handle this situation.

Talking Agents Sensitive to World Events in CSAL AutoTutor

CSAL (Center for Study of Adult Literacy) AutoTutor was developed to help adult learners improve their reading comprehension skills (Graesser et al., 2017; Lippert, Walker, Davis, & Clewley, 2017). CSAL AutoTutor has 35 lessons focusing on different comprehension strategies. While these lessons allow natural language input, just as the above systems, the target learners have relatively low writing ability and limited knowledge in computer use. To complement natural language interaction, most lessons allow learners to interact with talking agents through interactive pages, including clicking on buttons, selecting items, dragging and dropping objects, highlighting texts, filling in blanks, etc. There are two talking agents: a tutor agent, Christina, and a peer student agent, Jordan. Similar to Operation ARA, Christina, Jordan, and a human learner collaborate through trialogue conversations to construct solutions to problems.

Christina's responsibilities include: 1) presenting a problem; 2) asking a question; 3) judging actions and natural language inputs from the human learner; and 4) providing feedback and correct answers. Jordan's responsibilities include: 1) acting on the media pages; 2) talking about the actions; 3) responding to actions from human learners as a peer student at a level similar to the human learner.

Similar to trialogue agents in Operation ARA, Christina and Jordan function as "one mind, two heads". The difference is that they must also respond to users' actions which results in more complicated conversation rules (Cai, Graesser, & Hu, 2015).

All examples above assume that one single human learner collaborates with one or more talking agents. Our last example is a system developed for team learning.

Group Mentoring in AutoMentor

AutoMentor simulates human mentoring for teamwork on a land science epistemic simulation (Shaffer, 2006; Shaffer et al., 2009). Teams of learners (four per team) collaborate on planning land use, considering various constraints, such as water, housing, waste, pollution, and so on. Learners interact with team members and human mentors through online chats, which provide rich information about the quality of the teamwork, as well as the emerging roles and personalities of team members (Dowell, Nixon, & Graesser, 2018). AutoMentor is a computer agent residing between learners and human mentors. AutoMentor evaluates learners' individual work and team work and suggests feedback to human mentors. Human mentors select and modify AutoMentor's suggestions and deliver them to learners. AutoMentor learns from the human mentors' selections and modifications to make better suggestions.

AutoMentor is more of a “hidden” agent because no avatar appears in the system. AutoMentor’s full responsibilities (not yet fully implemented) include: 1) evaluating design quality in simulated environment; 2) grading individual notes and messages; 3) participating in and monitoring team chats; and 4) guiding teamwork. When a mentor agent participates in a team chat, the conversation rules needed for the agent differ greatly from agents in AutoTutor EMT conversations. For example, because it is a team chat, the mentor agent needs to select complex conversational actions like when to take a turn and whom to address.

So far we have talked about agents that could analyze and respond to learners’ actions and natural language input. Agents can be aware of learners’ emotional states. We present such agents in the next example.

Affect-Aware Talking Agents that Adapt Interaction Style

If an intelligent tutoring system features methods for detecting a variety of user emotions, it can vary the style in which it delivers feedback to users accordingly. For example, MetaTutor changed the level of encouragement based on an estimate of the learner’s negative emotions (Azevedo et al., 2009; VanLehn et al., 2014). Yang and Dorneich (2016; in review) developed a prototype mathematics intelligent tutoring system that dynamically adapts the interaction style between a talking agent and the learner. The talking agent provides feedback delivered in distinct etiquette strategies when the learner struggles solving college-level mathematics problems. Etiquette strategies in human-human communication influence participants’ sense of comfort in social contexts (Brown & Levinson, 1978; Mills, 2003)—influence that human tutors have employed for pedagogical benefit (Person, Kreuz, Zwaan, & Graesser, 1995). The intelligent tutoring system varies etiquette strategy according to the current emotional and performance state of the learner. Specifically, after each math problem, the system measures the learner’s performance, and the learner provides feedback on his or her level of motivation, confidence, and satisfaction. These variables, as well as the learner’s frustration level, determine the agent’s etiquette strategy for the next math problem from among the following possibilities: bald, positive politeness, negative politeness, and off-record (Yang & Dorneich, 2016).

With a bald strategy, the agent displays no consideration for the level of imposition on the learner. Positive politeness minimizes the level of imposition on the learner by the agent, expressing statements of friendship, solidarity, and compliments. Negative politeness, on the other hand, recognizes an imposition while remaining respectful. Finally, the off-record strategy attempts to communicate indirectly, requiring the learner to infer the true intention of the agent.

During a math problem, each one of the six steps of the problem-solving process (as defined by Gordon, 2008) prompts system feedback. For each step, one to three feedback utterances is possible, with at least one proactive (e.g., “Define the variables.”) and at least one reactive (e.g., “It’s not the appropriate formula.”) utterance. The content of the feedback does not change for a specific problem, but the etiquette strategy in which the feedback arrives does. Thus, four versions of each feedback statement of the talking agent are available, one for each etiquette strategy.

Triggers for the strategies depend on the four steps identified by Keller (2009) for encouraging and sustaining students’ motivation in the learning processes: attention, relevance, confidence, and satisfaction (ARCS). Varying feedback etiquette strategy has the potential to improve not only performance, but also motivation, confidence, and satisfaction, even identifying and targeting the factor among these most in need of improvement (Yang & Dorneich, in review). The ability of a talking agent to adapt its interaction style can be used in conjunction with the more traditional tutoring system adaptations of task difficulty or learning content to create tutors that support both the cognitive and affective needs of learners.

Conclusions and Recommendations for Future research

Intelligent computer agents play different roles. An *expert agent* has domain knowledge that allows monitoring of a learner's knowledge state, providing the best learning path, explaining concepts, and correcting misconceptions. In other words, a domain expert agent needs a built-in domain knowledge space (Doignon & Falmagne, 2015) and an understanding of a learner's cognitive state. A *learning strategy agent* needs knowledge of learning strategies to suggest at different phases in the learning process. A *tutor agent* in an intelligent tutoring system could be a combination of a domain expert agent and a learning strategy agent. That is, at any moment in a learning process, an ideal tutor agent knows the best content and the best way for a learner to learn. A more complex *team mentor agent* needs the capability of evaluating group interactions and guiding healthy and productive team interactions, in addition to domain knowledge and pedagogical strategies.

Peer student agents present potentially the most complex instantiation of all. In order to simulate students with different cognitive states, metacognitive states, and personalities, a peer student agent needs an integrated brain of a domain expert agent, pedagogical strategy agent, and a social cognitive agent. A peer student agent is one that knows everything but pretends that it only knows something. The peer student agent may give wrong or partial answers to questions, bad solutions to problems. When a peer student agent participates in a conversation, it often uses language different from a tutor agent. For example, when giving positive feedback to a learner's good answer, a tutor may say: "That is correct!" while a peer student agent may say: "That is what I thought!" Peer student agents may simulate learning from human responses. One peer student agent may "remember" many learners' responses and strategically use the learned responses in future conversations.

Online collaborative learning environments make it possible for learners to interact with other learners, as well as teachers, mentors, and experts from distant locations. The use of intelligent virtual agents in online learning environments has many advantages. For example, it makes the collaborative learning available all the times, bypassing the need to coordinate schedules by providing an omnipresent virtual agent. A team with virtual agents may be adaptively formed according to the learner's personality and performance on target tasks. Peer student agents can be configured with knowledge and skills that optimally compensate for the learners. Effective pedagogical strategies may be more easily and consistently implemented with the help of virtual agents. For example, vicarious learning needs a medium knowledge peer student to interact with a tutor; learning in teachable agent mode needs a peer student with low knowledge. With virtual agents, selecting these tailored team members presents no problem.

The Generalized Intelligent Framework for Tutoring (GIFT) is capable of integrating different types of learning resources to create engaging and effective collaborative learning environments. GIFT emphasizes one-to-one and one-to-many tutoring. We recommend that many (agents)-to-one (learner) and many (agents)-to-many (learners) tutoring should be considered in GIFT. Currently AutoTutor is the only module that has deep conversation agents. Conversational agents should be considered as a "universal interface" for all tutoring modules. This implies three challenges. The first challenge is the natural language processing. While speech recognition technologies have made great improvements in recent years, understanding the meaning of the learner's language input (usually ungrammatical) is still difficult. GIFT should promote further research and development in this direction. The second challenge is the authoring tool of conversational agents. Authoring conversational agents is always difficult and expensive. Standardized authoring tool for conversational agents could be a way to go. GIFT should put more effort in this direction. The last challenge is automatic conversation script generation. Currently agent speeches are authored by human experts. Developing agents that can learn from human learners may lead to better performance and lower development costs.

Conversational agents play important roles in online collaborative learning environments. We are expecting GIFT to provide improved ways of integrating conversational agents into learning modules.

Acknowledgements

The research on was supported by the National Science Foundation (DRK-12-0918409, DRK-12 1418288), the Institute of Education Sciences (R305C120001), US Army Research Laboratory (W911NF-12-2-0030), and the Office of Naval Research (N00014-12-C-0643; N00014-16-C-3027). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, IES, or DoD. The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of researchers from psychology, computer science, and other departments at University of Memphis (visit http://www.autotutor.org).

References

Azevedo, R., Landis, R. S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J. M., … Hossain, G. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7315 LNCS, 212–221. https://doi.org/10.1007/978-3-642-30950-2_27

Azevedo, R., Witherspoon, A., Graesser, A., McNamara, D., Chauncey, A., Siler, E., … Lintean, M. (2009). MetaTutor: Analyzing self-regulated learning in a tutoring system for biology. *Frontiers in Artificial Intelligence and Applications*, *200*(1), 635–637. https://doi.org/10.3233/978-1-60750-028-5-635

Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. Goody (Ed.), Questions and Politeness: Strategies in social interaction (pp. 56-310). Cambridge: Cambridge University Press.

Cai, Z., Graesser, A. C., & Hu, X. (2015). ASAT: AutoTutor Script Authoring Tool. In R. Sottilare, A. C. Graesser, X. Hu, & K. Brawner (Eds.), Design Recommendations for Intelligent Tutoring Systems: Authoring Tools (pp. 199–210). Orlando: Army Research Laboratory.

Cai, Z., Graesser, A., Forsyth, C., Burkett, C., Millis, K., Wallace, P., … Butler, H. (2011). Trialogue in {ARIES}: User input assessment in an intelligent tutoring system. Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems, (March 2016), 429–433. https://doi.org/10.13140/2.1.4284.5446

Doignon, J.-P., & Falmagne, J.-C. (2015, November 20). Knowledge Spaces and Learning Spaces. arXiv [math.CO]. Retrieved from http://arxiv.org/abs/1511.06757

Dowell, N., Nixon, T., & Graesser, A. (2018). Group Communication Analysis: A Computational Linguistics Approach for Detecting Sociocognitive Roles in Multi-Party Interactions. Retrieved from http://arxiv.org/abs/1801.03563

Golub, J. (1988). Focus on Collaborative Learning. Classroom Practices in Teaching English, 1988. National Council of Teachers of English (NCTE).

Gordon, T. (2008). Parent effectiveness training: The proven program for raising responsible children. Harmony.

Graesser, A. C. (2016). Conversations with AutoTutor Help Students Learn. *International Journal of Artificial Intelligence in Education*, *26*(1), 124–132. https://doi.org/10.1007/s40593-015-0086-4

Graesser, A. C., Cai, Z., Baer, W. O., Olney, A. M., Hu, X., Reed, M., & Greenberg, D. (2017). Reading Comprehension Lessons in AutoTutor for the Center for the Study of Adult Literacy. In S. Crossley & D. S. McNamara (Eds.), Adaptive Educational Technologies for Literacy Instruction (pp. 288–294). New York: Routledge.

Graesser, A. C., Chipman, P., Haynes, B., & Olney, A. M. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4), 612-618.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research, 1*(1), 35–51. https://doi.org/10.1016/S1389-0417(99)00005-4

Graesser, A., & McNamara, D. S. (2010). Self-regulated Learning in Learning Environments with Pedagogical Agents that Interact in Natural Language. Educational Psychologist, 2579, 4. https://doi.org/10.1080/00461520.2010.515933

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity, 7*(2), 93–100. https://doi.org/10.1016/j.tsc.2012.03.006

Keller, J. M. (2009). Motivational design for learning and performance: The ARCS model approach. Springer Science & Business Media.

Landauer, T., McNamara, D., Dennis, S., & Kintsch, W. (2007). Handbook of Latent Semantic Analysis. LSA A Road Towards Meaning. Retrieved from http://books.google.co.uk/books?hl=en&lr=&id=jgVWCuFXePEC&oi=fnd&pg=PR9&dq=latent+semantics+analysis&ots=wX1WCWsuEm&sig=t2UoB1sOpfphIwCRLab26Qr6LvY

Leelawong, K., Viswanath, K., Davis, J., Biswas, G., Vye, N., Belynne, K., & Bransford, J. (2003). Teachable Agents Learning by Teaching Environments for Science Domains. Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence, (August 2015), 109–116.

Lippert, A., Walker, B., Davis, R., & Clewley, D. (2017). The Effect of CSAL AutoTutor on Deep Comprehension of Text in Low-Literacy Adult Readers. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), Proceedings of the 39th Annual Meeting of the Cognitive Science Society (p. 3766). Austin, TX: Cognitive Science Society.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011). Operation ARIES!: A Serious Game for Teaching Scientific Inquiry. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.), Serious Games and Edutainment Applications (pp. 169–195). London: Springer London. https://doi.org/10.1007/978-1-4471-2161-9_10

Mills, S. (2003). Gender and politeness (Vol. 17). Cambridge University Press.

Person, N. K., Graesser, A. C., Kreuz, R. J., & Pomeroy, V. (2003). Simulating human tutor dialog moves in AutoTutor. *International Journal of Artificial Intelligence in Education (IJAIED)*, *12*, 23-39.

Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and instruction, 13*(2), 161-188.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. Computers and Education, 46(3), 223–234. https://doi.org/10.1016/j.compedu.2005.11.003

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., … Mislevy, R. J. (2009). Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning. *International Journal of Learning and Media*, *1*(2), 33–53. https://doi.org/10.1162/ijlm.2009.0013

So, Y., Zapata-Rivera, D., Cho, Y., Luce, C., & Battistinui, L. (2015). Using trialogues to measure eEnglish language skills. *Educational Technology and Society*, *18*(2), 21–32.

VanLehn, K., Burleson, W., Girard, S., Chavez-Echeagaray, M. E., Gonzalez-Sanchez, J., Hidalgo-Pontet, Y., & Zhang, L. (2014, June). The affective meta-tutoring project: lessons learned. In International Conference on Intelligent Tutoring Systems (pp. 84-93). Springer, Cham.

Wallace, P., Graesser, A., Millis, K., Halpern, D., Cai, Z., Britt, M.A., Magliano, J. & Wiemer, K. (2009). Operation ARIES!: A Computerized Game for Teaching Scientific Inquiry. 602-604. 10.3233/978-1-60750-028-5-602.

Yang, E., & Dorneich, M.C. (2016).Evaluation of Etiquette Strategies to Adapt Feedback in Affect-Aware Tutoring. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Washington DC, September 19–23, 2016.

Yang, E. & Dorneich, M.C. (In Review). Affect-Aware Adaptive Tutoring Based on Human-Automation Etiquette Strategies. *Human Factors*.

Chapter 17 – Automating the Assessment of Team Collaboration through Communication Analysis

Peter W. Foltz
Pearson and University of Colorado, Boulder

Introduction

Regular monitoring of teams of decision-makers in complex collaborative environments requires effective and efficient tracking. Current methods of assessing team and group performance often must rely on temporally delayed outcomes or global metrics that are insufficiently detailed to detect the cause of failures. For example, human monitoring of teams typically occurs after the fact, or requires large investments of manpower from supervisors or trainers. A largely untapped source of timely and diagnostic information lies in ongoing communications among team members. Communications can reflect cognitive and task states, knowledge, situation awareness, stress, vigilance, and who is communicating. With appropriate analyses, the communication data can be tied back to both the team's and each individual's abilities and knowledge.

Automated analysis of communication offers the possibility of real-time diagnosis and warning along with assessment of team and user states and abilities. This chapter describes a series of studies in which we apply speech recognition, speech analysis, and statistical natural language-based techniques to analyze the communications of teams in training and operational environments. While the method has been tested across a broad range of domains and tasks, the chapter provides details on performance in one domain (convoy operations). The techniques are able to provide accurate predictions of the overall team performance, make reliable judgments of the type of statements each team member is making, and predict team performance problems based on the language and patterns of communication among team members. Such performance problems include loss of situation awareness, knowledge gaps, workload, and critical incidents.

Overall, the application of the approach demonstrates that the methods can have broad use for varied teams in different types of collaborative situations. The chapter concludes with implications for applying the approach within small and large-scale training situations, considerations for the level of effort for customized development for different training situations, and types of feedback for individuals and trainers.

Communication in Teams

Communications in Teams

Teams performing collaborative tasks provide a rich source of information about their performance through their verbal communication. The communication can provide information both about the structure of their social network and the content and quality of information flowing through the network. This information reflects team member roles, connectedness, as well as how they are performing their tasks. Additionally the content and manner of the communication provides information about the team's cognitive states, knowledge, errors, information sharing, coordination, leadership, stress, workload, intent, and situational status. Thus, communication provides a window into team cognition.

Studies of team communication have shown that the content and pattern are predictive of performance (e.g., Contractor & Grant, 1996; Cooke, Duchon, Gorman, Keyton & Miller,2012). Such studies have analyzed categories of communication being passed, frequencies of counts of communications, units of meaning, and patterns of discourse among team members. For example, Achille, Schulze and Schmidt-Nielsen (1995) found that the use of military terms, acknowledgments, and identification statements increased with experience. Similarly, Jentsch, Sellin-Wolters, Bowers and Salas (1995) found that teams that were able to identify typical flight problems faster made more leadership statements and more observations about the environment. However, many of these studies relied on hand-coding of the data and/or laborious post-hoc analyses.

Language as Data

With the advent of digital systems for capturing and interpreting spoken audio, communication can be treated as a formal type of data about teams and can be analyzed to reveal information about team performance. Automated transcription through speech recognition, along with machine learning, language processing, speech analysis and probabilistic modeling can be applied to distill and understand critical features in teams. For instance, computational linguistic methods are used to break down individual words and phrases from communication streams in order to characterize the types of communication or the quality of a team member's knowledge. Patterns of communication can be analyzed through Markov models or probabilistic chains to determine where and how communication flows and whether the patterns of interaction are expected given the nature of the task. These methods have been shown to be able to predict overall team performance (e.g. Gorman, Foltz, Kiekel, Martin & Cooke, 2003), detect patterns of communication and locations of communication breakdown (Kiekel, Gorman & Cooke, 2004; Gorman, Martin, Dunbar, Stevens & Galloway, 2015), categorize types of communications and how those types relate to team tasks (Foltz & Martin, 2008).

The Communication Analysis Pipeline

In the present work, we have developed a generalized framework, or pipeline for analysis of communication data. A major goal of the work has been to apply the same framework across varied types of data. The communication analysis pipeline is designed to take communication data as input, format it into usable means (e.g., transcribing spoken text to data), convert it into sets of language features, and then apply those features in machine learning-based modeling to generate predictions about the performance of the team. This communication analysis pipeline is shown in Figure 1. We outline the key steps used in the pipeline and then illustrate its use across a range of domains and studies.

Due to the complexity of teams, modeling often must account for individual knowledge, team knowledge, team skills, team dynamics, and interactions with tasks as well as team members. One of the simplifying assumptions in the communication analysis pipeline is that the goal is not to build complete and specific domain models for each domain. Instead, the process relies on using information about team performance, such as ratings of performance by experts, objective measures of success, and indicators of troubles and success on individual tasks. The process then uses computational language methods to break the communication into raw features about communication and then applies machine learning to learn to associate those features to the information about team performance. Thus, it "learns" what features in language are indicative of different kinds of performance. This learned model can then be used to predict performance on new teams or the same team in subsequent tasks. In a sense, the process learns to model what humans do when they listen to teams to distill useful characterizations about the domain and situation.

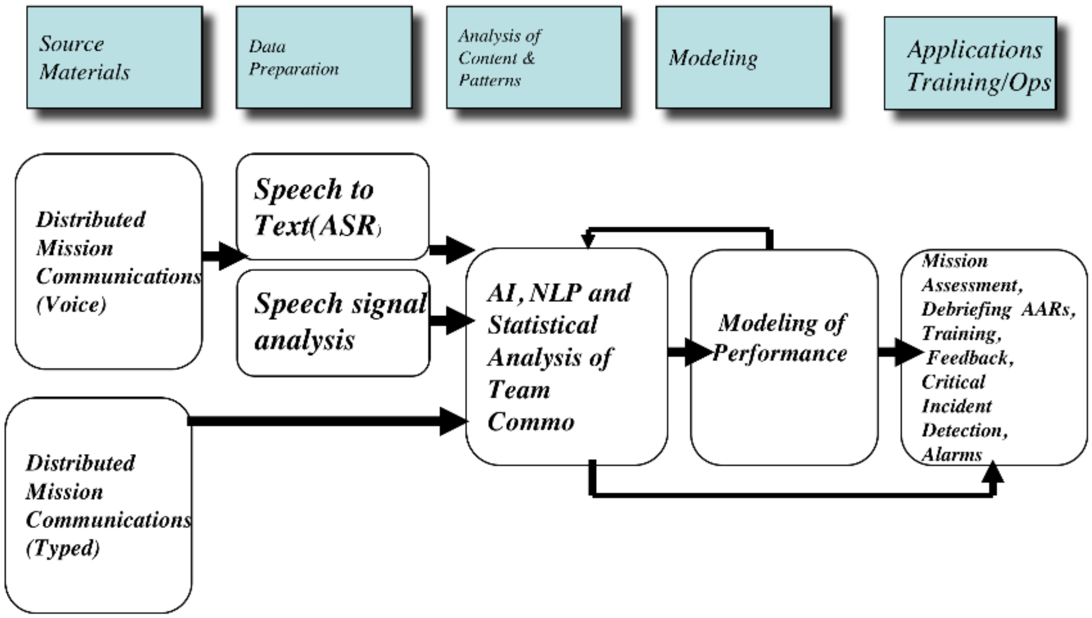


Figure 2. The Communication Analysis Pipeline.

Speech processing

Since communication is typically spoken, automatic speech recognition (ASR) systems can be applied for converting speech to text for input into processing stream. Speech recognition can be imperfect, most noticeably in noisy environments (e.g., Schmidt-Nielsen et al., 2001). However, even with errors in transcription, modeling can often still perform very well at predicting performance. For example, even with typical ASR systems degrading word recognition by 40%, the modeling prediction performance degraded less than 10% (See Foltz, Laham & Derr, 2003). This is because verbal interactions in such situations are highly constrained by the actions currently being taken and by the current execution status of the mission plans. Since verbal interactions are largely routinized, the difficulties of both automatic speech recognition and machine learning-based modeling understanding are greatly reduced. Moreover, because these techniques derive meaning from whole utterances, not from individual words, it tends to be more immune to high word level error rates typically found in speech recognition systems. Along with performing ASR, speech data can be analyzed for signal properties to detect voice stress. For example, Foltz, Rosenstein, LaVoie, Oberbreckling, Chatham and Psotka (2008) analyzed the speech power, pitch, change over time, rate, duration and frequency component to build an excitement detection algorithm. They found that the algorithm was accurate over 80% of the time and could predict significant variance related to ratings that Subject Matter Experts (SMEs) made about teams when they encountered critical events.

Content modeling

Within the analysis of content and language patterns features and modeling, a range of techniques have been applied. One notable technique has been to use Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998) to analyze and interpret the content of what is being said during communication. Since the technique analyzes *meaning* rather than the direct words used, it is able to account for variability that may reflect how different people express similar events or situations in team situations. The details of the approach are beyond the scope of this chapter, but the methods applied to team analyses are described in detail elsewhere (Foltz, Martin, Abdelali, Rosenstein & Oberbreckling, 2006; Gorman et al.,2003; Kiekel, Cooke, Foltz, Gorman & Martin, 2002).

Domains Analyzed

The approach discussed in this chapter has been applied across a wide range of studies in varied domains, with different types of tasks and different types of communication. Table 1 describes these domains and provides pointers to relevant work.

Table 1. Domains analyzed.

Domain/Task	Nature of communication data	Representative publications
AFRL MESA DMT F-16 simulators	229 engagements of 4 Ships + AWACS	Foltz, Oberbreckling & Laham (2012)
NAVY TADMUS Command and Control	~80 hand transcribed Command and Control missions	Foltz, LaVoie, Oberbreckling and Rosenstein (2008)
UAV simulated missions	~300 Hand and ASR Transcribed missions	Kiekel et al., (2002) Gorman et al. (2003), Foltz et al. (2006)
ARL Intelligence decision making during STTR missions	480 hand transcribed missions	Foltz, Lee, Bond & Martin (2009)
Convoys in Ambush! virtual environment and National Training Center live STX training	--100s of hours of missions with team size ranging from 10 to 50. ASR and hand transcribed	Foltz, Rosenstein et al., (2008), Lavoie et al. (2008)
Navy Non Combatant Extraction Operation (NEO)	32 Face to Face and Asynchronous hand transcribed missions	Foltz, Lavoie, Oberbreckling & Rosenstein (2008)
Army/Air Force Officers in chat room-based planning exercises	1000s of chat interactions	Lavoie et al. (2010)
NPS command and control planning task	100s of Predictions of Situation Awareness (ONR)	Bolstad et al. (2010)
US/Singapore planning operation to evaluate systems interoperability	One multi-day mission	Pierce et al. (2006)

Communications in Convoy Training

As a more specific example, Foltz et al., 2009 developed a system to predict team performance and detect critical events in convoy operations. The data collected included teams of up to 50 soldiers using the Ambush! virtual convoy training (Diller, Roberts & Willmuth, 2005) and teams of up to 40 participating in live convoys performing STX training at the National Training Center. In all cases, company-sized elements conducted mounted tactical road march along specified routes and encountered various events including a civil disturbance, RPG attack, IEDs, a near ambush, negotiation with Iraqi police and complex attacks (IED and ambush).

All audio data from the team members was recorded and ASR transcribed, logs of events that occurred through the training were collected and results from the After Action Reviews (AARs) were recorded. A team of domain experts listened to all the recordings and rated teams on performance as well as marking and rating particular critical events. Teams were rated by the SMEs on a 5 point scale for measures including: Battle drills, Command & control, Situation understanding, Adherence to Standard Operating Procedures (SOPs), and overall team performance. SMEs correlated with each other with an r = 0.34 to 0.59 for individual events and an r = 0.76 to 0.85 averaged across events.

Critical incident detection model

A computational model to detect when teams were having critical incidents was built. The method used a moving window time series analysis that generated semantic, statistical, and syntactic properties of segments of communication. The model analyzed a 1 minute window of communication and then moved ahead by 10 seconds. Classification algorithms compared segments to discourse properties of the communication from critical incidents identified by SMEs to predict if an event was happening. The results showed that the model could correctly classify 84% of the events and detect more than half of the critical events (See Foltz, Rosenstein et al., 2008; Lavoie et al., 2008).

Predicting Performance

Performance predictions were generated for each rating metric (e.g., situation understanding, command and control, etc.) for each team, both for overall missions as well as for each critical event during the mission. In each case, the system used a machine learning-based approach in which the system was trained on a subset of team mission data in which it learned to associate communication variables with SME ratings. Then the performance predictions were generated for held-out sets of team missions. The results showed that the system could correlate with SME ratings for individual events with correlations ranging from 0.37 to 0.43 (within the range of human rating agreement). For predicting overall team performance on a mission, the model correlated with SMEs ranging from 0.71 to 0.83. These results indicate that the approach both accounts for a large amount of the variance of team performance as well as corresponds closely to the intuitions of SMEs.

Conclusions and Recommendations for Future Research

Communication is a natural part of the team process and provides one of the best windows into team cognition. Applying automated methods to analyzing the streams allows near-realtime analysis and modeling of real, complex tasks. The communication analysis pipeline breaks spoken as well as written communication into features, then models and predicts objective and subjective metrics of performance. A distinct advantage of the approach is that the models can be automatically derived, not requiring large efforts in up-front tasks analyses. Overall, the use of the approach allows analysis of teams while they perform in natural environments without intrusive data collection methods.

As an application, the approach allows the development of tools that can be integrated into systems to monitor teams and provide timely feedback. It can detect teams and individuals who are performing poorly. Thus, it can be used to provide AARs or determine when retraining is necessary. It can further detect critical events that may significantly change the course of a mission. This allows supervisors to monitor multiple teams and know which teams need the most attention as well as to be able to quickly locate parts of team records which may be most useful for feedback. Finally, the approach has the potential for contributing to an adaptable training system, which can monitor team performance and, in realtime, change the simulation in order to best train individuals and teams.

References

Achille, L. B., Schulze, K. G., & Schmidt-Nielsen, A. (1995). An analysis of communication and the use of military terms in Navy team training, *Military Psychology, 7*, 95-107.

Bolstad, C. A., Cuevas, H. M., Connors, E. S., González, C., Foltz, P. W., Lau, N. K. C., & Warwick, W. J. (2010). Advances in modeling situation awareness, decision making, and performance in complex operational environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 54, No. 13, pp. 1047-1051). SAGE Publications.

Contractor, N. S., & Grant, S. J. (1996). The emergence of shared interpretations in organizations: a self-organizing systems perspective. In J. H. Watt and A.C. VanLear (Eds.), *Dynamic Patterns in Communication Processes*, pp. 215-230. Thousand Oaks, CA: Sage Publications.

Cooke, N. J., Duchon, A., Gorman, J. C., Keyton, J., & Miller, A. (2012). Preface to the special section on methods for the analysis of communication. *Human Factors*, *54*(4), 485-488. DOI: 10.1177/0018720812448673

Diller, D. E., Roberts, B. & Willmuth, T. (2005). DARWARS Ambush! A case study in the adoption and evolution of a game-based convoy trainer with the U.S. Army. Presented at the *Simulation Interoperability Standards Organization*, 18-23 September.

Foltz, P. W., Lavoie, N., Oberbreckling, R. & Rosenstein, M. (2008). Automated performance assessment of teams in virtual environments. In D. Schmorrow, J. Cohn & D. Nicholson, (Eds.), *The PSI handbook of virtual environments for training and education: Developments for the military and beyond*. Westport, CT: Praeger Security International.

Foltz, P. W., Lee, A. Y., Bond, G., & Martin, M. J. (2009). Models, Methods, and Tools for the adaptability of teams. In P. McDermott and L. Allender (Eds.). *Advanced Decision Architectures for the Warfighter*. Army Research Lab, Aberdeen, MD.

Foltz, P. W., Laham, R. D. & Derr, M. (2003). Automated Speech Recognition for Modeling Team Performance. In *Proceedings of the 47th Annual Human Factors and Ergonomic Society Meeting.*

Foltz, P. W. & Martin, M. J. (2008). Automated Communication Analysis of Teams. In E. Salas, G. F. Goodwin, & S. Burke (Eds.) *Team Effectiveness in Complex Organizations and Systems: Cross-disciplinary perspectives and approaches*. New York: Routledge.

Foltz, P. W., Martin, M. A., Abdelali, A., Rosenstein, M. B. & Oberbreckling, R. J. (2006). Automated Team Discourse Modeling: Test of Performance and Generalization. In *Proceedings of the 28th Annual Cognitive Science Conference.*

Foltz, P. W. Oberbreckling, R. J., Laham, R. D. (2012), Analyzing Job and Occupational Content Using Latent Semantic Analysis. M. A. Wilson, W. Bennett, Jr., S. G. Gibson, G. M Alliger (Eds.). *The Handbook of Work Analysis Methods, Systems, Applications and Science of Work Measurement in Organizations.* Routledge. 691-709.

Foltz, P. W., Rosenstein, M., LaVoie, N., Oberbreckling, R., Chatham, R., & Psotka, J. (2008). DARCAAT: DARPA Competence Assessment and Alarms for Teams. *Proceedings of the Interservice/Industry Training Simulation, and Education conference I/ITSEC,* Orlando, FL

Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. A. & Cooke, N. J. (2003). Evaluation of Latent Semantic Analysis-based measures of team communications content. In *Proceedings of the 47th Annual Human Factors and Ergonomic Society Meeting.*

Gorman, J. C., Martin, M.J., Dunbar, T. A., Stevens, R. H., Galloway, T. (2013) Analysis of Semantic Content and Its Relation to Team Neurophysiology during Submarine Crew Training. *Foundations of Augmented Cognition: 7th International Conference, AC 2013, Held as Part of HCI International 2013,* Las Vegas, NV, USA, July 21-26, 2013.

Jentsch, F., Sellin-Wolters, S., Bowers, C. & Salas, E. (1995). Crew coordination behaviors as predictors of problem detection and decision making times. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting.*

Kiekel, P. A., Cooke, N. J., Foltz, P. W., Gorman, J & Martin, M. J. (2002). Some promising results of communication-based automatic measures of team cognition. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting.*

Kiekel, P., Gorman, J., & Cooke, N. (2004). Measuring Speech Flow of Co-located and Distributed Command and Control Teams During a Communication Channel Glitch. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 683-687.

Landauer, T. K, Foltz, P. W. & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes*, *25*(2&3), 259-284.

LaVoie, N., Foltz, P. W., Rosenstein, M., Oberbreckling, R., Chatham, R., & Psotka, J. (2008). Automated support for AARs: Exploiting communication to assess team performance. *Proceedings of the Interservice/Industry Training Simulation, and Education conference I/ITSEC*, Orlando, FL

LaVoie, N., Streeter, L, Lochbaum, K, Wroblewski, D, Boyce, L, Krupnick, C., Psotka, J., (2010) Automating Expertise in Collaborative Learning Environments, *Journal of Asynchronous Learning Networks*, v14 n4 p97-119

Pierce, L. G., Sutton, J. L., Foltz, P., Lavoie, N., Scott-Nash, S., & Lauper, U (2006) Technologies for augmented collaboration. *In Proceedings of the CCRTS Command and Control Research and Technology Symposium*. Cambridge, UK.

Schmidt-Nielsen, A., Marsh, E., Tardelli,, J., Gatewood, P., Kreamer, E., Tremain, T., Cieri, C., Strassel, S. Martey, N., Graff, D. & Tofan, C. (2001). *Speech in Noisy Environments (SPINE2) Part 1 Audio*. Linguistics Data Consortium catalog: LDC2001S04.

Acknowledgements

This work has broadly benefited from collaboration with many colleagues including Mark Rosenstein, Rob Oberbrecking, Noelle Lavoie, Kyle Habermehl, Nancy Cooke, Jamie Gorman, Preston Kiekel, Melanie Martin, Ahmed Abdelali, Cheryl Bolstadt, Joseph Psotka, Linda Pierce, Ralph Chatham, and Fred Flynn. The work has been funded by the AFRL, ONR, ARI, ARL, and DARPA.

SECTION IV - SYSTEM DESIGN OF TEAM TUTORS

Dr. Robert A. Sottilare, Ed.

Chapter 18 – System Design for Intelligent Tutoring of Team Taskwork

Robert A. Sottilare
US Army Research Laboratory

Core Ideas

The chapters in this section focus on the ***system design*** aspects for the tutoring of team taskwork using intelligent tutoring systems (ITSs). When we discuss system design, we are specifically addressing how ITSs are designed to assess and interact with learners. Since much more effort has been focused on ITSs to support individual learning rather than teams, we are also attempting to address two question:

- *What should be fundamentally different about ITSs for teams from their individual learner ITS counterparts*?
- *What does the ITS need to know about the team to make effective decisions about feedback, support, and guidance for the team*?

During team instruction, team members interact with each other as well as with the tutor. Based on this interaction, we can separate team attitudes, behaviors, and cognition into those related to team taskwork and teamwork. Team taskwork is fundamentally task dependent and tied to the goals, measures, and assessments of learning and performance of the task under instruction. Teamwork is fundamentally task independent in that it is tied to the inner workings of the team: communication, collaboration, conflict management, cognition, coaching, coordination, and the conditions under which the team works together. In this section of the *Design Recommendations for Intellligent Tutoring Systems – Volume 6*, we will discuss how both team taskwork and teamwork influence ITS design, but will focus more heavily on team taskwork.

In general, the system design process accounts for the definition of the architecture (framework), components (hardware and software), interfaces, and data distribution mechanisms for a system to satisfy a set of specified requirements. In our case, ITSs are adaptive instructional systems and the requirements might be distilled down to: *a computer-based system that guides the learning experiences of teams by tailoring instruction and recommendations based on team goals, team roles, team and individual learning needs, and individual preferences in the context of domain learning objectives*.

It might also be noted that the structure of teams and the nature of team tasks could and probably should influence the design of a team ITS. There are many types of team configurations, but most are variations on a set of themes:

- Single leader or authority
- Multiple leader
- Collaborative or non-hierarchical

Teams by their nature as a system require energy to maintain their focus and accomplish their missions. Since team structures involve a division of labor, there are processes required to coordinate the various roles within the team and communicate goals and progress toward those goals. Authority within the team, planning and reporting mechanisms, rules and policies should all be well understood by its members, and may also influence system design.

Teams may be focused on tasks in which the mission varies so there are consequences on our systems approach to adaptive instruction:

- *Work or process teams*. Individuals working together to accomplish a task or complete a process
- *Collaborative teams* – individuals working together to solve a problem
- *Review teams* – individuals working together to insure a standard level of quality in a product or process
- *Distributed or virtual teams* – geographically separated individuals working together in a shared digital environment to accomplish a goal
- *Adhoc teams* – individuals rapidly assembled to perform a short-term task

The individuals working on a team may also take on a variety of roles and responsibilities, which also affect interaction design between the tutor and team members:

- *Task related roles* – cost/schedule/performance/risk manager, reporter, information or opinion giver/seeker, clarifier, summarizer
- *Social or maintenance roles* – motivator, gatekeeper, harmonizer, issue tracker
- *Dysfunctional roles* – dominator, aggressor, sabbeteur

The importance of the role of leadership within a team cannot be overstated. The patterns of behavior used to influence the behaviors of others on the team run the gamut across task related, social, and maintenance roles. At their simplest, leadership can be divided into directive and supportive behaviors. Directive leadership behaviors (e.g., guidance, control, supervision) are used in setting goals, organizing work, communicating priorities, setting timelines, clarifying roles, and modeling tasks. Supportive leadership behaviors (e.g., listening, praising, and facilitating) are used in coaxing input from team members, encouraging the team, sharing experiences, easing team member problem solving, and suggesting solutions.

Leadership plays a large role in ITSs for teams. The tutor may take on a leadership role in guiding, facilitating, and recommending new goals to the team. The system design of the team ITS is influenced by the team's need to see the ITS as an credible and effective leader to encourage engagement and learning. Both the competence and commitment of the tutor may be weighed by the team members. The competence element of leadership may be assessed via demonstrations of the tutor's task knowledge, task skill, experience in the domain, and ability to transfer knowledge and skill to other domains. The commitment element of leadership is weighed in the team member's perception of the tutor's motivation to help them learn and the confidence with which the tutor facilitates the learning experience.

The chapters in this section cover topics related to team tutoring including authoring, review of a prototype system design for teams, a model of dynamic team formation and taskwork, and experimentation tools to evaluate team tutoring designs.

Individual Chapters

The chapter by ***Burke, Sottilare, and Gilbert***, "Leveraging Team Taxonomic Efforts for Authoring" asserts that all teams are not created equal and provides a detailed evaluation of these differences, and their influence on the design of effective authoring systems. Their review of the literature highlights the taxonomic structure of teams and how this work might be used to facilitate the development of an authoring system designed around the key features common to all teams.

The chapter by ***Brawner, Sinatra, and Gilbert***, "Lessons Learned Creating a Team Tutoring Architecture:

Design Reconsiderations" discusses a prototype team ITS developed using the Generalized Intelligent Framework for Tutoring (GIFT) as an adaptive instructional driver for a commercial game, Virtual Battle Space. The prototype team ITS was developed to tutor a small team on the task of surveillance and was created in part to unify disparate research into a tangible model of team tutoring design principles. This chapter reports on the experience of constructing a team ITS using GIFT, and discusses the challenges and limitations of using an ITS architecture originally designed to tutor tasks for individual learners.

The chapter by ***Huang, Folsom-Kovarik, Hu, Du, and Yang,*** "Team Taskwork for Knowledge Building in Cyberlearning" discusses knowledge building in developing teams via digital technologies (cyberlearning). The authors draw heavily from the computer-supported collaborative learning literature to consider three modes of learning: individual, collaborative, and collective. Based on an analysis of social learning, knowledge building and taskwork, the authors propose a structural model for dynamic team formation and taskwork based on the literature. The proposed model will enable evaluation of the process and effect of knowledge building and team taskwork by considering each individual's contribution to the team.

Finally, the chapter by ***Boyce, Sinatra, Gilbert, and Sottilare***, "Developing the GIFT Event Report Tool to Support Experimentation for Teams" describes the growing need for an experimental mechanism to collect and analyze interaction data with team tutors as they evolve. The mechanism for collecting team-based data uses the GIFT event report tool (ERT) as a starting point for structuring, exporting, and analyzing data and reports.

Chapter 19 – Leveraging Team Taxonomic Efforts for Authoring

C. Shawn Burke[1], Robert Sottilare[2], and Stephen Gilbert[3]
University of Central Florida[1], US Army Research Laboratory[2], and Iowa State University[3]

Introduction

The use of teams has become ubiquitous within organizations of all types. However, organizations often falter in the design and implementation of teams as there is often an assumption that a team of experts will automatically culminate in an expert team capable of collaborating and coordinating their actions in an interdependent manner. Unfortunately, practical examples abound to indicate this is simply not true. Research on team training has a long history, but most recently there has been an increased focus on building intelligent tutoring systems for teams whereby instructor burden could be lessened. Building upon prior research, ITSs are only as good as the science upon which their development lies (Salas & Burke, 2002). To be useful, systems need to be systematically developed based on the science of learning and teams, yet also remain flexible to handle the complexity inherent in teams.

Research has suggested that not all teams are created equal and while a general set of team competencies can be identified, task/team specific competencies also exist (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995). Thereby, one is faced with the question of how to build a system that can not only provide instruction with relation to transportable competencies, but guide those charged with authoring to additional competencies which are more task/team specific. A first step in this process is to understand the ways in which teams differ, which can then be used to develop a set of systematic questions to guide scenario development/selection. In this vein, the current chapter will highlight taxonomic work on teams (e.g., Sundstrom et al., 1990; Wildman et al., 2012). It will then utilize that work to illustrate how such taxonomic efforts can facilitate the development of an efficacious authoring system designed around core features common to all teams, yet diagnostic in its guidance. Finally, the chapter will conclude with a description of challenges and key decisions to be made along the way.

Understanding Teams

Teams have been defined as a set of two or more individuals working collaboratively and interdependently toward a common valued goal (Salas et al., 1992). There are a number of important components within the definition of teams which can provide some initial insight that can assist in the development of a front-end architecture for team ITSs that can be used to guide practitioners in their choice of the competencies to be trained (and correspondingly, training content or scenarios to be used). Perhaps the first question is the degree of teamness that is actually present in the collective entity whose members are to be trained. Determining whether the collective is really a team is the first step in understanding whether or not a team ITS is truly needed, as well as those competencies that might need to be targeted.

To determine the degree of 'teamness', one can look to team definitions to extract core features. One such core feature, revolves around a question as to whether the targeted collective is comprised of two or more individuals who are working towards a common shared goal? A second, and perhaps more complex feature, is the degree to which members must rely on one another to complete their tasks. This feature has been argued to be one factor driving the particular competencies that are important to train as well as serving as a distinguishing feature between groups and teams. Seminal work in this area conducted by Saavedra, Earley, and Van Dyne (1993), delineated four types of task interdependence which vary in the degree to which team members are coupled with one another: pooled, sequential, reciprocal, and team. Pooled interdependence refers to tasks in which members all make a contribution to the group output, but there does not need to be direct interaction amongst members. In this type of interdependence, Saavedra et al. (1993) argue that each member typically has the ability to complete the whole task (i.e., performance is the sum of the individual performances). Moving up a level, sequential interdependence represents a one-way work flow whereby one member must complete an action prior to another member being able to act. The typical example provided for this type of interdependence would be an assembly line whereby members must rely on one another to complete the task, but coordination is dyadic in nature. In this case, members perform different aspects of the task (i.e., no one person can perform the whole task), but in a prescribed order. Greater coordination amongst members is needed in the third type of interdependence, reciprocal. Reciprocal interdependence has been descsribed as, "…temporally lagged, two way interactions", whereby "Person A's output becomes Person B's input and vice versa" (Saavedra et al., 1993, p. 63). It is dyadic coordination whereby members perform different portions of the task (similar to sequential interdependence), but do so in a flexible manner (i.e., bi-directional workflow). Finally, under team interdependence the task requires the greatest degree of coordination and tight coupling of team members in order to accomplish the task. Specifically, Saavedra et al. (1993) argue that this type of task interdependence requires members to, "…jointly diagnose, problem solve, and collaborate to complete a task" (p. 63) and in its truest form members have the freedom to design their own jobs.

The above work could be used by those designing the architecture to assist in first determining whether or not the targeted collective is truly a team versus a more loosely coupled group. In combination with this initial set of questions, is a subsequent set of questions which begin to provide insight on a baseline set of competencies to be trained. For this later set, we turn to the literature on the science of teams.

Establishing a Baseline

Early literature on teams argued that in order to effectively train teams, two tracks of skills must be mastered: taskwork skills and teamwork skills (Morgan, Salas, & Glickman, 1993). Traditionally, taskwork skills have represented the bulk of the focus within intelligent tutoring systems and other training-related interventions. However, in building intelligent tutoring systems for teams, those responsible for building the architecture underlying the system must move beyond taskwork skills to include teamwork skills. Building intelligent tutoring systems for teams represents a complex endeavor with a number of challenges, not the least of which is to build tools on the front end of the architecture which will facilitate the proper selection of training scenarios (and their corresponding metrics), such that the most efficacious set of training experiences can be provided based on the particular needs of the trainee.

An early consideration in moving towards this goal is the nature of the team to be trained. There has been much work conducted on the science of teams over the past thirty years which can help in this vein. For example, in delineating the competencies that are needed by organizational teams, researchers have argued that team competencies vary along two dimensions: task and team (Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995). The task dimension represents that competencies can be task-generic or specific. Task-generic competencies are ones that should facilitate performance regardless of the specific task engaged. Task-specific competencies, in contrast, are driven by the particular task in question and therefore must be trained with that task as a baseline. With regard to the team dimension, team-specific competencies are those that are important based on the specific composition of the team. In contrast, team-generic competencies are those that would influence performance regardless of the specific team members involved. These later competencies are transportable across teams and should serve members well, irrespective of the particular team members who comprise the team.

The above work not only serves to highlight a subset of broad considerations that should be considered by those responsible for developing the content within an intelligent tutoring system for teams, but combining the task and team dimensions produces a 2 x 2 matrix. The 2 x 2 matrix highlights four broad types of team competencies: context driven (i.e., task and team specific), task contingent (i.e., task specific, team generic), team contingent (i.e., team specific, task generic), and transportable (i.e., team and task generic)(Cannon-Bowers et al., 1995). Examples of context driven team competencies include: (1) knowledge - task-specific teammate characteristics, team-specific role responsibilities, cue-strategy associations, (2) skills - mission analysis, compensatory behavior, and (3) attitudes - collective efficacy, shared vision. Examples of task contingent competencies include: (1) knowledge - task specific role responsibilities, task sequencing, team-role interaction patterns, (2) skills - mission analysis, compensatory behavior, information exchange, (3) and attitudes - task-specific teamwork attitudes. Examples of team contingent competencies include: (1) knowledge - teammate characteristics, relationship to larger organization, (2) skills - conflict resolution, motivation of others, compensatory behavior, and (3) attitudes - team cohesion, mutual trust. Finally, examples of transportable competencies include: (1) knowledge of teamwork skills, (2) skills - morale building, conflict resolution, information exchange, and (3) attitudes - collective orientation, belief in importance of teamwork. The above knowledges, skills, and attitudes are but a sample of those delieneated, those interested in the complete list are referred to Cannon-Bowers et al. (1995) or later work by Salas, Rosen, Burke, and Goodwin (2009). For a practitioner-oriented version of core processes and states which facilitate team performance the reader is referred to Salas, Shuffler, Thayer, Bedwell, and Lazarra (2015).

This work, in turn, can be leveraged for the building of an architecture to guide the user in identifying the most efficacious scenario set based on the competencies which need to be trained. For example, based on the above work, several questions which begin to describe the nature of the team within which the trainees will be embedded could be imagined:

- **Do you expect to be working within the same team across tasks or different teams?**

- **Do you expect to be working on the same type of tasks or different tasks?**

These questions begin to highlight the types of competencies that need to ultimately be trained. Therefore, they could then be used to link practitioners to sets of scenarios which target said competencies. In essence, not leaving scenario content to chance but using the science of teams to drive a diagnosis of front-end needs, which have a priori been tied to specific content within the program. However, in many cases there may be time and/or resource constraints so one could imagine a third-tier of questions that serves to delineate for the most efficacious set of competencies out of this larger set to begin targeting. This is where the team taxonomic efforts that have been conducted over the years become instrumental. Specifically, in providing a set of defining task, team, and contextual features that have been argued to drive a need for different types of interaction and corresponding competencies. Next, a few of the more prominent taxonomic efforts will be highlighted along with how they might be used to inform the development of a front-end architecture to assist practitioners in the selection of ITS content.

Team Type Taxonomies

In an effort to better understand teams, there have been a multitude of team taxonomies and classification systems developed throughout the years (e.g., McGrath, 1984; Hackman, 1990; Sundstrom et al., 2000; Devine, 2002). Taxonomies have been argued to serve several functions, including the parsimonious specification of the phenomena of interest, critical structural components and, as such, the range of generalization (Fleishman, Mumford, Zaccaro, Levin, Korotkin, & Hein, 1991). In the case of teams, taxonomic efforts have been broadly used to describe the types of teams that exist in organizational settings and corresponding structural features. A recent review of team taxonomic efforts, identified approximately seventeen team taxonomies that have appeared in the literature (Wildman et al., 2012). While there is no single agreed upon team taxonomy within the literature, the predominant number of team taxonomies are heavily influenced by the type of team and the nature of the tasks that are engaged in and, correspondingly, the underlying dimensions upon which these tasks vary. The lack of agreement on a single team taxonomy may, in part, be due to the complexity of teams.

Despite the multitude of team taxonomies and lack of an agreed upon gold standard, much can be leveraged from existing taxonomic efforts that can help to guide the building of a front-end architecture for team-based ITSs. Specifically, these taxonomies can provide insight into the types of defining features upon which teams are often classified. These defining features have been argued to be important based on them driving the need for different sets of teamwork competencies. Therefore, the defining features could potentially be used to build a set of primer questions within the front-end architecture of a team-based ITS to facilitate the identification of the types of teamwork competencies that should most appropriately be trained. Next, we briefly describe a few of the more prominent taxonomies and highlight their potential contributions in light of developing pieces of an architecture that could guide users to a better understanding of the types of competencies (and later corresponding scenarios) that may be most beneficial to train.

McGrath (1984) presents one of the earlier team taxonomic efforts whereby eight team types are distinguished based on the types of tasks typically performed. The eight team types (i.e., planning, creative, intellective, judgment, cognitive, mixed motive, contests, and psychomotor) are classified based on the degree to which they reflect: conflict-cooperation, conceptual-behavioral, and choose-execute. While this classification provides an initial leverage point, later taxonomies classify teams based on contextual characteristics. The contextual characteristics, or defining features, in turn are most often argued to have an impact on performance requirements and correspondingly the team competencies that might be most efficacious to train. Similar to McGrath (1984), in a primary focus on the nature of the team's task is Devine's (2002) work which delinates an integrative taxonomy of work teams consisting of fourteen team types which vary along the degree to which their tasks are intellectual (i.e., executive, command, negotiation, commissions, design, advisory) or physical (i.e., service, production, performance, medical, response, military, transportation, sports). However, Devine (2002) pushes beyond the level of task type to identify a set of contextual characteristics which cut across intellectual and physical team tasks and through which teams can be classified. Specifically, Devine (2002) argues that teams vary along the following defining features: fundamental work cycle, physical abilities, temporal duration, task structure, active resistance, hardware dependencies, and health risk. These contextual features have been, in turn, argued to impact the nature of coordination required and the types of competencies that might be most efficacious to train.

Also of note is the work of Sundstrom and colleagues (e.g., Sundstrom, DeMeuse, & Futrell, 1990; Sundstrom & Altman, 1989). Similar to earlier efforts, Sundstrom and Altman (1989) classified teams into types (i.e., advice/involvement, production/service, project/development, and action/negotiation). However, these authors highlight contextual features which have not explicitly been delineated by the previously mentioned taxonomic efforts, but would most likely fall within task structural features. Specifically, Sundstrom and Altman (1989) classified teams based on the degree of differentiation and external integration. Again, for the current purposes, we are less interested in the 'team type' than the contextual features which have been argued to be important in that they drive a need for different types of competencies. Differentiation refers to, "degree of specialization, independence, and autonomy of a work team in relation to other work units" (Sundstrom et al., 1990, p. 124). In contrast, external integration refers to the degree to which the team must coordinate and synchronize its actions with others outside the immediate team. This can refer to the integration of the team with other entities in the same organization, but outside the immediate team or entities outside the organization. The degree of differentiation and external integration, in turn, are argued to create different demands for effectiveness. This distinction is important as it begins to point to the degree to which team competencies associated with boundary spanning are important. In addition, for those teams which have low levels of differentiation it may also begin to highlight the importance of a subset of team membership being trained on competencies that are emerging out of the literature on teams-of-teams or multiteam systems (see Zaccaro, Marks, & DeChurch, 2012; Shuffler, Rico, & Salas, 2014). As emerging work in this area is beginning to highlight that when teams are working within a multiteam system it is not enough to be trained in team competencies, but additional competencies may be needed to foster performance within the larger system. In some cases what fosters effectiveness at the team level, may actually detract from coordinated interaction and corresponding effectiveness at the multiteam system level.

Most recently, Wildman and colleagues (2012) developed an integrative taxonomy, in part, to combat weaknesses in prior taxonomies which do not elucidate team-level properties which serve to distinguish different team types. Similar to earlier taxonomies, Wildman, Thayer, Rosen, Salas, Mathieu, and Rayne (2012) argue for an integrated set of task types: managing others, advising others, human service, negotiation, psychomotor action, defined problem solving, and ill-defined problem solving. However, going beyond prior work, this taxonomy focuses on team-level attributes which distinguish the various team types and the corresponding implications for performance. Specifically, Wildman et al. (2012) argue for the following set of team-level characteristics: task interdependence, role structure, leadership structure, communication structure, physical distribution, and team life span.

In looking at the team taxomic efforts that have been briefly described above, one can envision a series of prompts which could be incorporated into a front-end diagnostic tool and used to point those not experts in team science toward the types of competencies that may be needed to be covered within targeted teams. For example, (1) To what degree does the team need to rely on entities outside the immediate team?, (2) What is the expected life span and temporal duration of the team (e.g., ad-hoc, intact)? (3) What is the level of task interdependence?, (4) To what degree is the team distributed in time and/or space?, (5) What is the fundamental work cycle?, and (6) What is the task structure? Responses to the prompted questions serve to provide the system with a better understanding of the contextual, team, and task characteristics within which the targeted team(s) are most likely to be operating. The science of teams could then be further leveraged to link defining features as pulled from the taxonomic efforts to the specific competencies underlying the type of team interaction needed as argued for within the taxonomic efforts. Within this front-end architecture specific competencies or perhaps sets of competencies would then be linked to scenario sets from which practitioners and users could choose. It might also be possible to use this information to provide those who choose to use the team-based ITSs with a set of guidelines that would assist in creating new scenarios to tap into specific competencies once the initial front end mapping had been done.

Conclusions and Recommendations for Future Research

There is no denying that the design and delivery of team training is a complex endeavor, even when done outside the context of intelligent tutoring. The complexity increases when attempting to design ITSs for teams not only due to the multitude of moving parts, but also in part due to the fact that a team of experts does often not equal an expert team. In essence, ITSs that are seeking to target the training of teams must move beyond training those taskwork skills which have traditionally been the focus of ITSs, to training the more complex teamwork skills. In doing so, there are implications and challenges from the front end design of the architecture, to the building of content, to assessment and feedback on the back-end. Team training has a reputation for being resource intensive and it has traditionally been noted that it often takes a team to assess a team. The use of team-based ITS is not attempting to remove the instructor, but to reduce the instructor's burden. It can also serve to increase the systematic design of training by providing instructors with a set of tools that can be leveraged by individuals who have the domain content knowledge, but may not be experts in team dynamics.

While there are a number of challenges in designing ITSs for teams which have been covered elsewhere, there is also a large literature on the science of teams that can be leveraged. Herein, we have attempted to examine a small piece of that existing literature to highlight how it could begin to be used to build a high-level diagnostic capability into the front-end of the ITS architecture that could be used to more systematically map training needs to scenario design and content. Here we provide a high-level vision. To substantitate this vision, future work needs to further use the science to drill down into the questions such that meaningful responses (i.e., meaningful to those outside team science) to the prompted questions can be provided within the architecture. Additionally, while taxonomic work highlights the features argued to drive performance requirements, most do not explicitly systematically lay out how the different features drive the need for specific competencies. For some of the features highlighted that information exists at least, in part, based on meta-analytic efforts that have been conducted. In other areas it may point a need for future meta-analytic efforts or alternatively thematic analysis of the literature to map the specific linkages. In the cases where there are gaps, most notably on the knowledge or cognition side it points the need for additional targeted research. It is our hope that we have served to pique the interest of those in the ITS community and highlight that there is a literature that can help the development of ITSs for teams.

References

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco, CA: Jossey-Bass.

Devine, D. J. (2002). A review and integration of classification systems relevant to teams in organizations. *Group Dynamics: Theory, Research, and Practice*, 6(4), 291-310.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. *The Leadership Quarterly*, 2(4), 245-287.

Hackman, J. R. (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork*. San Francisco, CA: Jossey-Bass

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

Morgan, B. B. Jr., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *The Journal of General Psychology*, 120(3), 277-291.

Saavedra, R., Earley, P. C., & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology*, 78(1), 61-72.

Salas, E., & Burke, C. S. (2002). Simulation for training is effective when… *Quality Safety in Health Care*, 11, 119-120.

Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2009). The wisdom of collectives in organizations: An update of the teamwork competencies. In E. Salas, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pgs. 39-79). New York, NY: Psychology Press.

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human Resource Management*, 54(4), 599-622.

Shuffler, M. L., Rico, R., & Salas, E. (2014) (Eds.). Pushing the boundaries: Multiteam systems in research and practice. *Research in Managing Groups and Teams*, Vol. 16, UK: Emerald Group Publishing.

Sundstrom, E. & Altman, I. (1989). Physical environments and work group effectiveness. *Research in Organizational Behavior*, 11, 175-209.

Sundstrom, E., DeMeuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, 45(2), 120-133.

Wildman, J. L., Thayer, A. L., Rosen, M. A., Salas, E., Mathieu, J. E., & Rayne, S. R. (2012). Task types and team-level attributes: Synthesis of team classification literature. *Human Resource Development Review*, 11(1), 97-129.

Zaccaro, S. J., Marks, M. A., DeChurch, L. A. (2012) (Eds.). *Multiteam systems: An organization form for dynamic and complex environments*. New York, NY: Taylor and Francis.

Chapter 20 – Lessons Learned Creating a Team Tutoring Architecture: Design Reconsiderations

Keith Brawner[1], Anne M. Sinatra[1], Stephen Gilbert[2]
[1]Army Research Laboratory, [2]Iowa State University

Introduction

Intelligent Tutoring Systems have been constructed for many reasons by numerous groups of researchers and practitioners. The Generalized Intelligent Framework for Tutoring (GIFT) was created in part to unify disparate research into a commonly shared research output. Unlike individual tutoring, however, few groups have constructed an Intelligent Team Tutoring System (ITTS) (Robert A Sottilare, 2018; R.A. Sottilare, Holden, Brawner, & Goldberg, 2011). While GIFT was created in order to address the problems of individual tutoring systems, small group (i.e., squad level) team tutoring is one of its goals, and it was modified to support the instruction of teams of individuals. These modifications were consistent with the original GIFT design goals for individual tutoring systems (individual assessment/feedback files), but this ultimately proved cumbersome for the construction of a reusable set of modules and processes for teams. This paper reports on the experience of constructing an ITTS software architecture using GIFT, but expands on the challenges encountered during the design process, and what could be done differently during redesign. Perhaps more important than discussion of the original team tutoring design or an improved team tutoring design is a discussion of the reasoning behind using certain principles and implementations. As among the first to try to construct a reusable team tutoring structure, the lessons learned should be considered as a basis for future implementations.

Types of Team Tutoring

Team tutoring is a complicated process, and requires updating existing ITS software architecture to support a team (Gilbert, Dorneich, Walton & Winer, 2018). However, the updating of the architecture is more than just duplicating what has been done for an individual multiple times. Careful consideration needs to be given to how the team will be tutored, how the team's performance will be assessed, and the type of task that the team will be tutored on. In the case of GIFT, since it is a domain-independent ITS architecture, updating it to support teams is particularly difficult. The framework needs to be able to support multiple types of domains/content areas, multiple configurations of teams, as well as multiple types of team performance assessments. In addition to this, GIFT must be able to support tutors that are focused on teamwork, taskwork, and collaborative learning.

Further, team tutoring must support different team task configurations, and the different team pedagogies that might be used. Teamwork and taskwork tutoring differ from each other, as teamwork focuses on the general overarching principles of working together with a team, and taskwork is specific to the team task that is being engaged (Salas, 2015). In order to engage in teamwork tutoring, there should be a representation of generalized positive team behaviors that need to occur to reach a goal, as well as methods of measuring those desired team behaviors. To engage in taskwork training, however, the focus is more on the actual performance of the task itself and the role that each of those team members play in it; performance on the task is usually more easily measured. The structure and roles within the team are going to vary greatly based on what the overall task is that is being tutored. Additionally, some team tutoring tasks including collaborative problem solving or collaborative learning may have a different learner configurations. Collaborative learning theory (Roschelle, Suthers, & Grover, 2014) characterizes teams as a community in which some team members might be core members, while other members are more peripheral. Thus, a key element of monitoring team dynamics is measuring the degree to which each team member is involved in the community, e.g., team members' motivation for collaborating, the level of joint attention to the task, and the degree to which the team members reflect on the teamwork process. From the point of view of ITTS architecture, this theory creates a software requirement that the ITTS architecture support monitoring of each team member individually, as well as each member's interactions with teammates.

GIFT

GIFT is made up of a series of interchangeable modules which are linked together by a message bus (Robert A Sottilare, Brawner, Sinatra, & Johnston, 2017). The core Modules of GIFT are the Domain, Learner, Pedagogical, Gateway, and Sensor. However, GIFT is an architecture, and on top of this module foundation is built a number of less standardized components – items such as authoring tools (i.e., the GIFT Authoring Tool), dashboards (i.e., the learner panel), and post-processing tools (i.e., the Event Reporting Tool). This ecosystem also includes less commonly used, or more commodity-manufactured, components such as a user database (i.e., User Management System), content repository (i.e., Nuxeo), and software libraries for external applications (i.e., XML Remote Procedure Calls).

GIFT is relatively simple, containing a few key modules, which now seek standardization through the IEEE Adaptive Instructional Systems group. Each of the core modules relates to the learner, e.g., a model of the learner's state, content to teach the learner, or expert information needed to evaluate the learner. It is tempting to design a "Team GIFT" to be the same thing – a model of the team's state, content to teach the team, and expert information needed to evaluate the team. As GIFT was built for an individual learner and then built out as a collection of differing tools in order to support the learner – the hope is that many of the tools can be re-used and not reinvented for team tutoring purposes. However, as has been reported in the literature, there are challenges involved with scaling up intelligent tutoring systems for teams (Bonner, Gilbert, et al., 2016; Bonner, Slavina, et al., 2016; S. Gilbert et al., 2015; S. B. Gilbert et al., 2017). Since a team's performance depends strongly on the team skills of the members', not only on the individuals' task skills, the complexity of the ITTS's model of the both the learners and the state of the training scenario grows exponentially. Additionally, there is the challenge of keeping GIFT domain-independent, therefore, any tools that are developed for authoring in team tutoring in GIFT need to support multiple types of team configurations.

Team GIFT

An initial team surveillance task tutor was developed using GIFT, and has been detailed in the literature (S. B. Gilbert et al., 2017). One of the main goals in the development of the tutor was demonstrating that GIFT could be adapted for use by multiple players in a shared environment. The first version of the tutor consisted of two players working together to monitor their own portions of an environment in a scenario in the computer-based software Virtual Battlespace 2. Challenges that needed to be overcome were how to assess the team members as individuals and as a team, and how to have GIFT assess their performance as well as provide feedback simultaneously. The goals of this initial tutor creation included creating a shared terminology; creating a militarily-relevant team task; and establishing an approach that can facilitate the development of team tutors. The challenges were overcome primarily by adapting the existing feedback and assessment structure (Domain Knowledge Files) to be able to track performance by both individuals and by a team. Primarily this was done by having a separate DKF for each individual player, and an additional DKF that assessed the team tasks and could provide feedback to both players if needed. The synchronization challenge was solved in this particular instance by utilizing a "lobby" in which both players checked in prior to the beginning of the tutoring session, as is used in many multiplayer commercial games. This lobby method is generalizable, as shown by the multitudes of commercial games, but the approach may have to differ based on the training applications in use with GIFT.

The initial team surveillance task was used to create two tutors. The first of these was the Surveillance Tutor, which had two team members with the same role. The second tutor was called "Surveillance Tutor with Sniper," which was a three player version which involved creating an additional DKF file for the new team role and their interactions with other team members. These two tutors represent a first attempt to engage in team tutoring in the GIFT framework. They can be thought of "successful" as they were operational enough to accomplish the task of team tutoring while being designed flexible enough to capture new tasks, domains, and team structures. However, many of the approaches that were used in the creation of the two tutors were tightly coupled with the task that existed, and the addition of tasks lacked in both generalizability and scalability. The simple addition of a new team member involved the addition of a role, its assessment, its communication, and the assessment of the communication; further additions would face exponential complexity. While the initial two ITTS met their goals, but it is now time to consider how the approach could be improved; are there other approaches that could achieve the same goals but have additional scalability?

First Pass Design and Walkthrough

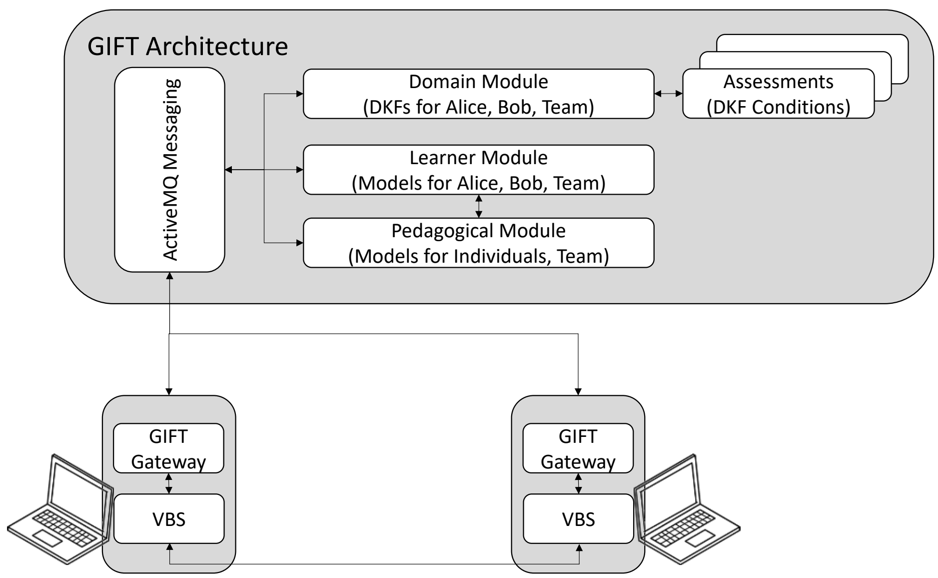


Figure 3. Initial Team Tutoring System Design (S. B. Gilbert et al., 2017)

As part of the initial GIFT ITTS approach, the team decided that the team tutoring system and architecture would be built on top of the existing GIFT structure. The tutoring for a team consisted of the tutoring for each team member in the same manner as an individual tutor (based on an individual DKF assessment), and that the team members would additionally receive tutoring as part of team tutoring components (based on a team DKF assessment). See Figure 1 for a visual representation of the approach used in the initial GIFT team tutor. This decision to treat "the team" in a similar way that GIFT would treat an individual had ramifications for the core GIFT modules, but also for the less standardized components and for the commonly-used components. The changes to individuals' modules in order to support this decision are discussed next, on a module basis. For an individual application, a single Gateway Module was needed to link the external simulation to the rest of GIFT. For an individual learner, the Gateway Module translated simulation messages into GIFT messages and sent them to the Domain Module for assessment. For a team of learners, the Gateway Module remains unchanged, except that messages for each player were appropriately routed to the individual Domain Modules for the individual learners, and messages regarding each player should also be copied to a Team Domain Module (or, more accurately, the existing Domain Module with a team configuration). As mentioned above, the need to have multiple players within the same simulation also imposed the requirement to have a shared start time, a “game lobby” functionality and other general synchronization logic.

As noted above, for an individual learner, the Domain Module has a configuration called a Domain Knowledge File (DKF). The DKF contains the information about how the learner will be assessed in the external application, and the domain specific feedback that they are to receive; it has links to content, assessments, and feedback which are controlled by pedagogical direction. For a team of learners, the GIFT architecture was tweaked to allow the existence of multiple simultaneous DKFs. Thus, each learner had an individual DFK, and a team DKF was added containing similar information. The team DKF processed *team* performance, *team* assessment, and *team* feedback (for both teamwork and taskwork). For a team with three team members, one can imagine six DKFs: one for each team member, one for the team of three, and one for each pair of team members so that pair communications can be monitored. This architecture served the purpose of monitoring both individual and team performance, but it leads to the challenge that each DKF functions like an independent tutor that does not know the actions of its tutoring colleagues. See below for more detail on the implications.

For an individual learner, the Learner Module gathered, stored, and reported information about the learner. In individuals, these attributes are items such as personality type, learning preference, as well as information about performance, learning, and affect. Following the logic described above, to extend this for teams, the Team Learner Module (Learning Module with a "team" configuration coming from the Domain Module), would gather, store, and report information about the team. Relevant team specific learner module components would include team performance and team-level assessments of state, such as measure of communication or trust. However, in the two surveillance ITTSs, this team Learner Module did not add any specific value, as no domain-general team items were programmed to be observed. This design decision required results in the Domain Module to drive team instructional decisions. The result of this is that it is not easy to reuse this approach between domains, as many assessments rely on the content of the messages that team members communicate to each other. In theory, however, the configuration of team assessments, such as frequency of communication acts, may be transferrable between domains.

For an individual learner, the Pedagogical Module controls the instruction, including actions such as ordering content, requesting assessment, changing scenario difficulty, providing hints and feedback, and other instructional actions. Following the logic above, for a team, a team pedagogical module would need to make the important decision of when to give feedback to specific team members and when to give feedback to the entire team. In the surveillance ITTSs, the normal pedagogical module was fed information tagged with "team" simply passed the feedback onwards. Future implementations should allow for team-specific models, modeling of differing team members, and other pedagogically-specific items.

Also, a key requirement for tutoring teams is monitoring the overall amount of feedback received by team members. With the multiple-DKF architecture mentioned above, it is easy for learners to be inundated with too much feedback with messages arriving from their individual DKF, the team DKF, and any sub-team DKFs that might exist. As a specific example, a learner who does not report an observation of an enemy ship in a timely manner may be graded poorly on their individual observation performance, graded poorly as a team on a communication metric, and graded poorly as a team for reporting observations – resulting in three pieces of feedback individually and two pieces of feedback for the team. A better team tutoring system would handle individual and team feedback differently, such as providing individual task feedback during the event and saving team feedback for an After Action Review (AAR). Ideally the team pedagogical module would filter feedback messages from the multiple DKFs, correctly model a chain of command, or other more complex items.

In the surveillance tutors, an additional Feedback Filter component was created. Normally, GIFT offers feedback to the learner after every action that changes the learner's overall assessment state (Above, At, or Below Expectation). In tasks in which every learner action is noteworthy pedagogically, e.g., where learner actions consist of answering questions or completing tasks, this approach makes sense. But if the granularity of the learners' tasks is smaller, such that the overall pattern of learner actions is of pedagogical interest, but not the individual actions, feedback need not be given for each action. The surveillance tutors fit the latter category, with learners performing many repetitive small actions. Thus, the Feedback Filter allowed only every *nth* feedback message to reach the learners, where *n* depended on the subtask.

Design Failures

One of the design goals of GIFT is to provide a pedagogically sound approach to authoring a tutor that does not require the subject matter expert to understand the elements that make up a tutoring system. The tutor author brings his or her content to the system, and the system is primarily responsible for configuring the tutor. A part of this design trade-off is that the *individual* components of the tutoring system can be *more* complex, while reducing complexity overall. As a specific example, consider the creation of a domain-dependent model of content applicable to multiple new and unknown application tasks (e.g., modeling how to dress wounds in multiple application contexts such as in field, in the home, and in the workplace); this task is more complex than modeling the domain specifically for one application (e.g., how to dress wounds just at home), but did not require the additional creation of learner modeling techniques, instructional techniques, feedback delivery mechanisms, and other tutoring components. Although the individual model of the domain was more complex, the overall complexity required to make a tutoring system decreased. This approach, while adding work for an individual component (the wound dressing model), gives an overall reduction to system authoring cost as applications are expanded.

The architecture of the Feedback Filter, while a good first attempt, presented several problems. First, it was based on the assumption that feedback about individual performance and team performance were of equal weight. In reality, the nature of the task and coaching pedagogy would determine this weighting. For a small team with tightly interdependent tasks, the individual feedback may be critical for the overall team. For a team of people who act more independently, but share a common overall purpose, individual feedback may play a stronger role. Also, the approach of simply "passing along feedback" belies the careful strategic approach to giving feedback taken by a professional team coach. For the individual, every feedback statement is a criticism, even if taken constructively. While sometimes the bitter pill of receiving feedback can be mollified by a framing of "It's for the good of the team," that approach can be used only a limited number of times. Thus, team feedback statements and individual feedback statements are not equivalent in impact on the learner, and should not be counted as such. Finally, different learners have different tolerances for receiving feedback. Ideally, this filter would be personalized for the team member based on a profile of whether the learner benefits from more or less feedback (ex. frequent feedback for novices, infrequent feedback for experts)..

This complexity is compounded through the interactions of the various components. The original GIFT architecture called for a single sequential round trip from the learner to the training environment to the Gateway Module to the Domain Module to the Learner Module to the Pedagogical Module and then back to the Domain Module, to the Gateway Module, to the training environment and to the learner. This represents a single loop of human-system interactions. The above design calls for a *parallel* linear interaction (adding “Team” to each Module), which is tied back into the training environment for *multiple* learners. The human-system interaction is then a product of *two* loops that interact: one for individual tutoring and one for team tutoring. This makes testing, debugging, and tracking activities more complex, even without considering the challenges of time-delayed feedback, the construction of after-action review reports, or real-time feedback.

One approach to handling this is to create a broader, more generalized version of the Feedback Filter that one might call the Learning Experience (LX) Module (Cushard, 2012). Just as companies seek to optimize the customer experience they offer, the LX Module would maintain awareness of each team member's experience in the simulation: their initial learning profiles, how successful they are, their cognitive load, the amount of feedback they've received, how their performance is changing, how much impact they have on other team members, etc. – all the factors that a quality professional coach takes into consideration. This "executive" module of GIFT would be based on LX models of different team tasks and different team structures, so that it could be re-usable from team to team and task to task of the same genre.

GIFT Design Principles

GIFT was initially designed to solve complex but definitive problems. The first of these is the problem of lack of standardization, causing increased initial construction cost of the ITS. The second is the difficulty of modifying an existing system to suit new purposes. These problems were addressed through the separation of logic and the creation of a common infrastructure. The goals for team tutoring *include* the goals for individual tutoring with the added goal of being able to provide tutoring to multiple individual team members and to the team as a whole.

Separation of Logic

Meeting the two goals discussed above starts with the separation of logic. Separation of logic, or dividing the system architecture into maximally separate independent modules, so that each could be updated separately for new training situations, is the primary design goal. This approach accommodates shared authoring tools, interchangeable instructional domains, keeping experimental control for instructional modifications, and other items. Separating the logic into modules additionally allows for the topic area (domain) of a tutor to be changed to another topic (domain) area with relatively minor edits. Similarly, it allows for the improvement of pedagogy across *all* topic areas through the replacement of a pedagogical module with an improved pedagogical module without further change.

For a team, this issue matters – team assessment is not likely to be held constant across multiple team structures. A good basketball team and a good aircraft maintenance team have both a) different jobs (score points, fix planes), b) different measures of what good taskwork is (pass ball and create openings, delegate out work), and c) different models of what good teamwork is (communicate frequently, leave people alone to work). It seems possible that a model of team instruction may not be held constant over multiple domains, though it could be that, if well designed, a relatively small set of team models could satisfy many domains (e.g., the Pareto 80/20 principal).

Another approach to consider is a parent-child inheritance hierarchy within modules. For example, some pedagogical principles, like "If the learner is getting frustrated, try to offer him/her some small successes," might generalize across domains. Other pedagogical principles might be specific to the domain. For example, if learning to write complex Excel formulas with multiple nested IF conditions, a principle might be, "If the learner can't debug the complex Excel formula, suggest breaking it up across multiple spreadsheet cells and building it up piece by piece." This more specific pedagogical principle could be considered to be a child class instance of the parent class principle of offering small successes. Thus, a new ITS or ITTS could be authored with a generic template of parent class principles, saving significant time, and then add more specific child instances as necessary. This same hierarchical inheritance approach could apply to domain models and feedback instances as well.

Figure 4 illustrates an example of an ITTS architecture that would use this inheritance approach to guide members of a medical team: a doctor and two nurses. The tutor contains domain models for general medical task work, along with more specific models of doctor knowledge and nurse knowledge. Also, it contains a domain model for teamwork generally, and a more specific instance of that model focused on teamwork on medical teams. The domain models would like generate many possible feedback messages for any given configuration of world states and learner actions, and pass those messages to Pedagogical Prioritization for winnowing. The Pedagogical Prioritization module contains a model of the most general pedagogy, as well as more specific models for team pedagogy, and even more specifically, medical team pedagogy. Also, it contains pedagogical techniques designed specifically for doctors and nurses. This module would prioritize feedback messages based on pedagogical guidelines and pass the surviving options to the Feedback Filter or LX Module, which contains models of each of the learners, as well as a team model and sub-team models. This overall module is responsible for making sure the learners are not overwhelmed by feedback, and would maintain a history of their learning experiences so far so that it can optimize each person's learning. This module finally chooses among the multiple feedback messages passed to it and sends them along to the learners.

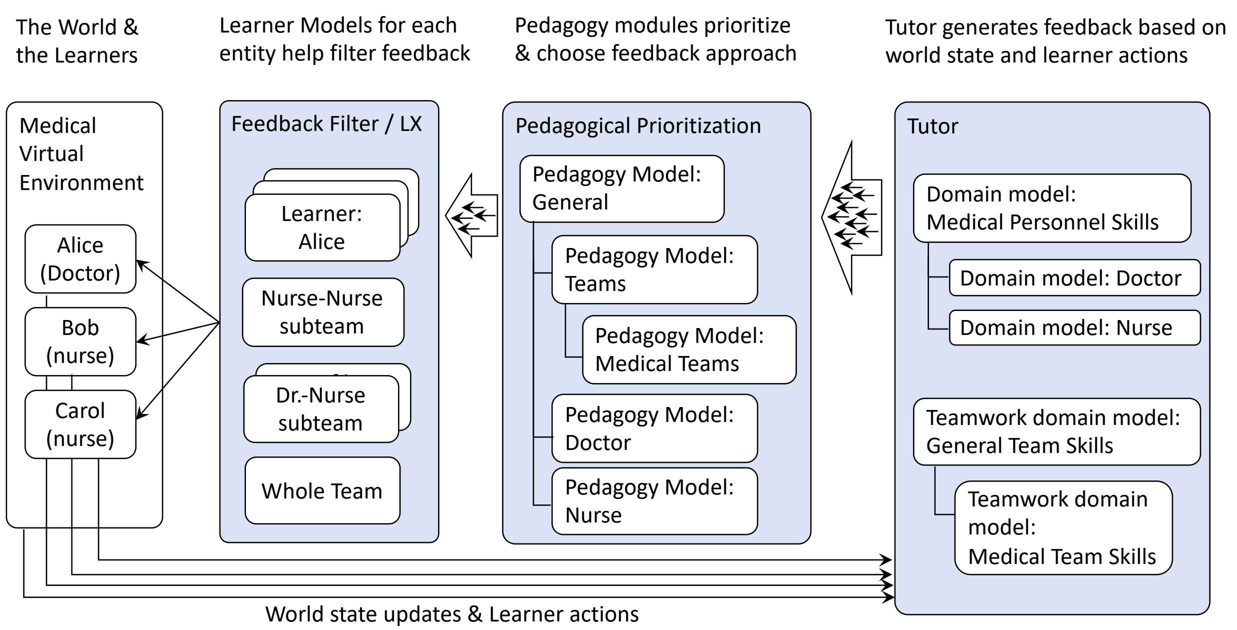


Figure 4: Example of an ITTS architecture featuring hierarchical inheritance models. A doctor and two nurses are the team.

Common infrastructure and Pipeline

In the same manner as GIFT for individual tutoring is built upon interchangeable modules for the purpose of both production and experimentation, a team tutoring architecture should mirror this functionality. Achieving the first goal of logic and process separation should allow for the achieving of a second goal of a common infrastructure. Advantages of construction in this manner is the ability to share a common set of authoring tools, a data processing pipeline, and standardization of requirements for construction. While these are not explicit architectural principles, they are a desired product of the other principles.

Individual *and* Team Tutoring

For individuals, tutoring is synonymous with individualized feedback and path planning through content. For teams, however, teaming items are divided among both taskwork and teamwork. Taskwork is the tasks that have to be accomplished (it is domain dependent). Teamwork is how the team is performing with respect to itself (it is mostly domain independent). These items are somewhat separate, as a team can be internally consistent and work well together but still fail to accomplish the task. As a concrete example, consider a losing World Cup team; the team is clearly very skilled and works well together, but in this specific instance did not accomplish their task. Similarly, it is possible to have poor team performance but accomplish mission or task objectives. As a concrete example, consider a middle school soccer team with a single exceptional athlete. The team may perform poorly as a team, but still win because of the individual athlete's performance. Tutoring a team to perform well (taskwork) and to be a functional unit (teamwork) are two *separate* items, but a flexible domain-independent team tutoring architecture should strive to accommodate both goals.

Team Design Examples – Training Different Types of Teams

The structure of teams can vary significantly. Teams might have members with different roles, the same role, or a mix. Teams may have a central hierarchical leader, shared leadership, or no explicit leader. Team members may be co-located or distributed. These are just a few of the different characteristics of team structure described in (Bonner et al., 2015). Also, team tasks themselves can vary significantly in interdependence across team members, independent subtasks performance, and other criteria. In the following section we describe two different team tasks with different configurations of roles.

These two examples highlight the differences that can exist in the roles, responsibilities and tasks of team members in different domains. A generalized ITTS framework needs to be flexible enough to allow for the authoring of tutors in both of these areas as well as other needed configurations.

Anti-Submarine Warfare (ASW) Helicopter Team

Anti-submarine warfare (ASW) helicopter team members have distinct roles with separate tasks. Physically, using active sonar reveals the position of the sonar system to those within range. An alternative to the use of surface ship mounted active sonar, which effectively announces ship position to enemies, is to use helicopter-mounted "dipping" active sonar. As it is technically difficult to hit a helicopter with an underwater-launched weapon without revealing the submarine's position, this approach to active sonar is preferred in many contexts. The goal of an ASW Helicopter team is to find hidden submarines from the (relative) safety of an airborne platform. The ASW Helicopter team has three key members – a pilot, a copilot, and a sonarman, with the primary tasks of flying the helicopter, coordinating intelligence with the host ship, and operating the sonar to find submarines, respectively. Theoretically there is no overlap between the tasks in each of these roles, while cross-training is common in practice.

Combat Outpost Guard Team

A combat outpost guard team has a very different configuration than an ASW helicopter team. The primary goal of a combat outpost is to extend the depth of the security area, or provide safety to forward observing positions who may be encircled by enemy forces. There can be many members of a stationed combat outpost guard, but a minimum of two posted guards is common. The mission of the posted guards is to observe and report on relevant activities they observe outside the secure area until ordered to withdraw, advance, or some other order change. Functionally, each member of the team has the same mission during this time period.

Team Architecture Design Goals

Design Goals for Reuse

GIFT, as originally designed for individuals, can be reused to provide tutoring for teams, provided that the tutoring system adequately represents each team member's task role. As a design approach, this means that the individual tutoring for the combat outpost guard team can be identical without any significant change – the system can apply the assessments written for the individual for each team member. Reuse of individual tutoring is wholesale. Similarly, changing the individual tutoring from tutoring two individuals to tutoring ten individuals playing similar roles involves little, if any, changes. Similarly, the performance of the team can be a more simplistic roll-up of individual performance metrics; the security of the outpost is mostly a summation of the security at each point. The assessment configuration of teamwork performance (e.g., communication, coordination, etc. (R. Sottilare et al., 2017)), however, may be able to be reused from another domain which is heavily communication-oriented.

The ASW team, however, has different roles. Each role has very different responsibilities – piloting aircraft; tactics and procedure; operation of underwater acoustic sensors. The individual tutoring that each of these learners receives is significantly different; there is little opportunity for reuse of domain content. GIFT, however, may reuse all of the other components of tutoring – Learner Module, Pedagogical Module, Gateway, etc. The team performance metrics cannot be a simplistic roll-up of individual performances; there is not an additive relationship between 20 seconds of smooth hovering and spotting an enemy submarine. Alternatively, the performance of the team, or teamwork measures, such as communication and coordination (Sottilare, 2017), may be able to be reused from a training environment similar to the above example guard team task.

The intention behind a team tutoring framework is that, at a minimum, the teamwork tutoring component can be reused across instructional domains. Further, that the generated team tutor can be expanded to accommodate additional team members and roles easily. Additionally, similar to how GIFT does not change the Learner/Pedagogical/etc. Modules in order to instruct a new domain, the team tutoring system should be able to perform these actions on the team level – keep a history of the team, have a consistent instructional model, etc. This design goal for a team tutoring architecture to support the training of many teams has two derived requirements: for the team component to be agnostic to both the domain of interest and the structure of the team.

Agnostic to the Domain

Firstly, the structure to support team modeling and instruction must be agnostic to the domain of instruction. If correctly designed, this is a direct product of modularity. The ability to shift a team tutor from one type of team and task to another is critical to the creation of a team tutor; what good is a team tutoring architecture if it can only tutor ASW Helo teams? One way to think about team taskwork is to think of it as domain-dependent teamwork.

Agnostic to Team Structure

It is possible that different structures of teams should be instructed differently. The authors, however, hypothesize that there are basic models of instruction or principles of teamwork that are applicable regardless of team structure. The hope for such models is that they can be applied as baseline models for many types of teams, assuring that all teams get at least some instruction. An example of a simplistic model of team instruction, similar to the initial simplistic models for individual tutoring GIFT, is "feedback on underperformance given immediately upon observation." Another simplistic model of team pedagogy is a "speak the same language" model which monitors for vocabulary overlap and gives corrective feedback when vocabulary overlap is not observed. Both of these models are applicable regardless of team size and structure, albeit simplistic. More complex models can be constructed with similar data, as enabled by architectural design.

Design Tradeoffs Considered for Taskwork and Teamwork

Design Tradeoff Evaluation

The design tradeoffs will be considered with their ability to provide for the following list of desired features, discussed above:

1. Separation of logic among modules
2. Ability to provide common infrastructure, such as for authoring tools and database structure
3. Ability to tutor individually
4. Ability to tutor team taskwork
5. Ability to tutor teamwork
6. Ability to accommodate differing team structures
7. Ability to accommodate differing team assignments
8. Ability to be authored

Option 1 – Do Nothing: The Crowd of Tutors.

GIFT is capable of team tutoring by default through the use of multiple simultaneous domain models: one for each player, one for the team, and others for subteams, if necessary. This represents the failures of the first design, discussed earlier. This approach provides good scores on the separation of logic and common infrastructure. Upon use of this design, the team is instructed as though it is an individual learner – with content mapped to its joint traits, scenarios assigned in an escalating difficulty mastery-learning paradigm, and feedback given to all team members. A direct result is that no individualized feedback is available from the system (except on individual tasks created in the individual tutors), but that team taskwork performance is the primary means of adaptation. No team-specific modeling pedagogy is assigned, although authoring considerations could provide teamwork items as authored domain-dependently within the Domain Module. The structure does accommodate for differing team structures and roles. One of the primary disadvantages of this approach is that each of the learner components acts essentially as a separate, independent tutor. Thus, each team member on a three-person team might get feedback from his or her individual tutor and the team tutor, without those two tutors knowing what each other's feedback was. If the three-person task involved important pair-wise relationships, then it might require a learner module for each individual (A, B, and C), a learner module for each pair of individuals (A-B, B-C, A-C) and a learner module for the team (A-B-C). This "crowd" of six independent tutors could lead to significant feedback overload and even mixed messages that conflict in their direction. However, much like Selfridge's classic Pandemonium model of perception (Selfridge, 1958), in which we recognize objects by having a crowd of neural "demons" each simultaneously shouting about the features they see, and listening to the demons that shout loudest, a Crowd of Tutors approach could have a filter that quiets all but the "loudest" tutor feedback.

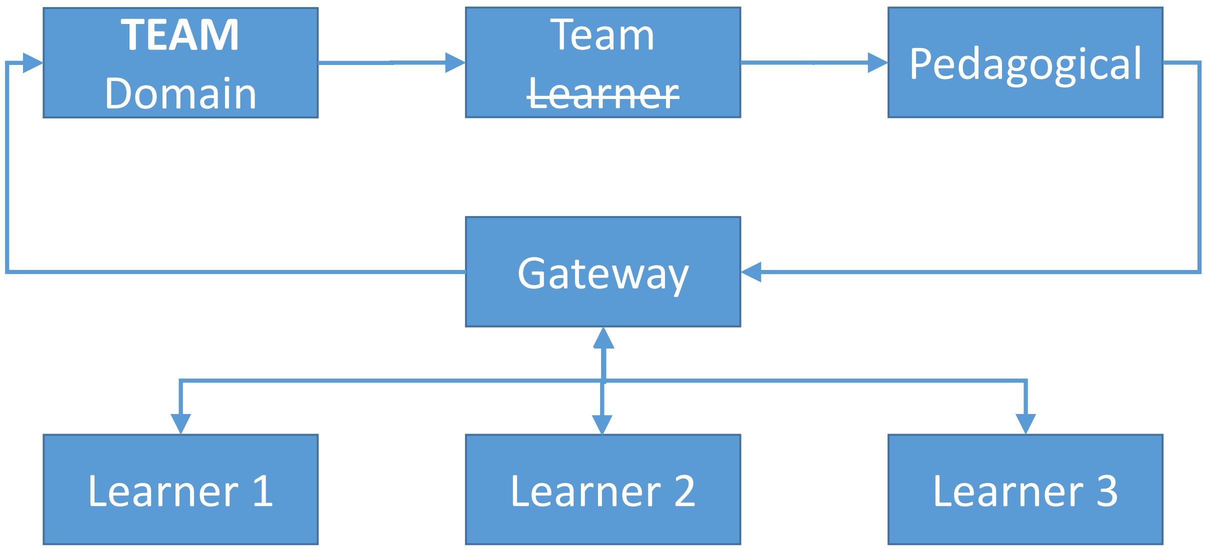


Figure 5. Do Nothing option for ITTS design

Option 2 – Individual Tutors for Individuals, Information Reported to Single Team Tutor Information (Initial Design).

The drawbacks of this type of design have been more thoroughly discussed in the earlier sections. Suffice to say that the design allows for individual interchangeability and significant flexibility, but that it does so at the cost of the maintenance of many modules. While the system is *capable* of tutoring the individual team members on taskwork and teamwork, the creation of the logic to assess, give feedback, and post-process imposes significant burden on the author. The author must answer the questions of what each team member should be performing with respect to his or her individual task, with respect to the overall team task (taskwork), and with respect to being a member of the team (teamwork). This is more appropriate for teams such as the ASW team where roles are sufficiently different with little taskwork overlap.

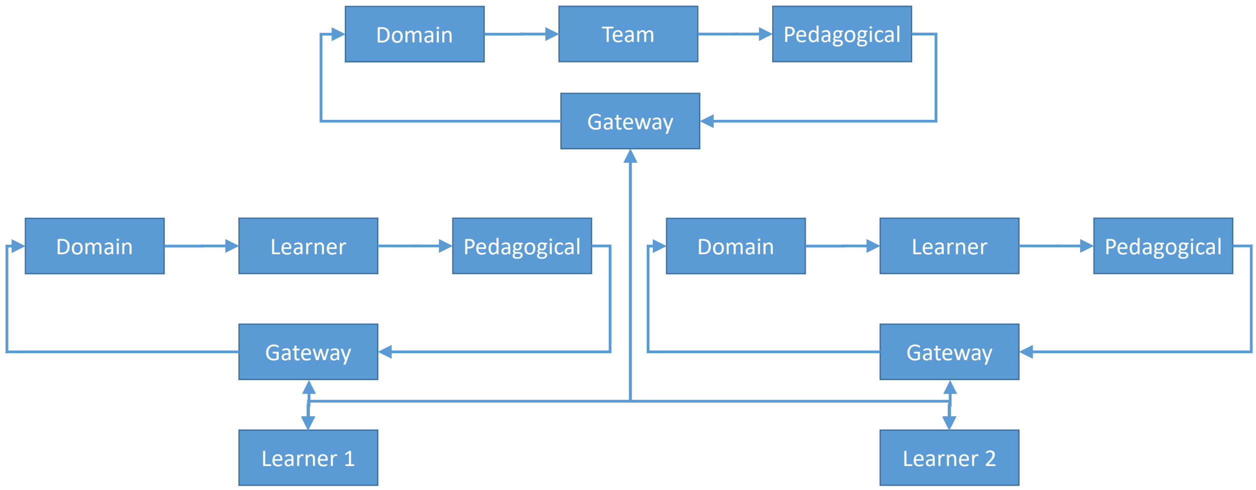


Figure 6. Initial Design for ITTS

Option 3 – "Team" Modules for Team Tasks

In an effort to moderate the complications of authoring team and individual tasks, and to standardize team taskwork and teamwork items, a structure of "team" Domain items, team modeling and team pedagogy can be considered. This mitigates many of the problems faced in the initial design. Further, it allows the common tools for team items without being particularly concerned about the individual domains. Authoring tools may be able to be templated for authoring team items, asking streamlined questions for individual domains such as "how often should team members communicate?" and "what should elements of a shared language or mental model be?".

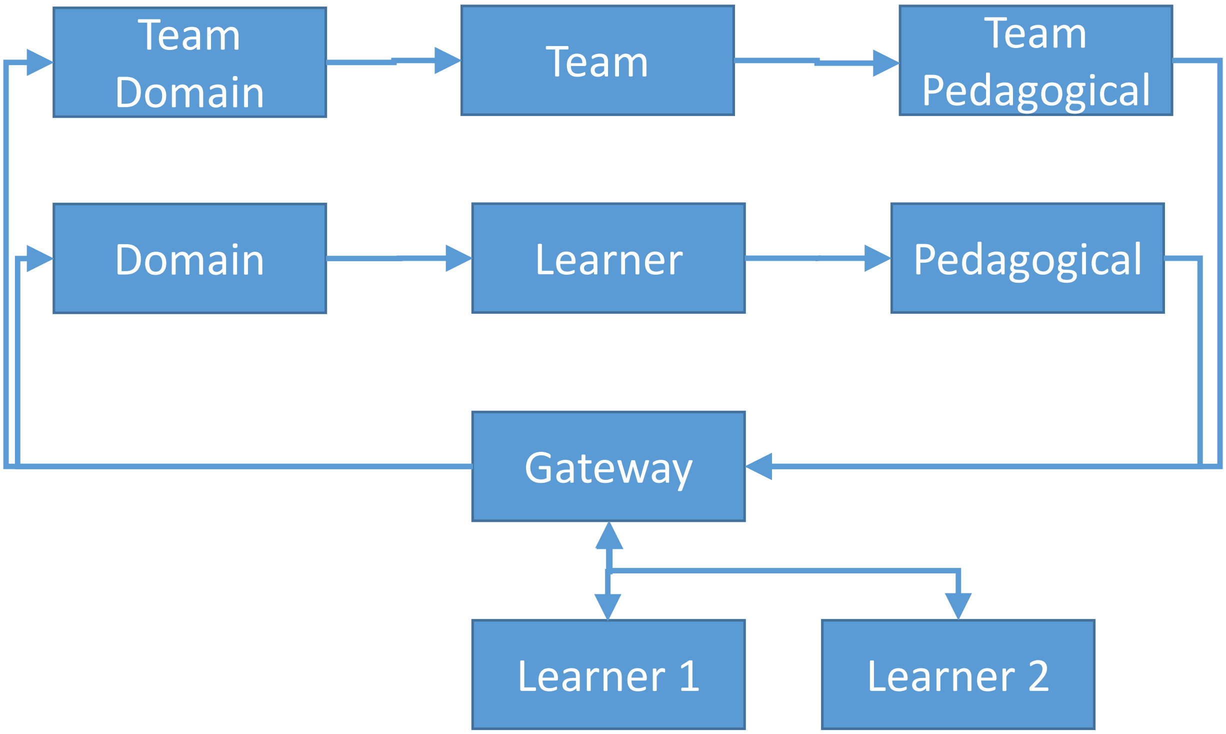


Figure 7. Model of Team Tasking overlaid to individual tutoring

Option 4 – "Role" Modules for Tasks, Team Roll-up

In an effort to moderate the complications of authoring team and individual tasks, and to standardize team taskwork and teamwork items, a structure of "team" Domain items, team modeling and team pedagogy can be considered. This mitigates many of the problems faced in the initial design. Further, it allows the common tools for team items without being particularly concerned about the individual domains. Authoring tools may be able to be templated for authoring team items, asking streamlined questions for individual domains such as "how often should team members communicate?" and "what should elements of a shared language or mental model be?". Further, it allows for team member to switch roles and not be penalized for accomplishing team goals; as a concrete example if the ASW Pilot and Co-Pilot switch seats but the total team mission is still accomplished. This type of design allows for an ease of authoring, as individual tasks can be mapped to one/more roles and evaluated at the "world" level for their accomplishment. Further, it allows for the portability of team pedagogy and models, abstracted from the domain.

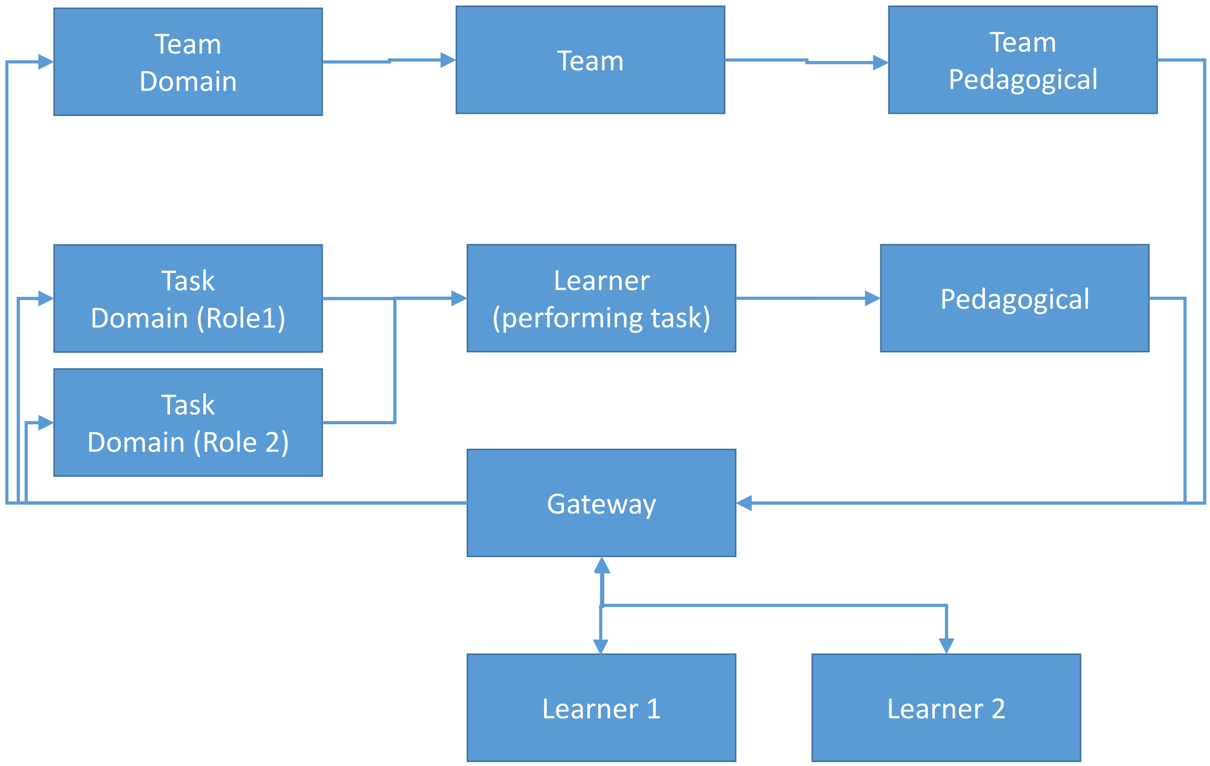


Figure 8. "Role" Based Model of ITTS

Design Recommendations for a Team Tutoring System

Table 1. Table of Design Tradeoffs

Individual Tutoring	Poor	Good	Good	Good
Team Taskwork	Good*	Good	Good	Good
Team Teamwork	Poor*	Good	Good	Good
Differing Team Structure	Fair	Good	Good	Good
Differing Team Roles	Fair	Good	Good	Good
Ability to be Authored	Poor	Fair	Fair	Good
	*Dependent upon authoring			
	**Many modules, difficulty in commodizing			

As we have described in this chapter creating a team tutoring framework is a difficult challenge. Creating a team tutor is difficult, but trying to account for all types of configurations of teams and retain flexibility makes it harder when designing an ITTS framework; achieving generalizability and scalability while not compromising to individual items is fundamentally difficult. We have discussed different possible approaches of creating an architecture that can support multiple types of team tutoring and different domains. At the time of writing, it seems as though the creation of individual roles, logic for mapping individual team members to roles, assigning traditional tutoring over roles, and transferring team instructional models between domains as domain-general team training provides the greatest level of scalability and generalizability. This is what this paper recommends as a path moving forward, although implementation details such as "team member to role disambiguation" may prove more difficult than initially assessed.

References

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A.,…Holub, J. (2016a). The Challenges of Building Intelligent Tutoring Systems for Teams. Paper presented at the Human Factors & Ergonomics Society (HFES) Annual Meeting, Washington, D.C.

Bonner, D., Slavina, A., MacAllister, A., Holub, J., Gilbert, S., Sinatra, A. M., … Winer, E. (2016b). The Hidden Challenges of Team Tutor Development. In R. Sottilare & S. Ososky (Eds.), *Proceedings of 4th Annual GIFT Users Symposium (GIFTSym4)* (pp. 49-60): U.S. Army Research Laboratory.

Bonner, D., Gilbert, S., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., . . . Holub, J. (2016). *The Challenges of Building Intelligent Tutoring Systems for Teams.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Bonner, D., Slavina, A., MacAllister, A., Holub, J., Gilbert, S. B., Sinatra, A. M., . . . Winer, E. H. (2016). The hidden challenges of team tutor development.

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

Cushard, B. (2012). Learning Design: When You Just Don't Know Where to Start. from http://www.bill-cushard.com/2012/10/learning-design-when-you-just-dont-know.html

Gilbert, S., Winer, E., Holub, J., Richardson, T., Dorneich, M., & Hoffman, M. (2015). *Characteristics of a Multi-User Tutoring Architecture.* Paper presented at the Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3), Orlando, FL.

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

Gilbert, S., Dorneich, M., Walton, J., & Winer, E. (in press). Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach. In J. Johnston, R. Sottilare, A. M. Sinatra & C. S. Burke (Eds.), Building Intelligent Tutoring Systems for Teams: What Matters. Bingley, UK: Emerald Publishing.

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

Salas, E. (2015). *Team training essentials: A research-based guide*: Routledge.

Selfridge, O. G. (1958). Pandemonium: a paradigm for learning in mechanisation of thought processes.

Sottilare, R., Burke, C., Salas, E., Sinatra, A., Johnston, J., & Gilbert, S. (2017). Towards a design process for adaptive instruction of Teams: A Meta-Analysis. *International Journal of Artificial Intelligence in Education*.

Sottilare, R. A. (2018). A Comprehensive Review of Design Goals and Emerging Solutions for Adaptive Instructional Systems. *Technology, Instruction, Cognition & Learning, 11*(1).

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT).

Sottilare, R. A., Holden, H. K., Brawner, K. W., & Goldberg, B. S. (2011). *Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training*.

Chapter 21 – Team Taskwork for Knowledge Building in Cyberlearning

Ronghuai Huang [1], Jeremiah T. Folsom-Kovarik [2], Xiangen Hu [3,4], Jing Du [1], Junfeng Yang [5]
Beijing Normal University[1], Soar Technology, Inc.[2], University of Memphis[3], Central China Normal University [4],
Hangzhou Normal University[5]

Introduction

With digital technology and the Internet infused in education, the research and practice of learning by means of digital technologies, or cyberlearning, is growing quickly. In order to understand how people learn in technology-rich learning environments, the computational modeling of learners and groups are of great importance to promoting learning science. Considering the interactions of the three forms of cyberlearning (individual learning, collaborative learning, and collective learning), we draw from the research of social learning, knowledge building, and computer-supported collaborative learning (CSCL). Based on the analysis of the character of social learning, knowledge building and taskwork, we propose a structure model for dynamic team formation and taskwork. The model will enable evaluating the process and effect of knowledge building and team taskwork by considering each individual's contribution to the team.

Infusing Cyberlearning into Classroom Learning

With the expansion of the digital world, every sector of society has been influenced by information and communications technology (ICT). The digital world is a combination of the real world and virtual world, thus, people's life style and ways of working have changed from real world to the blended world, as well as the ways of learning (Liu, et al., 2017).

Nowadays, learning is not confined to only physical settings, like a classroom, but also happens in cyber settings, in either a synchronous or asynchronous approach. Informed by learning science, cyberlearning is the use of new technology to create effective new learning experiences that were never possible or practical before (CIRCL, 2017).

In 2008, a report on "Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge" (Borgman, et al., 2008) was presented by the US National Science Foundation (NSF), which stated the background for cyberlearning, strategies for building a cyberlearning framework, and opportunities for action. Cyberlearning offers new learning and educational approaches and the possibility of redistributing learning experiences over time and space, beyond the classroom and throughout a lifetime (Borgman, et al., 2008). From the report, it is clear that "cyberlearning" reflects a growing interest in managing the interactions of technology and education, especially concerning the use of Internet and information technologies.

China has implemented the ConnectSCS (connecting Schools through the broadband network, connecting Classes with quality digital resources, and connecting Students in cyber learning spaces) project since 2011. During this period, school Internet access rose from 25% in 2011 to 89.8% in 2017; 85.1% of schools were equipped with multimedia classrooms in September 2017; 55% of teachers have used cyber learning spaces, and 38.4% students have used cyber learning spaces by September 2017 (MOE China, 2017). ICT in education in China has changed from emerging and applying to integration and infusion.

In China, the "Guidance on active promotion of 'Internet +' Initiative" was initiated by State Council in July 2015 (China State Council, 2015), with the aim to integrate mobile Internet, cloud computing, big data and the Internet of Things with modern manufacturing, to encourage the healthy development of e-commerce, industrial networks, and Internet banking, and to help Internet companies increase their international presence. Part 6 of the action plan is "'Internet Plus' Services for the benefit of people", with the No. 5 of this part focuses on "Explore new service supply mode for education". The main foci are: (1) encourage Internet enterprises and private educational institutions to develop digital education resources and to provide educational services via the Internet according to market demands; (2) encourage public schools to explore the new model of education by using digital education resources and educational services platforms, expand the coverage of quality education resources and promote education equity; (3) encourage schools to cooperate with Internet enterprises to provide new ways of public educational services for basic education, vocational education and so on, by connecting online learning resources and offline learning resources; and (4) promote the sharing of online courses in colleges and universities, expand the new cyberlearning methods (such as Massive Open Online Courses), explore the system of learning credits recognition and learning credits transfer, and accelerate the reform of service mode of higher education.

From the US to China, and from west to east, ICT in education has been paid more and more attention and is often regarded as the transformative force to innovate education. With ICT infused in education, the learning environment has become more blended, and learning in cyber spaces should be connected with classroom learning. It is urgent to understand how to connect different forms of learning in both physical spaces and cyber spaces.

Collaborative Knowledge Building and Cyberlearning

We consider three types of learning: individual learning, cooperative learning, and collective learning. Individual learning is structured by creating no interdependence among learners; individuals perceive that the achievement of their goals is unrelated to other individual performance (Johnson and Johnson, 1978). Cooperative learning is the instructional use of small groups through which students work together to maximize their own and each other's learning (Johnson et al., 1994). It is similar to collaborative learning, which emphasizes that learning occurs as an effect of community (Johnson and Johnson, 1999). Collective learning refers to a conceptualization of learning that takes the structures and processes of social cooperation into account as a "reality sui generis" (Miller, 1987). Collective learning represents a macro concept that addresses learning at the levels of the team, the organization, and society, which can be conceived as an evolutionary process of perfecting collective knowledge (Garavan and Carbery, 2012).

Cyberlearning, technologies are used to facilitate the efficiency of individual learning, cooperative learning, and collective learning, by exploring connections among human learning, cognition, collective intelligence and technologies, and even artifacts. Cyberlearning not only addresses the next-generation genres (types) of learning technologies, but also how people learn in technology-rich learning environments through data collection and computational modeling of learners and groups of learners that can be done only in such environments (Borgman, et al., 2008).

Modeling of learners and groups of learners is the key to understanding how people learn in the individually, through cooperative learning, and collective learning. Since the emergence of intelligent tutoring systems (ITS), the modeling of learners has been explored as an important part of the tutoring process. In this Internet age, individual learning, cooperative learning and collective learning could happen hand in hand in a cyberlearning setting, and the modeling of the group should get insights from this collaborative knowledge building.

Computer Supported Collaborative Learning (CSCL) was first described as a pedagogical approach wherein learning takes place via social interaction using a computer or through the Internet (Stahl, Roschmann, & Suthers, 2006). This kind of learning is characterized by the sharing and construction of knowledge among participants using technology as their primary means of communication or as a common resource (Stahl, Roschmann, & Suthers, 2006). However, since 2005, CSCL is increasingly deployed in support of other pedagogy, rather than as a distinct pedagogical approach (Dillenbourg, Järvelä, & Fischer, 2009). Collaborative activities are becoming integrated within comprehensive environments that include non-collaborative activities stretching over the digital and physical spaces and in which the teacher orchestrates multiple activities with multiple tools (Dillenbourg, Järvelä, & Fischer, 2009). In the research and practice of CSCL, the cultural differences of eastern world and western world should be noted. As an example, students in eastern world attached great importance to maintaining a harmonious atmosphere and maintaining relationships, which normally leads to a superficial discussion in CSCL. However, students in the western world are more focused on the learning task, and when their opinions are different from others, they usually directly expressed what they thought, which will lead to relatively in-depth discussion (Yang et al., 2014).

In cyberlearning, the three different forms of learning are dynamically interconnected in both physical and virtual spaces, and CSCL provide valuable insights to the collaborative knowledge building in cyberspace. In order to deeply understand the group forming process and collaborative mechanism, the following section will focus on social learning, knowledge building and team taskwork.

Team Taskwork for Knowledge Building in Cyberspace

Social learning was defined as “a learning theory that emphasizes how people learn from others through observation, imitation, and modeling” (Bandura, 1963). In order to stress the dynamic interaction between people and the context in the construction of meaning and identity, Reed et al. (2010) defined social learning as a change of understanding that goes beyond the individual to become situated in wider social units or communities of practice through social interactions between participants within social networks.

Therefore, social learning must (1) demonstrate that a change in understanding has taken place in the individuals involved; (2) demonstrate that this change goes beyond the individual and becomes situated within wider social units or communities of practice; and (3) occur through social interactions and processes between actors within a social network (Reed, et al., 2010).

Knowledge building could be defined as "the creation, testing, and improvement of public knowledge/cultural artifacts" (Scardamalia, Bereiter, & Lamon, 1994), which is a social and collaborative process to advance knowledge and ideas of value to a community, with individual learning and growth of members as an important by-product (Zhen, 2012). Based on the ontological analysis of our existence as made up of three interacting worlds: World 1 (the physical), World 2 (the subjective) and World 3 (the locus of cultural products), Bereiter (2002) believed that learning took place in World 2, while knowledge is built in World 3. One component of knowledge building is the creation of “epistemic artifacts,” tools that serve in the further advancement of knowledge (Sterelny, 2005). These may be purely conceptual artifacts (Bereiter, 2002), such as theories and abstract models, or “epistemic things” (Rheinberger, 1997), such as concrete models and experimental setups.

Stahl (2000) proposed a model of collaborative knowledge building to describe the process and elements of interactions between personal knowledge and public knowledge, as shown in Figure 1. From the figure, the relationship of individual learning, collaborative learning, and collective learning could be easily understood. Social knowledge building was the process of collaborative creation of public, collective knowledge, while in the process individual learning is an essential and demonstrable by-product.

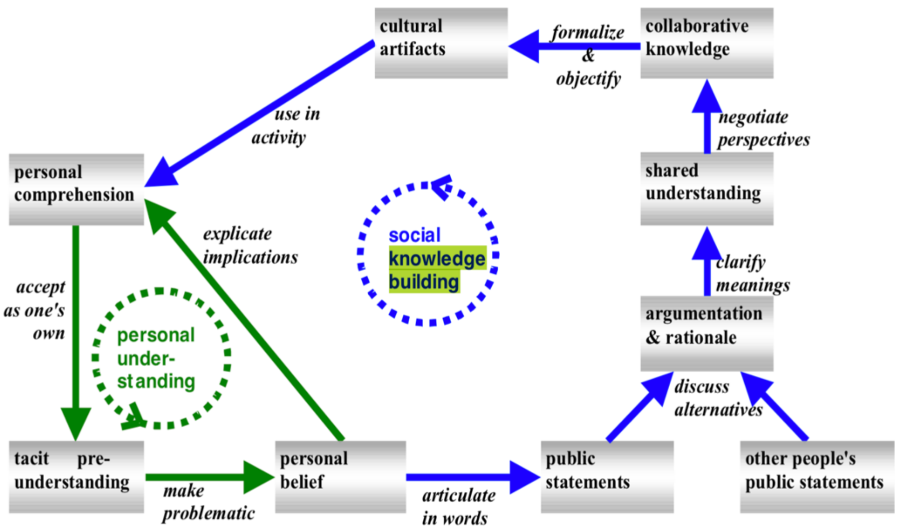


Figure 1. A diagram of knowledge building processes

Knowledge building stemmed from an early emphasize supporting individual intentional learning and expertise and has evolved to emphasize social production and improved collective knowledge. The change of this emphasis is of great importance to the design and use of computer-based environments to support collaborative learning and knowledge growth. In such a period, how individuals contribute to the development of public knowledge and at the same time gain their own individual knowledge are the key for learner modeling and group modeling for cyberlearning. Teamwork and taskwork are the two concepts that should be considered in the modeling of knowledge building.

Team taskwork refers to those relevant individual behaviors that directly lead to the successful accomplishment of collective goals (Salas, 2015), while teamwork on the other hand describes the activities performed by a group of people focusing on how they work together. Teamwork consists of behaviors that are related to team member interactions and are necessary to establish coordination among individual team members in order to achieve team goals, whereas taskwork consists of behaviors that are performed by individual team members and are critical to the execution of individual team member functions (Morgan et al, 1986). Team performance is comprised of both teamwork and taskwork. Measuring teamwork requires identifying dimensions of teamwork or processes that comprise the teamwork construct, while taskwork requires identifying individual behaviors and therefore the specific team functions (Yeboah-Antwi, et al., 2013).

The study of teamwork is becoming popular in service, hospitality, and tourism organizations because of its influence on service quality, productivity, and organizational effectiveness (Jong, Wetzels, & Ruyter, 2008); Hu et al., 2007). Teamwork can be used as a knowledge building activity or a creative learning strategy. How to model teamwork and taskwork to promote the best performance of groups is still a challenge for technology-enhanced group learning.

Intelligent tutoring frameworks like Generalized Intelligent Framework for Tutoring (GIFT) should include support for team taskwork (Sottilare et al., 2011). A typical Intelligent Tutoring System (e.g. GIFT) is considered to have four major components: the domain model, the student model, the tutoring model and the tutor-student interface model. An ITS could be enhanced to include a group model. The proposed dynamic group model consists of two parts. One part is distributing different students into groups in an online system relying on the structure of student model and tutoring model, which is mainly focusing on group formation problem. The other part is supporting team taskwork with the consideration of the structure of the domain model, which is mainly focusing on profiling taskwork and team assessment.

Profiling Group and Group Processing

The collaboration between people in a virtual environment can be used as an important part of knowledge building. This trend is also reflected in the field of education in which students collaborate via an Intelligent Tutoring System. The increasing popularity of Intelligent Tutoring Systems has enabled students with different characteristics, skills and background to collaborate with each other in a teamwork setting, especially when students need to collaborate in different short-term groups, but each turn changes their role in the team. So to improve team performance and service quality, guarantee productivity and effectiveness of a team, we will put forward a model to identify dimensions of teamwork or process in collaborating learning.

There exist several methods which solve a group formation problem in the educational domain (Isotani et al., 2009). The significant part of these methods focuses on long-term groups which collaborate on complex tasks during several days or even weeks. A minority of existing methods are aimed to propose short-term groups, but they usually consider only a single assignment of students into groups ignoring following collaboration. One of the most cited group formation models is Tuckman's small group development model. Tuckman proposed a model with four stages of group development: forming, storming, norming and performing in 1965. Then Tuckman and Jensen reviewed the original model and added a final stage called adjourning in 1977, as shown in table 1 (Tuckman & Jensen, 1977).

Table 1. Stages of Group Development (Tuckman, 1965; Tuckman & Jensen, 1977)

Attributes	**Group Structure** The pattern of interpersonal relationships; the way members act and relate to one another.	**Task Activity** The content of interaction as related to the task at hand.
Forming: orientation, testing and dependence	Testing and dependence	Orientation to the task
Storming: resistance to group influence and task requirements	Intragroup conflict	Emotional response to task demands
Norming: openness to other group members	In-group feeling and cohesiveness develop; new standards evolve and new roles are adopted	Open exchange of relevant interpretations; intimate, personal opinions are expressed
Performing: constructive action	Roles become flexible and functional; structural issues have been resolved; structure can support task	Interpersonal structure becomes the tool of task activities; group energy is channeled into the task; solutions can emerge
Adjourning: disengagement	Anxiety about separation and termination; sadness; feelings toward leader and group members	Self-evaluation

Tuckman's model has been already successfully applied to localized long-term study groups, but it is less suitable in online environments where storming and norming are sometimes short-circuited. However, Tuckman's model laid a solid foundation for many other specialized groups' lifecycle models as the attributes are key to long-term groups and distributed groups (e.g. those created in an ITS). Another model with a four-step scheme focuses on short-term group formation in a virtual learning environment (Daradoumis, et al. 2002). This model shows more details about how groups form in online learning environments (Table 2).

Table 2. Description of the group formation process (average duration time: 8 days)

	When	Why	How
Initiation	Initial action of the approach (duration: 2 days)	To initiate students into the new experience of virtual collaboration and to enable them to understand the notion and function of collaborative learning groups.	All students collaborate together to resolve a specific case study about what they need to know and do in order to construct effective virtual collaborative learning groups
Introduction	Second action of the approach (duration: 1 day)	To provide both one's own and other relevant information in order to enhance a deeper knowledge of each other and to promote a better interaction.	Students should work out a personal report with important information, such as personal data, expertise level, work pace, available working time, temporal coincidence, goals, and attitudes to collaborative learning, social aspects of collaboration, and previous experience in groupware.
Negotiation	Third action of the approach (duration: 4 days)	To form a learning group that satisfies both individual and group goals or to search for an open group that fits better one's personal goals and needs.	Each student initiates a negotiation process either with individual candidate members whose characteristics match with his/her own in order to form a reliable and effective learning group or with a possible open group in order to become an active member of it.
Group Proposal	Final action of the approach (duration: 1 day)	To inform and ask the tutor to approve the definite formation of a learning group and the initiation of the next collaborative phase.	A member of the recently constructed learning group informs the tutor about the group's constituent members, facilitates each member's data and asks for its final approval.

The main focus of our research is to support both short-term and long-term virtual groups. These dynamic groups exist especially in online learning systems in which students can choose the time, the content and the place they want to study, such as Intelligent Tutoring System (ITS), thus, groups need to be created in real time with consideration about students' skill, knowledge and learning context.

Considering both the characteristics of individuals and the characteristics of teams, we propose a structure model useful for dynamic team formation and evaluation. The model includes grouping, taskwork initiating, team working and group assessing. All information generated during collaboration can be stored in a database (in Figure 2). In this part, we describe our model of the grouping function first, and represent the taskwork part in the following section.

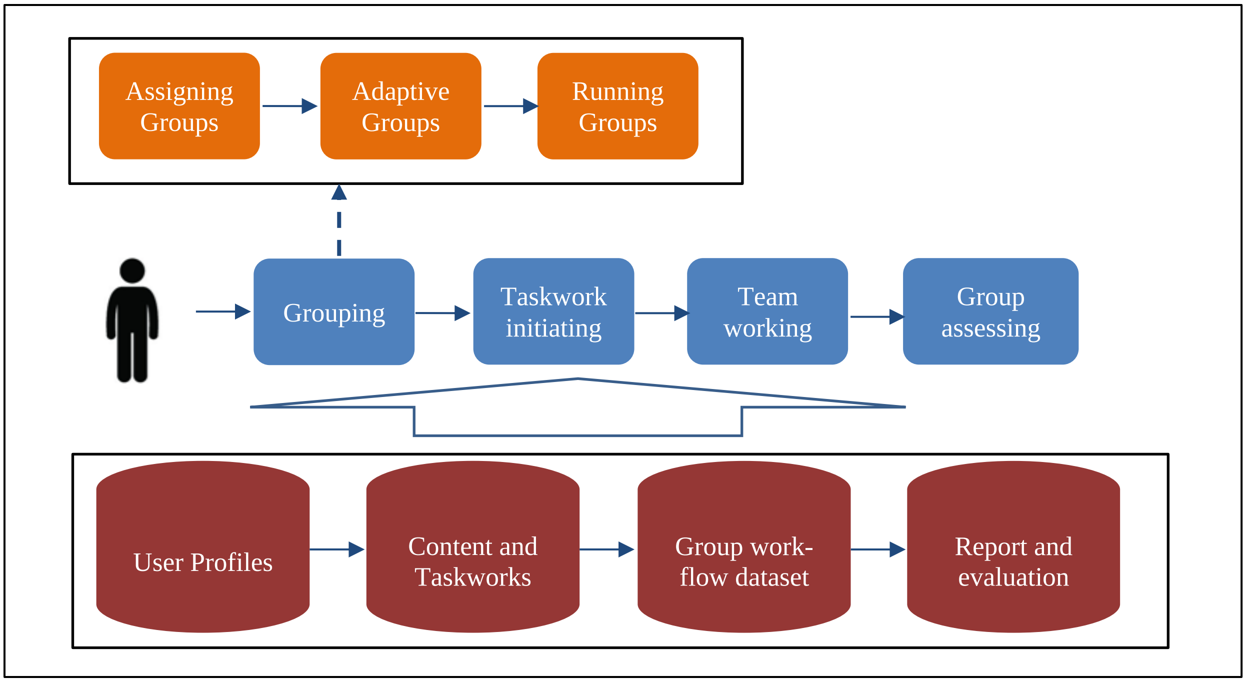


Figure 2. A structure model for dynamic team formation and evaluation (DTFE Model)

The main goal of the first stage of group development is how to distribute students into groups, known as group formation. Instead of selecting students randomly or grouping them manually by a teacher, automatic computer-supported methods are proposed. When identifying learning objectives (includes the learner's objective, peer's objective, and his/her group's objective), several individual factors need to be considered, including knowledge level, learning styles, student's preference. We can describe the user's characters by the ABC User model (Abhiraj & Rodney, 2013). The ABC User model can track affective states, behavioral and cognitive states and patterns of the users and analyze the states and patterns of the user which can provide an adaptive e-learning environment.

During the grouping process, each group is in one of three states, including assigning groups, adapting groups, and running groups. Assigning groups is at the beginning of grouping which focus on initiation action and introduction. Adapting groups focus on negotiation as the members of a group did not know each other before, and they need some time to identify and learn about their peer learners. In running groups, group members have a good interaction with each other and are well-prepared for the next step of collaborative study.

Dynamic modeling taskwork for cyberlearning

Taskwork results in observable group performance and by the method of team-based assessment, participants will learn how their team is doing and whether they have achieved the learning goal. The taskwork part of the model consists of three functions, including taskwork initiating, team working, and group assessing (Figure 2). It has a strong relation to the domain model and user-interface model in ITS.

"Taskwork initiating" describes the series of learning activities for accomplishing the collaborative task, including discussion, decision, peer instruction, and sharing documents. For a collaboration group, the grouping process organizes some learners together, and the learners form a set of a certain size and structure, but this collection does not have the interactive elements of the collaborative group. While the grouping process is transparent to learners, learners do not realize that they exist in a group and are about to establish a connection with others. At this stage, learners should be integrated into the collaborative group, not only to clearly recognize the various attribute parameters of the collaborative group, but also to identify the environment of the collaborative group. Specifically, the primary task in this phase is to identify learning goals, roles and principles, including the goals of individuals and the entire collaboration group, to understand the roles of individuals and others, to be familiar with the principles in which their collaborating groups interact and complete collaborative tasks, and so on to prepare for learning behavior.

″Team working″ describes how team members coordinate with each other, including goal specification, mission analysis, conflict management and progress monitoring. After the group preparation phase, the attribute parameters of the collaboration group and the control parameters related to the specific learning task have been determined, and the collaboration group members can collaborate on the realization of the common goal. In the process of learning specific content and filling a particular role, the members of the collaboration group need to provide a variety of content materials, courseware and other resources. The basic material, tools or corresponding functions of a collaborative learning task can be provided to support them. The interactive behaviors of collaborative members at this stage will be recorded as a basis for assessment.

″Group assessment″ refers to the criterion to evaluate a team's performance including both the accomplishment of the learning objective (test scores, academic performance), and the degree of collaboration in the group. After team working, the collaborative group members enter the test evaluation phase. The content of the test includes the completion of the learning goal and the degree of collaboration in the collaborative group learning process. The completion of the learning goal can be tested by means of general teaching quality, and an evaluation of the degree of collaboration that can be accomplished by the measurement of cohesion strength in the cooperative group. The results obtained through online testing are recorded in the report and evaluation database, which can be used to characterize the degree of collaboration of a collaboration group and be stored in the corresponding collaboration file.

The team assessment is a fully automated, web-based function that helps a team better understand its susceptibility, strengths, and weaknesses, as well as those areas where team members have significantly differing perspectives. For assessment, there exist some lessons in GIFT Analysis Construct. The GIFT Analysis Construct emphasizes the evaluation of the effect on learning, performance, retention and transfer. The goals of GIFT Analysis Construct are to support the conduct of both formative and summative assessment, to provide diagnose for identifying areas for individual's improvement, to understand how learning is progressing. These goals lay a solid foundation for why we put forward DTFE model.

At present, collaborations among learners happen in many web-based learning systems. Therefore, the study on group formation has a significant effect on knowing more details about the process of online collaborative learning. Although there are many existing methods for grouping formation, our concern is how to create dynamically both short-term and long-term groups repeatedly and automatically without the intervention by teachers. Future research will undoubtedly be needed to develop the system to realize the proposed DTFE Model and test the effectiveness of the model.

References

Abhiraj, T., & Rodney D.N. (2013). Affective-Behavioral-Cognitive (ABC) Learner Modeling. Design Recommendations for Intelligent Tutoring Systems (Volume 1 Learner Modeling). 75-85.

Baecker, R. M., Grudin, J., Buxton, W. A. S., & Greenberg, S. (1995). *Human-computer interaction: toward the year 2000*. Morgan Kaufmann Publishers.

Bandura, A., & Walters, R. H. (1963). Social learning and personality development. *American Sociological Review, 23*(5), 634–635.

Bereiter, C. (2002). Education and mind in the knowledge age. *Computer-Mediated Communication Magazine, 5*, 11-27.

Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., … others. (2008). Fostering learning in the networked world: The cyberlearning opportunity and challenge. Washington, DC: NSF. Retrieved from http://www.oxydiane.net/IMG/pdf_NSF_on_Cyberlearning_final.pdf

China State Council. (2015). Guidance on active promotion of "Internet +" actions. Retrieved from http://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm (in Chinese)

CIRCL. (2017), About CIRCL(The Center for Innovative Research in CyberLearning), retrieved from http://circlcenter.org/about/

Daradoumis, T., Guitert, M., Gimenez, F., & Marques, J. M. (2002). Supporting the composition of effective virtual groups for collaborative learning. *International Conference on Computers in Education*,1,332. IEEE Computer Society.

Dillenbourg, P., Baker,M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reimann (Eds.), Learning in humans and machine: Towards an interdisciplinary learning science. (pp. 189–211). Oxford, United Kingdom: Elsevier.

Dillenbourg, P., Järvelä, S., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (Eds.), *Technology-Enhanced Learning: Principles and Products*, 3–19. Springer Netherlands Publisher.

Garavan, T. N., & Carbery, R. (2012). Collective Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning*. (pp. 646–649. Boston, MA: Springer US.

Menglei [Hu, M. L. M. H., Jeoushyan, H., & Yuhua [Sun, Y. H. C. S. (2009). Hospitality teams: knowledge sharing and service innovation performance. *Tourism Management*, *30*(1), 41-50.

Isotani, S., Inaba, A., Ikeda, M., & Mizoguchi, R. (2009). An ontology engineering approach to the realization of theory-driven group formation. *International Journal of Computer-Supported Collaborative Learning*, *4*(4), 445–478. https://doi.org/10.1007/s11412-009-9072-x

Johnson, D. W., & Johnson, R. T. (1978). Cooperative, competitive, and individualistic learning. *Journal of Research & Development in Education*, *12*, 3-15.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1994). Cooperative learning in the classroom. Alexandria: Association for supervision and curriculum development.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into Practice*, 38(2), 67–73.

Jong, A. D., Wetzels, M., & Ruyter, K. D. (2008). Linking employee perceptions of collective efficacy in self-managing service teams with customer-perceived service quality. *International Journal of Service Industry Management*, *19*(3), 353-378.

Kearsley, G. (2013). The Theory Into Practice Database. https://icebreakerideas.com/learning-theories/

Liu, D., Huang, R., & Wosinski, M. (2017). Smart Learning in Smart Cities. Singapore: Springer Nature Singapore Pte Ltd.

Miller, D. L. (2003). The stages of group development: A retrospective study of dynamic team processes. Canadian *Journal of Administrative Sciences*, 20(2), 121–134.

Miller, M. (1987). Argumentation and cognition. In M. Hickmann (Ed.), Social and functional approaches to language and thought. (pp. 225–250. New York: Academic

MOE China (2017). Monthly report for ICT in Education of September 2017. Retrieved from http://www.moe.gov.cn/s78/A16/s5886/s6381/201710/t20171025_317440.html

Moreno, J. L. (1951). Sociometry, Experimental Method and the Science of Society: An Approach to a New Political Orientation. Beacon House.

Morgan BB, Glickman AS, Woodard EA, Blaiwes AS, Salas E (1986). Measurement of team behaviors in a Navy environment, NTSC Technical Report No. TR-86-014. Orlando, FL: Naval Training Systems Centre.

Resta, P. & Laferrière, T. (2007). Technology in Support of Collaborative Learning. *Educational Psychology Review*, 19, 65–83.
Reed, M., Evely, A. C., Cundill, G., Fazey, I., Glass, J., Laing, A., Stringer, L. C. (2010). What is social learning? *Ecology and Society*, 15(4).
Rheinberger, H-J. (1997). Toward history of epistemic things: Synthesizing proteins in the test tube. Stanford, CA: Stanford University Press.
Salas, E. (2015). Team training essentials: A research-based guide. Routledge.
Scardamalia, M., & Bereiter, C. (2003). Knowledge Building. In Encyclopedia of Education. (2nd ed., pp.1370-1373). New York: Macmillan Reference, USA.
Scardamalia, M., & Bereiter, C. (2010). A Brief History of Knowledge Building. *Canadian Journal of Learning & Technology*, 36(1), 1~16.
Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom into world 3. In K. McGilly (Ed.), Classroom lessons: Integrating cognitive theory & classroom practice. (pp.201-228). Cambridge, MA: MIT Press.
Sottilare, R., Holden, H., Brawner, K., & Goldberg, B. (2011, December). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. In Proceedings of the Interservice/Industry Training Simulation & Education Conference, Orlando, Florida, December 2011.
Stahl, G. (2000). A Model of Collaborative Knowledge-Building. In B. Fishman & S. O'Connor-Divelbiss (Eds.), Fourth International Conference of the Learning Sciences (pp. 70-77). Mahwah, NJ: Erlbaum.
Stahl, G., Roschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), Cambridge handbook of the learning sciences (pp. 409-426). Cambridge, UK: Cambridge University Press.
Staker, H., & Horn, M. (2012). Classifying K-12 Blended Learning. Innosight Institute. Retrieved from http://files.eric.ed.gov/fulltext/ED535180.pdf
Sterelny, K. 2005. Externalism, epistemic artefacts and the extended mind. In (R.Schantz, ed) The Externalist Challenge: New Studies on Cognition and Intentionality. Berlin: de Gruyter.
Tuckman, B. W, Jensen, M.A. (1977). Stages of small-group development revisited. *Group & Organization Studies*, 2(4), 419-427.
Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63(63), 384.
Windschitl, M. (1998). The WWW and classroom research: What path should we take? *Educational Researcher*, 27(1), 28-33.
Yang, J., Yu, H., Chen, S., & Huang, R. (2014). Strategies for Smooth and Effective Cross-Cultural Online Collaborative Learning. *Educational Technology & Society*, 17(3), 208–221.
Yeboah-Antwi, K., Snetro-Plewman, G., Waltensperger, K. Z., Hamer, D. H., Kambikambi, C., & Macleod, W., et al. (2013). Measuring teamwork and taskwork of community-based “teams” delivering life-saving health interventions in rural Zambia: a qualitative study. *BMC Medical Research Methodology*, 13, 1, 84-84.

Chapter 22 – Developing the GIFT Event Report Tool to Support Experimentation for Teams

Michael W. Boyce[1], Anne M. Sinatra[1], Stephen B. Gilbert[2], Robert A. Sottilare[1]
U.S. Army Research Laboratory[1], Iowa State University[2]

Introduction

The Generalized Intelligent Framework for Tutoring (GIFT) is an open source framework for creating Intelligent Tutoring Systems (ITSs). GIFT can provide tailored instruction and remediation that takes into account the current state of the learner, and learner attributes such as individual differences in various domains (Sottilare, Brawner, Goldberg, & Holden, 2012; Sottilare, Brawner, Sinatra, & Johnston, 2017). GIFT is available in both downloadable and in online form (known as GIFT Cloud at https://cloud.gifttutoring.org). GIFT includes authoring tools that can be used to create "GIFT courses," which are a sequence of materials, questions, and instruction that is presented to a learner. While GIFT is primarily a system for authoring ITSs, it can also be leveraged for use in experimentation in both traditional and ITS relevant experiments. For the purposes of experimentation, one of the major advantages of GIFT is its ability to extract participant data from GIFT courses through the use of either the desktop based Event Report Tool (ERT) or the GIFT Cloud Event Report Tool (Cloud ERT). Each time learners participate in a GIFT course, a log file is created that includes all of their entered data, responses to questions, and a record of their actions. Using the Event Report Tools, experimenters can select the specific GIFT data pieces of interest and export those as comma separated value files, which can be easily imported into Microsoft Excel. The Army has expressed a growing need for applying ITS approaches to teams, through Intelligent Team Tutoring Systems (ITTSs). There is also an increase in interest in developing GIFT Cloud to provide a proper mechanism for collecting team-based data. Part of creating a framework for ITTSs is not only providing guidance and authoring tools for the collection of team performance data, but also export tools that provide data in an understandable way. While both the team authoring and export aspects of GIFT are not currently implemented, this chapter's focus provides a starting point on how to make the export tools (ERT) more suitable for team-based data collection. The current chapter will focus on the team elements, while also providing recommendations for overall improvements to the ERT's flow and organization. Although the emphasis is on teams, the suggestions provided can help individual-based data collection as well.

The structure of this chapter begins with a discussion on the related literature to date on data reporting for teams. It then looks at challenges that are faced by researchers trying to run experimentation with teams on GIFT and the needs they have. This follows with some high level recommendations to fit those needs, and concludes with an initial mockup of potential future ERT functionality.

Related Literature

History of Intelligent Tutoring Systems to Support Teams

There is a vast literature expressing the characteristics of team training, team tutoring, and team performance metrics that is well beyond the scope of this chapter, but a high-level discussion can assist in providing context for the rest of the chapter. There have been a couple of attempts at developing team tutoring systems. The Advanced Embedded Training System was developed by the Navy to support team training on ships (Zachary et al., 1998). The system acted as a support tool to reduce workload. It performed less than optimally because the amount of feedback in a real-time team scenario turned out to exceed the capabilities of the instructor to provide remediation, thereby reducing performance. More recently, researchers using the Team Multiple Errands Task (TMET) were able to quantify and assess team performance without demonstrating ceiling or floor effects (Bonner et al., 2016; Walton, Gilbert, Winer, Dorneich, & Bonner, 2015). This computer-based task, which involved coordinating with a team and purchasing items from a list, had strong team characteristics and necessitated interdependence from the team members.

Sottilare and colleagues (Sottilare, Burke, et al., 2017) performed a large scale meta-analysis on connecting teamwork behaviors (communication, coordination, cognition, etc.) to the appropriate team outcomes (learning, performance, satisfaction, and viability). Their research focused on team measurements being represented by attitudes, behaviors, and cognition (Sottilare, Burke, et al., 2017). Understanding the states and traits of an individual and how those relate to the progression of goals can provide guidance on specific actions during a tutor event. The research resulted in identifying six sets of behavioral markers: trust, collective efficacy, cohesion, communication, and conflict / conflict management (Sottilare, Burke, et al., 2017), which can assist in measuring team behaviors. While this research focused on a large scale meta-analysis of the team literature, the majority of the articles available were in non-computer based and non-ITS related areas. Additionally, the behavioral markers identified must still be operationalized in order to be implemented in an ITTS. This research demonstrates the need for further investigation on learning with teams and the complexities that come with ITTSs.

Teams and Learning through Data

Teams of two or more are the building blocks for all military collective performance. It is important to be able to quantify how teams are performing relative to their stated goals, the solutions being generated to solve team problems, and how the complex challenges from the military are being met (Sottilare, Burke, et al., 2017). Past research identified some of the big challenges in developing ITTSs. This includes: measuring team performance, improving team performance, and studying team formation and development (Dorsey et al., 2009; Sottilare, Burke, et al., 2017). A defining characteristic of an ITTS is that it needs to account for individual interactions as well as team interactions with the tutor, and take into consideration the external factors associated with the environment of interest (Bonner et al., 2016; Gilbert et al., 2017). Since team skills contain social components like communication and coordination, it becomes harder to represent them quantitatively (Gilbert et al., 2017). In the military, communication and coordination can be manifested through collaborative learning or cooperative learning.

Irrespective of the type of learning, these additional components increase the amount of data available exponentially as each additional learner is considered (Bonner et al., 2016). Team data can come from a variety of sources such as real time learner data, learner states determined from classification, and long term attributes established from previous data (Sottilare, Burke, et al., 2017). Data can be used both as a source for experimentation or as a basis to shape the learner experience. A challenge for ITTSs is being able to transform data based on empirical hypotheses into a format that can be readily interpreted and understood (MacAllister et al., 2017).

When conducting research with an ITTS, there are four forms of data to collect and analyze. The first type is the low-level data such as users' interface clicks, movements of entities in a system, and timing between events. These data are typically passed along to the tutor to trigger real-time feedback. This process is what VanLehn (2006) called the "inner loop" of an ITS. To analyze team data, one might conduct a "team loop" process, in which the behaviors of the team as a whole (a second type of data) are sent to the tutor to generate team feedback. The third type of data is present in VanLehn' s "outer loop" process, in which data about users' accumulated skill profile or learner model (not their real-time clickstream) are used to choose the best next training scenario for them based on the skills that need bolstering. The fourth type of data are those that are averaged across multiple teams and used for statistical analysis by researchers to evaluate learning effectiveness or the usability of the system (sometimes called educational data mining or educational data analytics). Each of these four types represents four different forms of data-based decision making that need to be made by either the tutor or the researcher (Gilbert, Dorneich, Walton, & Winer, in press).

Past research using GIFT with teams made strides on increasing understanding and working with data through visualization. Data Visualization is the presentation of data in a graphical format so that it is easier to understand (Chen, Härdle, & Unwin, 2007). Data visualization can assist in organizing and grouping information to make it align with the mental model or schema of decision-makers with many different styles and techniques available (Bertini, Tatu, & Keim, 2011). Data visualization can be static (i.e. a snapshot of user / team performance) or it can be dynamic (e.g. a dashboard with consistently updating information). A data visualization can be passive, such as a view that can't be changed or it can be interactive where individuals can modify the data as needed and the data is rendered to represent those changes. Data visualization can also be expanded to encompass related information that is not directly represented within the data (Chen et al., 2007). As datasets become larger, individuals are forced to comprehend this data quickly and accurately.

Research conducted at Iowa State University used multiple techniques to manipulate and represent data for two-person teams. The researchers created customized post-processing solutions to be able to represent team variables (Gilbert et al., 2017; MacAllister et al., 2017). They also created a timeline chart, which provided information on the specific activities participants were doing during critical points in the experiment (MacAllister et al., 2017). While the output created was highly task-dependent, the approaches that they used (i.e. time series analyses), and data they extracted are relevant in determining the types and groupings of data that are relevant to extract from an ITTS (MacAllister et al., 2017).

Outside of the military environment, Dashi (2016, 2017) used Excel to generate macros to analyze student data. The data could be analyzed in a post processing format or during a class as they were interacting with an online learning platform, as well as creating pivot table solutions to better visualize the data. Specifically, student engagement was quantified through metrics such as mouse clicks, page views, and quiz scores. Data representations like these can provide insight on both the underlying empirical data and the complex relationships which often accompany team-based research studies.

Information and Data Visualization Foundations to Support Understanding

In developing concepts for improvements to the ERT, seminal research by two of the key contributors to the area of information and data visualization were examined: Ben Shneiderman and Edward Tufte.

Shneiderman, a leader in the field of human-computer interaction, developed a list of "Eight Golden Rules for Interface Design" such as consistency, informative feedback, and reducing short term memory load (B. Shneiderman & Plaisant, 2005). He also developed a task by data type taxonomy to support data visualizations where he created his mantra for visual information seeking: "Overview-first, zoom and filter, then details on demand." He breaks this down further into seven tasks when working with data visualizations:

- Overview – Provide an overview of the entire interface
- Zoom – Zoom in on areas of interest
- Filter – Remove non-relevant items
- Details on Demand – Provide ability to select specific items for more information
- Relate – Show relationships among items
- History – Maintain a history to support undo
- Extract – Allow for collections of subsets of the data (Ben Shneiderman, 1996)

These tasks are similar to the tasks that an individual might complete when using the ERT.

Edward Tufte is one of the key figures in maximizing the understanding of data representations. Tufte had four primary themes that echoed through his writings: graphical excellence, visual integrity, maximizing the data to ink ratio, and aesthetic elegance.

- Graphical excellence refers to expressing the greatest number of ideas in the simplest form as possible, using minimum amounts of space, and the fewest words.
- Visual integrity refers to having numerical scales that are proportionate to the values they represent. They should be tied directly to the data, rather than any sort of artistic interpretation.
- Maximizing the data-ink ratio refers to comparing the amount of ink needed to describe the data as opposed to the total ink used for illustrative purposes. The visualization should be less distracting and more useful for the user.
- Aesthetic elegance, in Tuft's view, is being able to clearly and simply display the complexity of data that is being represented via figures and tables (Tufte, 1983).

As organizing team data and output is a complicated and information-intensive task, it is important to consider visualization heuristics that will make the process easier to understand for users. It is through learning from previous research on teams, learning with ITTSs, and information visualization that we use to frame our concepts and mockups for improving the ERT for teams. However before going into improvements, there is a need to understanding the difficulties of running a team experiment in GIFT.

Challenges and Functional Needs for Team Experiments

When researchers retrieve data from an experiment, they face challenges to overcome and functional needs that must be fulfilled in order to support team based research.

Challenges

Tool Selection: Desktop vs. Cloud ERT

After collecting data, researchers can select the types of data that they want to extract from logs using GIFT's ERT. The ERT includes data categories such as survey responses, learner state, and more. However, researchers must decide whether to use the Desktop ERT or the Cloud ERT.

Both the Desktop and Cloud ERT architectures are currently configured to focus on individual learners. As mentioned earlier in the chapter, the shift from desktop to the Cloud is resulting in different requirements for the ERT and ultimately a redesign to improve usability. The Desktop ERT has can produce greater granularity and the ability to handle sensor based data, which the Cloud ERT cannot. The Cloud ERT focuses more on usability, but has limitations in regard to the types of data that can be extracted and the organization of the data.

For instance, the data from the Desktop version is pulled from an output folder in the GIFT installation folders, and allows experimenters to individually select the log files that they wish to include in the analysis. In the case of the Cloud ERT, all data from the specific instance of the course is housed online, and the output includes all logs relevant to the experiment. This may result in issues if there are participants with missing data or who had technical difficulty during the data collection. Experimenters need to realize that all participants were included and edit their output file appropriately to remove the data that should not be included. A current work-around for this issue is to download the log files from the cloud and import them into the Desktop version for analysis. However, this is not the ideal long-term solution. Below we discuss an improved approach for the Cloud ERT system.

Data Representation: Working with and Merging Data

Next, researchers need to decide the format for their data output. Currently, when experimenters want to collect data from an experiment run on GIFT Cloud, they have two options: download all of the raw data logs in order to export them using the desktop based ERT, or build a report by selecting the specific information of interest. If experimenters decide to build a report, they must first choose from frequently reported event types, training application event types, and other event types. Depending on the type of data that is included in the report, experimenters can select an option to combine all data for a single user onto a horizontal line. This option is useful in regards to survey output, and it facilitates the import of the output file into Excel or SPSS for further analysis.

It is important to note that in the Desktop ERT, experimenters have the option of merging data by certain characteristics like Use rid or Username, whereas in the Cloud version, the merging occurs based on the specific log and user session. While in most cases this would not be a problem, it may be helpful to include survey questions within a course which asks for a Use rid so that it is clear which user the data came from when the data file is output. Currently, one of the solutions for identifying how to group team data would be to include questions within the course that requires the team number and participant number to be entered. The data could then be sorted by the experimenter after it is output.

Participant number management is particularly important in the online ERT, as currently the only way to extract data from Cloud GIFT is through using the Publish Courses function and distributing links to participants that do not require logins. Due to this lack of a login requirement, the data logs that are being parsed are not associated with any particular participant number. It is important for the experimenter to realize this and include a question in their data set that asks for this information.

Functional Needs of the ERT

High-Level Needs

The high-level needs consist of those that could be applied across all experiments, which includes team-based experiments. They are discussed here to show both the potential of larger scale changes, and to delineate general changes from those that are especially relevant to teams.

Ontological Mapping of Across Levels of Data

To facilitate effective analysis, data needs to be encapsulated into a hierarchical model. At the lowest level is the raw data that is streamed from the system, it could be combined with data collected by human experimenters. Above that level are tools and methods to organize and present data for analysis. At the top level of the hierarchy is the collection of analyses compared against criteria for successful completion (Gilbert et al., 2017). Implementing this into the ERT will require a visual way for researchers to select and organize the log files that are being analyzed for the output.

Interface Enhancements to Work with Complex Experiments

One of the successes of the latest redesign of GIFT has been the incorporation of easier to use interfaces. This includes features like being able to drag and drop course objects in course authoring. In the same vein, an experimenter needs to be able to drag and drop experiment objects to represent their experiment design, much like what is possible in current leading experimental design software applications such as E-Prime (Schneider, Eschman, & Zuccolotto, 2002) and Open Sesame (Mathôt, Schreij, & Theeuwes, 2012). Rather than reinventing the wheel, integration with these software tools might provide the necessary infrastructure to better support experimenters using GIFT.

Matching Capabilities Between ERT Tools

As mentioned in the previous section, experimenters must choose between the Desktop ERT and the Cloud ERT. This presents a problem for most experimenters using GIFT because they are not going to have a clear understanding of the capability differences between the ERT tools unless they try to perform a function that exists only in one tool or the other. Since the ERT is a post processing tool, it does not have the runtime restrictions that experiments visualizing live data might require.

Scaffolding for First Time Users

The first time someone attempts to use the ERT, there needs to be scaffolding which demonstrates how the ERT works. This could be in the form of an instructional overlay with coach marks, where GIFT highlights a series of user interface features to show them how the ERT works. Although this is currently done using documentation and videos, a short action-based (non-voice) tutorial could help. It could be done by keeping track of every time a GIFT user enters a new part of GIFT that is unfamiliar. Then, after they see the tutorial once, they do not have to see it again. However, it could be retrievable again from a help menu or button on the screen if users feel that they need a refresher.

Linking to Data Sources

Linking to the sources of data can ensure that an experimental measure is being used as it is intended to be used. This could be done by providing references to previous data repositories, published research papers, or user's guides. It may also be of benefit to provide recommendations of related measures or data sets that might be of interest to the experimenter.

Team-Specific Needs

It is important to note that team experiments have different needs than individual learning experiments. The ERT can be improved and redesigned to allow researchers to include options to better frame team experiments, and provide easier to deal with data output. A few needs that we have identified include the following.

Team Variables

There needs to be a way to set team-specific variables that are dependent on multiple users before the data is requested from the ERT. This type of change can also have relevance to improving the log file analysis problems that exist in the current Cloud ERT. If specific user data logs could be selected in the ERT, and potentially grouped by the experimenter, it would assist in solving these problems. For instance, if the experimenter included questions such as "User ID" and "Team Number" in their questions, then perhaps these could be displayed to experimenters for selection as they begin analysis.

Pre-Processing of Experimental Data

Ideally, the ERT would begin the analysis process by populating the available logs on the screen for the experimenter, and instead of listing a title in the form of a string that does not have meaning to the experimenter. The title could pull specific values from the surveys in the file such as Use rid or Participant Number. This could be achieved in two ways: 1) creating standard questions that should be asked of all participants if it is indicated that an experiment is being created (e.g., "Use rid", "Participant Number") or 2) providing experimenters with a way to select specific survey answers that they want displayed as log titles for ease of use. Regardless, selectivity of specific logs and visibility of the participant identification are essential features as the Cloud ERT moves forward.

Considerations and Mockup for a TEAM ERT

While the overall ERT would benefit from a thorough redesign that is focused on usability and functionality, it would be helpful to start from a design that is both helpful for individual and team data. While researchers often take individual data and compile it in a single line of a large spreadsheet that has data from all participants, the design of the output file or features may look differently in a team setup. It might be helpful to have a way to easily determine which individuals were part of the same team, and to group their data close to each other in the output spreadsheet, or to even provide outputs that are specific to individual teams. The design of the ERT interface and functions needs to support multiple types of teams, multiple types of tasks, and different size teams, among other considerations. Therefore, it needs to be highly configurable and include highly generalizable functionality. Then a potential second level could be to represent those measures of team performance that may not necessarily be an aggregate of individual data.

It is important for the interface to elicit the following information from the researcher:

- How are team groupings identified in the data? (e.g., are they an entry in a specific survey field?)
- How are team roles represented in the data? Are team roles unique or duplicated?
- What are the team performance variables and what are the individual performance variables?
- Should data output be separated at the team level or the individual level?

Current Cloud ERT Design

Figure 1 represents the current interface screen of the Cloud ERT. . In the figure it can be seen that each event type requires a check box next to it to be included in the output report. Additionally, there's a single check box for merging each participant's events into a single row. However, all participants are included in the outputs and there's no way from the assigned log file numbers to tell which participant is which. Therefore, it would be helpful to have an earlier screen which allows for definition of the type of study (team or individual), and asks the user to define the above questions that will be used to help parse the data if it is a team study.

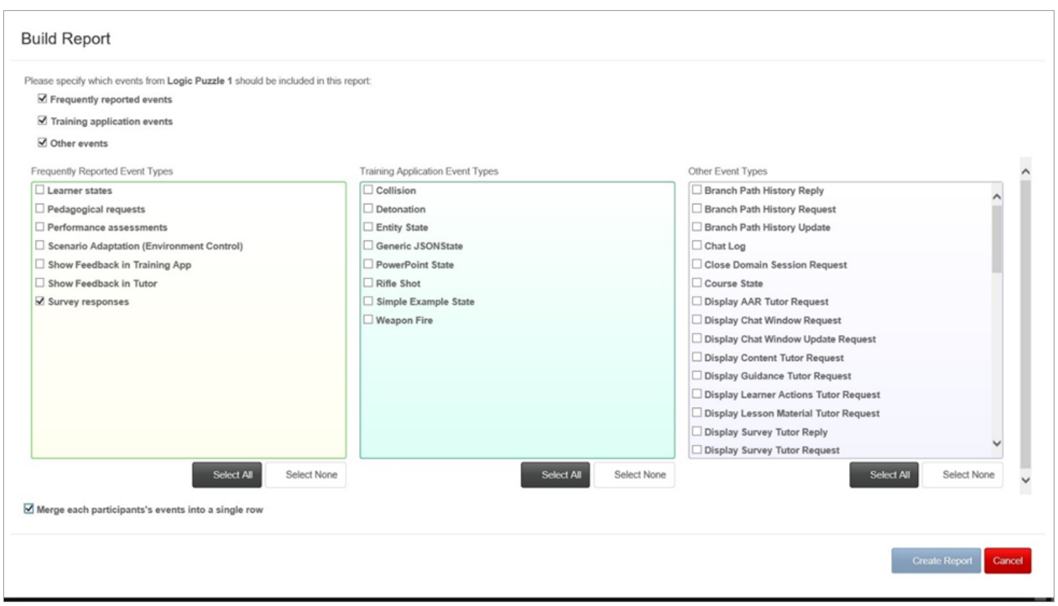


Figure 1. A screenshot example of the current Cloud Based Event Report Tool selection screen.

Mockup for ERT for Teams

A mockup for the ERT for teams can be seen in Figure 2. Attention was paid in the mockup to the design of the initial experimental set up screens to support the experimenter. The mockup is meant to provide an overview on the potential options that are available (per Shneiderman), and the screens are meant to be as simple and clear as possible (per Tuft's graphical excellence). As mentioned above, it would be to the benefit of the experimenter to tell the ERT the relationship among the participants and their relationship to the data. "Import your data" allows the researcher to use data files from various statistical or data management formats. "Define experimental conditions" allows the experimenter to set up relationships. "Create new variables" provides a way to build team-specific variables from existing data. For the purposes of this discussion, we will focus on the process for defining experimental groups.

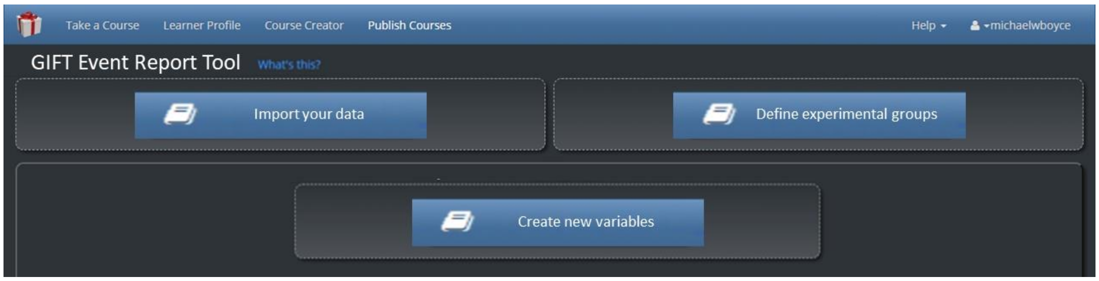


Figure 2. Mockup of ERT for Teams Selection Screen

The screen for defining experimental groups is shown in Figure 3. An experimenter would be able to set the relationship for each participant in terms of team and experimental condition. Participants could be assigned to more than one team, and they could also be assigned to more than one condition (in the event that the participant is going through the experiment more than once). As the experimenter would set the different groupings, GIFT would begin building a visual map of the structure. The assignment of groups and conditions could be modified to fit the researchers need (such as randomization). Once experimenters are finished making the selections, they would then move ahead and review their assignments.

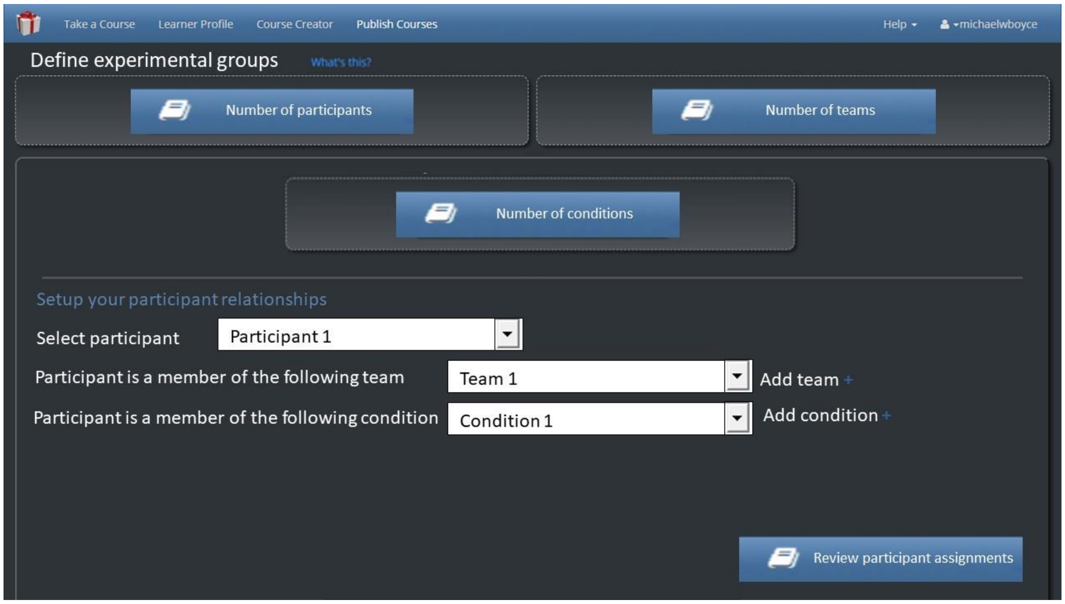


Figure 3. Mockup of the Define Experimental Groups Screen

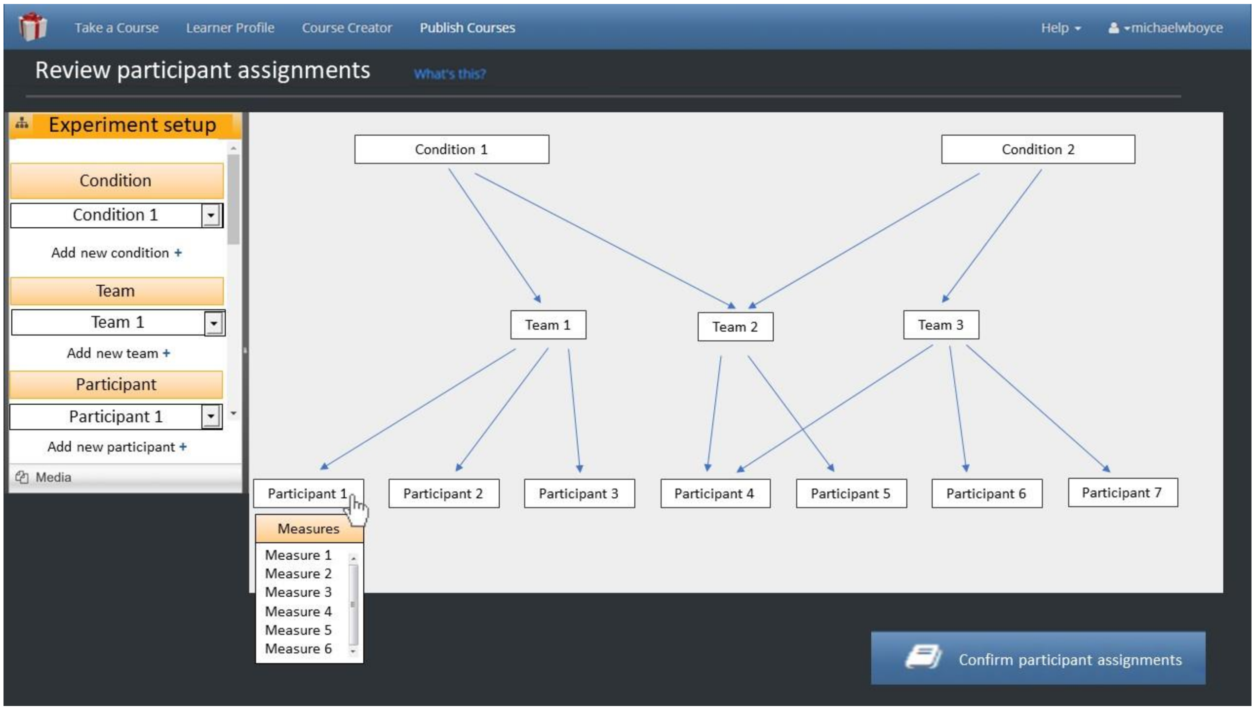


Figure 4. Mockup of the Review Participant Assignments Screen

Then the experimenter would have the chance to review and edit their assignments as necessary, which is shown in Figure 4. This is designed to mimic a flow chart where each relationship is defined by a line connector. Experimenters could add participants, move them between conditions and groups, and have a visual representation of how the experiment is set up. This design could potentially leverage a lot of existing GIFT functionality such as the zoom in and zoom out capability (make the diagram bigger or smaller when there is a need to focus on a specific participant or group of participants) of the authoring tools and the add / delete nodes of GIFT conversation trees. Also relevant here is the experimenter's ability to select each participant and view what measures are associated with them. This measures dropdown could be expanded to create and map measures similarly to experimental groups.

Conclusions and Recommendations for Future Research

This chapter offers suggestions for improving the current data export tools and ERT in GIFT so that they are more efficient and can be used to support team data extraction. The recommendations for updates to the ERT will not only be helpful from a team perspective, but will also provide researchers who are doing non-team research with more power and control over their data which is collected in Cloud GIFT. Improving usability in the ERT's design will ultimately make it more straightforward and result in increased use by the GIFT community. Additionally, allowing for flexibility in the way of defining teams within the ERT can also provide opportunities to leverage the team features for use by instructors in the classroom who are examining subgroups of student answers or in class team assessments. Designing an ITTS framework is a difficult challenge, but by focusing on identifying generalizable elements of team data analysis, and including tools that lessen the burden on the experimenter it is likely to be achieved. Although this is only a mockup, a first step with open questions still to be answered, this chapter could be, in the words of Ben Shneiderman: "A useful starting point for designing advanced graphical user interfaces…"(1996).

References

Bertini, E., Tatu, A., & Keim, D. (2011). Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics, 17*(12), 2203-2212.

Bonner, D., Gilbert, S. B., Dorneich, M. C., Winer, E., Sinatra, A. M., Slavina, A., . . . Holub, J. (2016). The Challenges of Building Intelligent Tutoring Systems for Teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 1981-1985. doi:doi:10.1177/1541931213601451

Chen, C.-h., Härdle, W. K., & Unwin, A. (2007). *Handbook of data visualization*: Springer Science & Business Media.

Dobashi, K. (2016). Development and Trial of Excel Macros for Time Series Cross Section Monitoring of Student Engagement: Analyzing Students' Page Views of Course Materials. *Procedia Computer Science, 96*(Supplement C), 1086-1095. doi:https://doi.org/10.1016/j.procs.2016.08.133

Dobashi, K. (2017). Automatic data integration from Moodle course logs to pivot tables for time series cross section analysis. *Procedia Computer Science, 112*(Supplement C), 1835-1844. doi:https://doi.org/10.1016/j.procs.2017.08.222

Dorsey, D., Russell, S., Keil, C., Campbell, G., Van Buskirk, W., & Schuck, P. (2009). Measuring teams in action: Automated performance measurement and feedback in simulation-based training. *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches*, 351-381.

Gilbert, S. B., Dorneich, M., Walton, J., & Winer, E. (in press). Five Lenses on Team Tutor Challenges: A Multidisciplinary Approach. In J. Johnston, R. Sottilare, A. M. Sinatra, & C. S. Burke (Eds.), *Building Intelligent Tutoring Systems for Teams: What Matters*. Bingley, UK: Emerald.

Gilbert, S. B., Sinatra, A. M., MacAllister, A., Kohl, A., Winer, E., Dorneich, M. C., . . . Ouverson, K. M. (2017). Analyzing Team Training Data: Aspirations for a GIFT Data Analytics Engine.

MacAllister, A., Kohl, A., Gilbert, S. B., Winer, E., Dorneich, M., Bonner, D., & Slavina, A. (2017). Analysis of Team Tutoring Training Data: MODSIM World.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods, 44*(2), 314-324.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*: Psychology Software Incorporated.

Shneiderman, B. (1996). *The eyes have it: A task by data type taxonomy for information visualizations.* Paper presented at the Visual Languages, 1996. Proceedings., IEEE Symposium on.

Shneiderman, B., & Plaisant, C. (2005). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Boston, MA: Addison-Wesley.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT).* Orlando, FL: Army Research Laboratory.

Sottilare, R. A., Brawner, K. W., Sinatra, A. M., & Johnston, J. H. (2017). An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT).

Sottilare, R. A., Burke, C. S., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2017). Designing Adaptive Instruction for Teams: a Meta-Analysis. *International Journal of Artificial Intelligence in Education*. doi:10.1007/s40593-017-0146-z

Tufte, E. R. (1983). *The visual display of quantitative data*.

Vanlehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.

Walton, J., Gilbert, S. B., Winer, E., Dorneich, M. C., & Bonner, D. (2015). *Evaluating Distributed Teams with the Team Multiple Errands Test*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL.

Zachary, W., Cannon-Bowers, J. A., Bilazarian, P., Krecker, D. K., Lardieri, P. J., & Burns, J. (1998). The Advanced Embedded Training System (AETS): An intelligent embedded tutoring system for tactical team training. *International Journal of Artificial Intelligence in Education (IJAIED), 10*, 257-277.

Biographies

Editors

Dr. Arthur C. Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is a Senior Research Fellow in the Department of Education at the University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. Dr. Graesser's primary research interests are in cognitive science, discourse processing, and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, education, memory, emotions, computational linguistics, artificial intelligence, human-computer interaction, and learning technologies with animated conversational agents. Dr. Graesser served as editor of the journal Discourse Processes (1996–2005) and Journal of Educational Psychology (2009-2014) and as president of the Empirical Studies of Literature, Art, and Media (1989-1992), the Society for Text and Discourse (2007-2010), International Society for Artificial Intelligence in Education (2007-2009), and the FABBS Foundation (2012-13). He has published over 500 articles in journals, books, and conference proceedings. Dr. Graesser and his colleagues have designed, developed, and tested software that integrates psychological sciences with learning, language, and discourse technologies, including AutoTutor, AutoTutor-Lite, MetaTutor, GuruTutor, DeepTutor, HURA Advisor, SEEK Web Tutor, Operation ARIES!, iSTART, Writing-Pal, AutoCommunicator, Point & Query, Question Understanding Aid (QUAID), QUEST, & Coh-Metrix.

In 2010, Dr. Graesser received the Distinguished Scientific Contribution Award (Society for Text and Discourse) and in 2011 he received the Distinguished Contributions of Applications of Psychology to Education and Training Award (American Psychological Association). In 2012, Dr. Graesser received the first Presidential Award for Lifetime Achievement in Research from the University of Memphis. This award is the University's highest level of research recognition given to its faculty. It was established as part of the University's Centennial fundraising campaign in order to recognize the vital role and impact of research at the University of Memphis. He served as Chair of the Framework group in PISA Collaborative Problem Solving 2015. In 2018 he received the Harold W. McGraw, Jr. Prize in Education.

Dr. Xiangen Hu is a professor at the University of Memphis in the Department of Psychology-Institute for Intelligent Systems, the Department of Computer Science and the Department of Electrical Engineering. Currently, he is also the Dean of School of Psychology at Central China Normal University (CCNU). He joined the IIS at the University of Memphis in 1993 and has played significant roles in all past IIS Intelligent Tutoring Systems (ITS) related projects. He was Director of the Advanced Distributed Learning Center for Intelligent Tutoring Systems Research and Development (ADL-C-ITS-R&D) at the IIS and is now the Director of ADL Partnership Lab at the University of Memphis. Dr. Hu's primary research areas include Mathematical Psychology, Research Design and Statistics, and Cognitive Psychology. More specific research interests include General Processing Tree (GPT) models, categorical data analysis, knowledge representation, computerized tutoring, and advanced distributed learning. Dr. Hu either currently receives or has had funding for the above research from the U.S. National Science Foundation (NSF), U.S. Institute of Education Sciences (IES), ADL of the U.S. Department of Defense (DoD), U.S. Army Medical Research

Acquisition Activity (USAMRAA), U.S. Army Research Laboratories (ARL), U.S. Office of Naval Research (ONR), the University of Memphis, and CCNU. He is a member of the PI Millionaire club at The University of Memphis for his successful efforts to attract multimillion-dollar funds from federal agencies as Principal Investigator (PI). He helped organize the annual meeting for the Society for Mathematical Psychology (2005), the E-Learning Summit meetings in 2005 and 2007, SCIP (2008-2009), and ALTTAI Consortium meeting in 2016 and 2017 and was the local chair for the 2017 AIED and general Co-Chair for 2017 EDM meeting. He currently serves as a technical editor for the PLOS One journal. In the past 10 years, Dr. Hu has been PI for over $5M in federal grants. He is responsible for creating several versions of ITS based on a previously funded project, including AutoTutor Lite (currently called Sharable Knowledge Object (SKO)) which was funded by a previous ADL grant. Dr. Hu led a team at the UM that won the ONR STEM Grand Challenge (2014).

Dr. Anne M. Sinatra is a Research Psychologist and Adaptive Tutoring Scientist in the Learning in Intelligent Tutoring Environments (LITE) Lab within the U.S. Army Research Laboratory – Human Research and Engineering Directorate (ARL-HRED). She leads the Team Tutoring research vector within the LITE lab, and works on the Generalized Intelligent Framework for Tutoring project. The focus of her research is in team tutoring, cognitive psychology, human factors psychology, and adaptive tutoring. She has specific interest in how information relating to the self and about those that one is familiar with can aid in memory, recall, and tutoring. Her work has been published in the journal Interaction Studies, the Journal of Artificial Intelligence in Education and in the conference proceedings of the Human Factors and Ergonomics Society and Human-Computer Interaction International. Dr. Sinatra received her Ph.D. and M.A. in Applied Experimental and Human Factors Psychology, as well as her B.S. in Psychology from the University of Central Florida.

Dr. Robert A. Sottilare leads adaptive training research at the US Army Research Laboratory where the focus of his research is automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems (ITSs). He is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT), an open source, AI-based adaptive instructional architecture. He is the lead editor for the Design Recommendations for Intelligent Tutoring Systems book series and the founding chair of the GIFT Users Symposia. Dr. Sottilare has authored over 165 technical publications. He is a program committee member and frequent speaker at the Defense & Homeland Security Simulation, Augmented Cognition, and AI in Education conferences. Dr. Sottilare is a member of the AI in Education Society, the Florida AI Research Society, the IEEE Standards Association, and the American Education Research Association. He is a faculty scholar and adjunct professor at the University of Central Florida where he teaches a graduate level course in ITS design. Dr. Sottilare is also a frequent lecturer at the United States Military Academy (USMA) where he teaches a senior level colloquium on adaptive training and ITS design. He has a long history of participation in international scientific fora including NATO and the Technical Cooperation Program. Dr. Sottilare is the recipient of the Army Achievement Medal for Civilian Service (2008), and two lifetime achievement awards in Modeling & Simulation: US Army RDECOM (2012) and National Training & Simulation Association (2015).

Authors

Dr. Michael W. Boyce is a research psychologist with ARL's adaptive training research program. For the past 3 years his emphasis has been in using technologies like GIFT to accurately assess learner knowledge and performance. Located at the United States Military Academy at West Point, his goal is to better inform the research progress of GIFT through interactions with a military student population. He received his Ph.D. in Applied/Experimental Human Factors Psychology from the University of Central Florida in 2014.

Dr. Keith W. Brawner is a researcher for the Learning in Intelligent Tutoring Environments (LITE) Lab within the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED). He has 12 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. His current research is in machine learning, cognitive architectures, learning technologies, and ITS architecture. He manages and advises research in adaptive training and architectural programs towards next-generation training.

Dr. C. Shawn Burke is a Professor (Research) and Director of the TRACE lab at the Institute for Simulation and Training, University of Central Florida. Her expertise includes teams and their leadership, team adaptability, team training, measurement, evaluation, and team effectiveness. Dr. Burke has published over 90 journal articles and book chapters related to the above topics and has presented/had work accepted at over 200 peer-reviewed conferences. Dr. Burke has received funding from the following agencies: Army Research Institute, Army Research Laboratory, Office of Naval Research, Army Research Office, National Science Foundation, Gulf Oil Research Program, and NASA. Most recently, Dr. Burke has been funded by NASA to investigate issues related to team leadership, team roles, and cultural diversity within the context of teams operating in isolated, confined environments (e.g., long duration space exploration) with an eye towards fostering resiliency within such teams. All of the above work is conducted with an interest in team leadership and the training of teams operating in complex environments. She has also recently been funded by the Army Research Institute to investigate issues related to leader and team resilience.

Dr. Burke earned her doctorate in Industrial/Organizational Psychology from George Mason University and is an Associate Editor for the Journal of Trust Research and Consulting Editor for the Journal of Business and Psychology. She also serves as an ad-hoc reviewer for several journals, including: Leadership Quarterly, Journal of Applied Psychology, Human Factors, Military Psychology, Small Group Research, and Human Resource Management. She has co-edited books on adaptability and advances in team effectiveness research. She is currently co-editing a book on intelligent tutoring for teams.

Zhiqiang Cai is a Research Assistant Professor with the Institute for Intelligent Systems at the University of Memphis. He has a M.S. degree in mathematics received in 1985 from Huazhong University of Science and Technology, P. R. China. After 15 years of teaching mathematics in colleges, he has worked in the field of natural language processing and intelligent systems. He is the chief software designer and developer of Coh-Metrix, OperationAries, CSAL AutoTutor and many other text analysis tools and conversational tutoring systems. He has co-authored over 70 publications**.**

Dr. Alan S. Carlin is a Principal Research Engineer at Aptima, within the Learning and Training Systems (LTS) division. He is interested in research and technologies related to artificial intelligence and machine learning. These include problems of decision making under uncertainty, communication between members of a team, generative models, and meta-reasoning for decision-makers. His publications include works on Decentralized Partially Observable Markov Decision Processes, communication under uncertainty, game theory, and distributed meta-reasoning in uncertain environments. At Aptima, he has developed intelligent training systems, aircraft pilot alert systems, and machine learning systems. Dr. Carlin received a Ph.D. in Computer Science from the University of Massachusetts, an M.S. in Computer Science from Tufts University, and a dual B.A. in Computer Science and Psychology from Cornell University. Prior to joining Aptima, Dr. Carlin was Associate Staff at MIT Lincoln Laboratory. As part of his M.S., he also completed the MIT Lincoln Scholar's Program.

Jody L. Cockroft is a Research Specialist at the University of Memphis in the Institute for Intelligent Systems. Prior to joining the University of Memphis, she was with the University of Tennessee Health Science Center in Memphis where she was involved in with various clinical trials and bench research for over twenty years. She earned her A.A. from the University of Tampa and her B.S. from the University of

Memphis. She has been working with Xiangen Hu and Arthur Graesser for the past two years on the Army Research Laboratory project on the Generalized Intelligent Framework for Tutoring (GIFT) and the Advanced Distributed Learning (ADL) Academy projects and the Advanced Learning Theories, Technologies, Applications and Impacts (ALTTAI) Consortium efforts.

Dr. Christopher Dede, also known as Chris, is the Timothy E. Wirth Professor in Learning Technologies at Harvard's Graduate School of Education. Mr. Dede is serving as the lead Time To Know Research Advisory Board member coordinating the digital teaching platform workshop at Harvard. Mr. Dede serves as a Member of the National Advisory Council of Ready to Learn Partnership. He serves as a Member of the Advisory Board of Critical Links, Inc. He serves as a Member of Research Advisory Board at Time To Know, Inc. He serves on Advisory Boards and Commissions for PBS TeacherLine, the Partnership for 21st Century Skills, the Pittsburgh Science of Learning Center, and several federal research grants. His fields of scholarship include emerging technologies, policy, and leadership. His funded research includes three grants from NSF and the US Department of Education Institute of Education Sciences to explore immersive and semi-immersive simulations as a means of student engagement, learning, and assessment. Mr. Dede served as a member of the National Academy of Sciences Committee on Foundations of Educational and Psychological Assessment and a member of the 2010 National Educational Technology Plan Technical Working Group. He has authored many books and research papers on the subject. Mr. Dede's primary area of interest is the role of new information technologies in knowledge creation, sharing and learning. He was a member of the U.S. Department of Education's Expert Panel on Technology. He served as chair of the Learning & Teaching area at HGSE. He is on the International Steering Committee for the “Second International Technology in Education Study” spanning approximately thirty countries. Mr. Dede is widely recognized as a global leader in the development of “technology in education” programs. His co-edited book, Scaling Up Success: Lessons Learned from Technology-based Educational Improvement, was published by Jossey-Bass in 2005. A second volume he edited, Online Professional Development for Teachers: Emerging Models and Methods, was published by the Harvard Education Press in 2006. In 2007, Mr. Dede was honored by Harvard University as an outstanding teacher.

Dr. Sidney D’Mello (PhD in Computer Science) is an Associate Professor in the Institute of Cognitive Science and Department of Computer Science at the University of Colorado Boulder. He is interested in the dynamic interplay between cognition and emotion while individuals and groups engage in complex real-world tasks. He applies insights gleaned from this basic research program to develop intelligent technologies that help people achieve to their fullest potential by coordinating what they think and feel with what they know and do. D’Mello has co-edited six books and published over 220 journal papers, book chapters, and conference proceedings (13 of these have received awards). His work has been funded by numerous grants and he serves(d) as associate editor for four journals, on the editorial boards for six others, and has played leadership roles in several professional organizations. https://www.colorado.edu/ics/sidney-dmello.

Dr. Michael C. Dorneich is an Associate Professor of Industrial and Manufacturing Systems Engineering at Iowa State University. He is a member of the Human-Computer Interaction Graduate Program, and has a courtesy appointment in Aerospace Engineering. He graduated from the University of Illinois at Urbana-Champaign with a Ph.D. in Industrial Engineering in the Human Factors Program. His research interests focus on creating joint human-machine systems that enable people to be effective in the complex and often stressful environments found in aviation, military, robotic, and space applications. He specializes in adaptive systems which can provide assistance tailored to the user’s current cognitive state, situation, and environment. Adaptive systems are becoming more necessary as intelligent assistants are spreading into every aspect of work, education, and home life. His recent work looks at the development of intelligent team tutoring systems, and the development of human-autonomy team frameworks. Prior to joining the faculty at Iowa State University, he worked in industry researching adaptive system design and human factors in a variety of domains. He was a visiting scientist at NASA Ames Research Center in 2004. He holds 28 US and international patents. He has authored over 150 professional, peer-reviewed papers, and is currently an Associate Editor for the Journal of IEEE Transactions of Human-Machine Systems.

Dr. Nia Dowell is a postdoctoral research fellow in the School of Information and Digital Innovation Greenhouse at the University of Michigan. She completed her Ph.D. at the Institute for Intelligent Systems in the University of Memphis. Her primary interests are in cognitive psychology, discourse processing, and learning sciences. Her research focuses on using language and discourse to uncover the dynamics of socially significant, cognitive, and affective processes. She is currently applying computational techniques to model discourse and social dynamics in a variety of learning environments including intelligent tutoring systems (ITSs), small group computer-mediated collaborative learning environments, and massive open online courses (MOOCs). Her current research highlights the practical applications of computational discourse science in the clinical, political and social sciences areas.

Jing Du is a Ph.D. candidate at Beijing Normal University under the direction of Dr. Ronghuai Huang. Her research focuses on improving individualized instruction in tutoring systems using data driven methods and the mining of educational data. She also interested in learning space design, computer supported collaborative learning. She earned her Master's degree from Central China Normal University, China.

Dr. Jeremiah T. Folsom-Kovarik is a Lead Scientist with Soar Technology, Inc. in the Intelligent Training research group. His research focuses on making intelligent automation capable and robust with advanced planning, user modeling, and contextual interpretation approaches to make technology meet individuals' needs. In learning systems, J.T. researches computer representations and algorithms that make adaptive training more effective via capabilities such as automation of instructor tasks that previously required laborious attention or technical knowledge engineering, and planning ahead under uncertainty in order to construct learning sequences that are more effective than individual choices. A natural continuation of his work is research into machine learning and planning approaches that extend across multiple learning systems to efficiently build a cohesive learning experience.

Dr. Peter Foltz is Vice President in Pearson's Advanced Computing and Data Sciences Laboratory and Research Professor at the University of Colorado's Institute of Cognitive Science. His work covers discourse processing, reading comprehension and writing skills, 21st Century skills learning, large-scale data analytics, artificial intelligence, and uses of machine learning and natural language processing for educational and clinical assessments. Much of his work has focused on techniques for automatically analyzing the meaning of language through writing and speaking. The approaches are used for assessing abilities, for providing feedback, and for understanding underlying cognitive mechanisms in the brain. The methods he has pioneered are used by millions of student annually to improve student achievement, expand student access, and make learning materials more affordable. Peter has served as the content lead for the framework development for several Organisation of Economic Cooperation and Development's (OECD) Programme for International Student Assessment (PISA) assessments, including the 2018 Reading Literacy assessment, the 2015 assessment of Collaborative Problem Solving, and a new assessment of reading literacy for developing countries. He has served as guest editor for a number of journals including *International Journal of AI in Education* and *Discourse Processes.* Peter has authored more than 100 journal articles, book chapters, conference papers, as well as several patents. He previously worked as a professor at New Mexico State University, and as a researcher at Bell Communications Research, the Learning Research and Development Center at the University of Pittsburgh and Yale University. Dr. Foltz holds doctorate and master's degrees in Cognitive Psychology from the University of Colorado, Boulder, and a bachelor's degree from Lehigh University.

Dr. Stephen Gilbert is an associate director of the Virtual Reality Applications Center and assistant professor of Industrial and Manufacturing Systems Engineering at Iowa State University. His research interests focus on technology to advance cognition, including interface design, intelligent tutoring systems, and cognitive engineering. He is a member of IEEE and ACM and works closely with industry and federal agencies on research contracts. From 2015-2018 he led a project supporting the U.S. Army Research Laboratory STTC in future training technologies for teams.

Dr. Benjamin Goldberg is a member of the LITE Lab at ARL-HRED in Orlando, FL. He has been conducting research in the M&S community for the past eight years with a focus on adaptive learning in simulation-based environments and how to leverage AI tools and methods to create personalized learning experiences. Currently, he is the LITE Lab's lead scientist on instructional management research within adaptive training environments and is a co-creator of GIFT. He is a PhD graduate from UCF in the program of M&S. His work has been published across several well-known conferences, with recent contributions to the Human Factors and Ergonomics Society (HFES), Artificial Intelligence in Education and Intelligent Tutoring Systems proceedings. He has also recently contributed to the Journal Computers in Human Behavior and Journal of Cognitive Technology.

Dr. Jamie Gorman received his PhD in Psychology from New Mexico State University and is an associate professor in Engineering Psychology at Georgia Tech. Dr. Gorman's research focuses on human performance in complex social and technological settings, including sports, medicine, and aviation. In particular, his research focuses on understanding and modeling human-systems interactions using dynamical systems theory. Research in Dr. Gorman's Systems Psychology Lab seeks to understand and enhance human performance using a variety of methodological approaches, including communication analysis, kinematics, physiological, and neural approaches. Dr. Gorman has over 50 refereed articles, proceedings, and book chapters, and his research has been funded by ONR, NSF, DARPA, and JUMP. Dr. Gorman is a member of the Human Factors and Ergonomics Society (HFES) and serves on the editorial board of the journal Human Factors. In 2011 he and his coauthors received the Jerome H. Ely award from HFES for the best paper published in the 2010 volume of Human Factors.

Dr. Tina Grotzer is a cognitive scientist whose research identifies ways in which understandings about the nature of causality impact our ability to understand complexity in our world. Her work has important implications for how we deal with global and ecological issues and is concerned with the environmental injustices that result from our inability to reason well about complexity. She is a member of the faculty at the Harvard Graduate School of Education and at the Center for Health and the Global Environment. She received a Career Award from NSF in 2009 and a Presidential Career Award for Scientists and Engineers in 2011. She is the author of *Learning Causality in a Complex World: Understandings of Consequence* (2012) as well as numerous articles, book chapters, and resources for teachers.

Dr. Mark Guadagnoli is a Professor of Neuroscience and Neurology at the UNLV School of Medicine where he serves as the Director of Learning and Performance as well as the Associate Dean of Faculty Affairs. He received his undergraduate and M.S. degrees at Texas A&M University, and Ph.D. at Auburn University in Human Performance/Cognitive Psychology. He has worked in industry and academia for over two decades. He specializes in optimizing performance, communication, leadership, and learning. He has received numerous awards for this work and has published more than 100 articles and abstracts as well as several books. His primary line of research is related to the Challenge Point Framework, which has been used to help corporate executives, medical professionals, and others who compete in high performance situations. His model of learning shows that appropriate short-term challenges result in long-term and stress resistant learning.

Dr. Andrew J. Hampton studied at the University of Central Florida and the Burnett Honors College on a full National Merit scholarship. After graduating with a B.S. in psychology and a minor in cognitive science, he entered the Wright State Human Factors/Industrial Organizational doctoral program in the fall of 2011 where he worked as a graduate research and teaching assistant and as a graduate teaching assistant, while also co-founding the Professional Development Group, a graduate student organization developed to encourage independent initiatives. While at Wright State, he earned his Master's degree in 2013, the program's Graduate Student Excellence Award in 2014, and his PhD in 2018. In 2016, he accepted a position as a researcher under Art Graesser at the Institute for Intelligent Systems, within the University of Memphis. There, in addition to his research duties, he has taken on responsibilities as coordinator for a grant with the Office of Naval Research, and project manager for ElectronixTutor, a platform integrating several intelligent tutoring systems in a common interface. His work won the 2016 Human Factors Prize for Excellence in Human Factors/Ergonomics Research, focusing on Big Data analytics. Andrew's research interests include technologically mediated communication, psycholinguistics, semiotics, intelligent tutoring systems, artificial intelligence, and political psychology.

Dr. Jiangang Hao is a senior research scientist at the Psychometrics, Statistics and Data Science (PSDS) division at Educational Testing Service. His current research centers on response process data modeling, collaborative problem solving, game and simulation-based assessment, educational data mining & analytics, and automated scoring. Jiangang is leading the computational psychometrics subinitiative of the FASP initiative at ETS in the FY 2017, and has been co-leading the infrastructure subinitiative of the game, simulation and collaboration initiative from 2014 to 2016. He is the principal investigator of several research projects at ETS for designing simulation-based assessments, web-based platform for collaborative assessments and data analytics packages for game-based assessments. Jiangang obtained his Ph.D. in Physics and M.A. in Statistics from the University of Michigan. Prior to joining in ETS, Jiangang worked on modeling and mining Terabyte-scale data in astrophysics at Fermi National Accelerator Laboratory. His work has been reported by leading technology media, such as the Wired and MIT Technology Review. He has published over 50 peer-reviewed papers, with over 3500 total citations and h-index of 29.

Michael Hoffman is a senior software engineer at Dignitas Technologies and the technical lead on the Generalized Intelligent Framework for Tutoring (GIFT) project. He has been responsible for ensuring that the development of GIFT meets the evolving customer requirements in addition to supporting both intelligent tutoring for computer based training and intelligent tutoring technology research of the growing user community. Recently this has involved altering the GIFT architecture from an individual based ITS to one that supports team tutoring. Michael manages and contributes support for the GIFT community through various mediums including the GIFT portal (www.GIFTTutoring.org), annual GIFT Symposium conferences and technical exchanges with ARL and their contractors. In addition he utilizes his expertise in integrating third party capabilities such as software and hardware systems to enable other organizations to integrate GIFT into their training solutions.

Dr. Ronghuai Huang is a professor of Beijing Normal University. He currently works as Co-Dean of the Smart Learning Institute of Beijing Normal University, Director of UNESCO International Rural Educational and Training Centre, Director of National Engineering Lab for Cyberlearning Intelligent Technology, Director of Digital Learning and Public Education Service Center as well as Director of Beijing Key Laboratory of Educational Technology. He also serves as Deputy Director of Professional Teaching and Guiding Committee for Educational Technology, member of expert group for educational informatization in the Ministry of Education, Vice-Chairman of China Association for Education and Technology as well as Chairman of International Smart Learning Environment Association. He has taken charge of some key projects, including research on the development of international educational informatization, strategic research on information-driven modernization for schools, IT curriculum standard revision for senior schools, etc.

Dr. Joan Johnston has been in U.S. government civilian service for 33 years. She is currently a Senior Scientist with the U.S. Army Research Laboratory (ARL) where she conducts research on training effectiveness and team training. She has produced over 80 professional papers, peer reviewed journal articles, and presentations. In 2016, she received the U.S. Army Civilian Service Achievement Medal for an innovative team training strategy to improve decision making under stress. Prior to her move to ARL, she was a senior research psychologist with the Naval Air Warfare Center Training Systems Division and a NAVAIR Fellow. For her innovations in Navy research, Dr. Johnston was awarded the ONR Dr. Arthur E. Bisson Prize for Naval Technology Achievement (2000) and the Society for Industrial and Organizational Psychology M. Scott Myers Award for Applied Research in the Workplace (2001). She received her B.S in Biology with Honors from Rutgers University, and an M.A. and Ph.D. in Industrial and Organizational Psychology from the University of South Florida.

Dr. Amy Kamarainen is a senior research manager and principal investigator at the Harvard Graduate School of Education where she collaboratively manages grant-based education research projects, most recently the AR Girls, EcoXPT, EcoMOD and EcoMOBILE projects. Amy is an ecosystem scientist who applies her understanding of ecosystems science and education research to the design and evaluation of technologies that support science learning inside and outside of the classroom. Amy's professional interests concern the application of these technologies to creative spaces like STEM learning, Citizen Science, and place-based education. The Ecological Society of America named Amy an Ecology Education Scholar in 2011.

Dr. Jong W. Kim is a postdoctoral researcher and adaptive tutoring scientist US Army Research Laboratory, Orlando, FL. Kim received his PhD in Industrial Engineering from Pennsylvania State University, and MS in Industrial Engineering from University of Central Florida. His research interests lie in the area of Cognitive Science and Adaptive Instructional Science. Particularly, Kim is interested in testing cognitive theories of learning (and forgetting) for the development of adaptive instructional systems. Previously, Kim has developed a theory of skill learning and forgetting (D2P: Declarative to Procedural) that is currently being applied to implement a series of intelligent tutoring systems for the Navy. After joining the GIFT team at ARL, Kim started to investigate an adaptive instructional system in a psychomotor domain that can be run beyond the desktop environment, pursuing maximizing training effectiveness in a simulated and range-based training.

Dr. H. Chad Lane is an Associate Professor of Educational Psychology and Informatics at the University of Illinois, Urbana-Champaign. Prof. Lane's research focuses on the design, use, and impacts of intelligent technologies for learning and behavior change. This work involves blending techniques from the entertainment industry (that foster engagement) with those from artificial intelligence and intelligent tutoring systems (that promote learning), as well as running studies to better understand whether and how the resulting learning experiences impact learners. He has over 70 publications and has hands-on experiences in informal and formal learning contexts. He earned his Ph.D. in Computer Science from the University of Pittsburgh in 2004. http://hchadlane.net

Dr. Shari Metcalf is a Senior Researcher at the Harvard Graduate School of Education and Project Director for EcoLearn. Her research centers on the design and evaluation of computer–based tools for learning through scaffolded environments in which students can engage in authentic, project-based, constructivist activities.

Dr. Benjamin D. Nye is the Director of Learning Sciences at University of Southern California Institute for Creative Technologies. Ben's major research interest is to identify best-practices in advanced learning technology, particularly for frontiers such as distributed learning technologies (e.g., cloud-based, device-agnostic) and socially-situated learning (e.g., face-to-face mobile use). His research interests include modular intelligent tutoring system (ITS) designs, modeling social learning and memes, cognitive agents, and educational tools for the developing world and low-resource/low-income contexts. He received his Ph.D. in Systems Engineering from the University of Pennsylvania in 2011.

Dr. Andrew Olney presently serves as Associate Professor in both the Institute for Intelligent Systems <iis.memphis.edu> and Department of Psychology <psyc.memphis.edu>. Dr. Olney received a B.A. in Linguistics with Cognitive Science from University College London in 1998, an M.S. in Evolutionary and Adaptive Systems from the University of Sussex in 2001, and a Ph.D. in Computer Science from the University of Memphis in 2006. His primary research interests are in natural language interfaces. Specific interests include vector space models, dialogue systems, unsupervised grammar induction, robotics, and intelligent tutoring systems.

Alec Ostrander is a Ph.D. student in Industrial Engineering and Human-Computer Interaction at Iowa State University's Virtual Reality Applications Center. He is currently exploring how intelligent systems can be designed to leverage principles and ideas from the teamwork literature to create effective human-agent teams.

Dr. Samantha (Baard) Perry is a Scientist at Aptima, Inc. has a decade of applied research experience with the Army, Air Force, NASA and EMTs. She has expertise in adaptation, training design and evaluation, survey development and implementation, and, in particular, the unobtrusive measurement of team processes, states and performance. Her work incorporates a multilevel approach to examining team process dynamics, targeted at creating more efficient and effective training, all with the goal of maximizing human performance. Dr. Perry holds a Ph.D. and M.A. in Organizational Psychology from Michigan State University and a B.A. in Psychology from George Mason University.

Dr. A. R. Ruis is a learning sciences researcher at the Wisconsin Center for Education Research and a historian at the University of Wisconsin–Madison. His work is primarily in the areas of food and nutrition studies, medical and surgical education, and learning analytics. He is the author of *Eating to Learn, Learning to Eat: The Origins of School Lunch in the United States*.

Dr. Vasile Rus is a Professor of Computer Science with a joint appointment in the Institute for Intelligent Systems (IIS) at The University of Memphis. He also serves as Director of the Data Science Center at The University of Memphis. Dr. Rus' areas of expertise are computational linguistics, artificial intelligence, software engineering, and computer science in general. His research areas of interest include question answering and asking, dialogue-based intelligent tutoring systems (ITSs), assessment of open learner answers, knowledge representation and reasoning, information retrieval, and machine learning. For more than a decade, Dr. Rus has been heavily involved in various dialogue-based ITS projects including systems that tutor students on science topics (DeepTutor), reading strategies (iSTART), writing strategies (W-Pal), and metacognitive skills (MetaTutor). Dr. Rus coedited three books, received several Best Paper Awards, and authored more than 100 publications in top, peer-reviewed international conferences and journals.

Dr. David Williamson Shaffer is the Vilas Distinguished Professor of Learning Sciences at the University of Wisconsin–Madison in the Department of Educational Psychology, the Obel Professor of Learning Analytics at Aalborg University in Copenhagen, and a Data Philosopher at the Wisconsin Center for Education Research. He studies how to develop and assess complex and collaborative thinking skills, and is the author of *How Computer Games Help Children Learn* and *Quantitative Ethnography*.

Genghu Shi is a Ph.D. student in Experimental Psychology and also pursuing a master's degree in Statistics in University of Memphis. He works as a research assistant for Center for the Study of Adult Literacy in the Institute for Intelligent Systems. His research interests lie in Adult Literacy and the characteristics of conversations in intelligent tutoring systems. He is also interested in Data Science and Machine Learning. He earned a master's degree in Cognitive Psychology from Central China Normal University and a Bachelor's Degree in Computer Science from China University of Geoscience (Wuhan).

Paul Shorter is an Operations Research Analyst with the Simulation Training and Technology Center (STTC). He has B.S. degree in mathematics received in 1989 from Virginia Commonwealth University, Richmond, VA. He has worked for the US Department of Army since 1989 in various areas including systems analysis and evaluation of major weapon systems, logistics modeling and simulation applications, field research relating to dismounted Soldier activity. His current assignments at STTC are in the areas of adaptive training and virtual reality applications.

Dr. Ron Stevens received his Ph.D. In Microbiology and Molecular Genetics from Harvard University. He is currently a Professor (Emeritus), UCLA School of Medicine, a member of the UCLA Brain Research Institute, and the CEO of The Learning Chameleon, Inc. He has published over 200 peer-reviewed studies in medical research, medical problem solving and team neurodynamics and is the author of three patents. Dr. Stevens early studies focused on the cellular and molecular defects in autoimmunity and immune deficiency, and included pioneering studies on Acquired Immune Deficiency Syndrome (before it was termed AIDS). Subsequently he developed of the technology-based UCLA-IMMEX™ problem-authoring and solving project. Funded in part by the National Science Foundation and the Department of Education, this project has engaged hundreds of teachers and tens of thousands of students, from middle school through medical school in scientific problem solving activities. This project was recognized by the Smithsonian Institute in its 'Search for New Hero's' program, by Zenith Corporation's 'Masters of Innovation' program and by the UCLA Medical School's 'Excellence in Education' award. The software has been licensed for commercial development. In his role as the CEO of The Learning Chameleon, Inc., his recent research has focused on using EEG-derived measures to investigate team neurodynamics in the complex and real-world settings of military and healthcare training. Dr. Stevens' research was the first to demonstrate the presence of neurodynamic organizations in teams. These are states of neurodynamic persistence that team members enter into when their rhythm can no longer support the complexity of the task and they must expend energy to re-organize into structures that better minimize the 'surprise' in the environment. These studies are leading to quantitative teamwork models showing how teams cognitively organize in response to environmental and task changes, and are paving the way for future real-time individual and team adaptive learning. The studies have been recognized by multiple awards from the Human Factors and Ergonomics Society's Augmented Cognition Group including the Admiral Leland Kollmorgen 'Spirit of Innovation Award.' This research has been supported by the National Science Foundation, DARPA, Office of Naval Research, Department of Education, as well as corporate and private funding. More information may be accessed at: www.teamneurodynamics.com.

Dr. Minhong Wang is associate professor and director of the Laboratory for Knowledge Management & E-Learning at The University of Hong Kong. She is also Eastern Scholar Chair Professor at East China Normal University. Her areas of expertise include e-learning design and evaluation, knowledge management, inquiry learning and problem solving, visualization for deeper learning, medical education, workplace e-learning, artificial intelligence, and business process management. She is editor-in-chief of *Knowledge Management & E-Learning*, and associate editor of *Information & Management*, in addition to guest-editor of *Educational Technology & Society*, and *Computers in Human Behavior*. She was previously a visiting scholar at Harvard Graduate School of Education, University of Cambridge, and MIT Sloan School of Management. More details can be found at http://web.edu.hku.hk/staff/academic/magwang.

Dr. Junfeng Yang is distinguished professor in Hangzhou Normal University, and he is the dean of department of Educational Technology in Hangzhou Normal University. He received his PhD from Beijing Normal University in 2014. His research interests include smart learning environments, blended synchronous cyber classroom, and the digital generation of learners.

Dr. Wayne Zachary is CEO and Managing Partner of Starship Health Technologies. He holds a Ph.D. in Social and Cognitive Anthropology and M.S. in Computer Science from Temple University. He has integrated this multidisciplinary background in a career-long thread of research, technology development, and entrepreneurship focused on cognitive modeling of individuals and organizations as applied to of decision support systems and training systems. Since founding Starship in 2012, he has focused on challenging and significant issues in human-computer interaction in health systems including improving clinical communications through intelligent virtual tutors. Prior to Starship, he was CEO of the award-winning CHI Systems, Inc. for over 20 years, a company which he also founded. He remains active in the research and technology communities, with over 110 technical and scientific publications.

Dr. Diego Zapata-Rivera is a Principal Research Scientist in the Cognitive, Accessibility, & Technology Sciences (CATS) Center at Educational Testing Service in Princeton, NJ. He earned a Ph.D. in computer science (with a focus on artificial intelligence in education) from the University of Saskatchewan in 2003. His research at ETS has focused on the areas of innovations in score reporting and technology-enhanced assessment (TEA) including work on adaptive learning environments and game-based assessments. His research interests also include Evidence-Centered Design, Bayesian student modeling, open student models, conversation-based assessment, supporting collaboration, virtual communities, authoring tools and program evaluation. Dr. Zapata-Rivera has produced over 100 publications including journal articles, book chapters, and technical papers. He has served as a reviewer for several international conferences and journals. He has been a committee member and organizer of international conferences and workshops in his research areas. He is a member of the Board of Special Reviewers of the User Modeling and User-Adapted Interaction journal and an Associate Editor of the IEEE Transactions on Learning Technologies Journal. Most recently, Dr. Zapata-Rivera has been invited to contribute his expertise to projects sponsored by the National Research Council, the National Science Foundation, NASA, and the US Army Research Laboratory.

INDEX

Design Recommendations for Intelligent Tutoring Systems

Volume 6
Team Tutoring

Design Recommendations for Intelligent Tutoring Systems (ITSs) explores the impact of intelligent tutoring system design on education and training. Specifically, this volume examines "Team Tutoring". The "Design Recommendations book" series examines tools and methods to reduce the time and skill required to develop Intelligent Tutoring Systems with the goal of improving the Generalized Intelligent Framework for Tutoring (GIFT). GIFT is a modular, service-oriented architecture developed to capture simplified authoring techniques, promote reuse and standardization of ITSs along with automated instructional techniques and effectiveness evaluation capabilities for adaptive tutoring tools and methods.

About the Editors:

- ***Dr. Robert A. Sottilare** leads adaptive training research at the Army Research Laboratory and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).*
- ***Dr. Arthur C. Graesser** is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is a Senior Research Fellow in the Department of Education at the University of Oxford.*
- ***Dr. Xiangen Hu** is a professor in the Department of Psychology at The University of Memphis and visiting professor at Central China Normal University.*
- ***Dr. Anne M. Sinatra is an** adaptive training scientist at the U.S. Army Research Laboratory.*

9 780997 725742