International Society of the Learning Sciences

Multimodal Learning Analytics Using Hierarchical Models for Analyzing Team Performance

Caleb Vatral, Gautam Biswas
caleb.m.vatral@vanderbilt.edu, gautam.biswas@vanderbilt.edu
Institute for Software Integrated Systems, Department of Computer Science, Vanderbilt University

Benjamin S. Goldberg
benjamin.s.goldberg.civ@army.mil
Simulation & Training Technology Center, US Army CCDC – Soldier Center

Abstract: This paper presents a comprehensive hierarchical model for teamwork by extending the well-known Affective, Behavioral, and Cognitive (ABC) approach for analyzing team performance. We develop a framework for interpreting individual and collective team processes, which allows us to create mappings from collections of actions to individual and team performance measures. The performance analysis algorithms use rich multimodal video, speech, and eye tracking data to monitor user activities in the teamwork scenario. We present an initial case study of soldiers training in small groups in a mixed reality training environment. Using data from the training scenarios, we demonstrate the use of multimodal learning analytics methods to analyze video data, infer soldiers' individual and collective actions as a scenario evolves, and establish a proof-of-concept for our analysis methods by comparing against expert judgment. We conclude that our hierarchical ABC model combined with MMLA presents an effective approach to analyzing and evaluating team performance in training applications.

Introduction

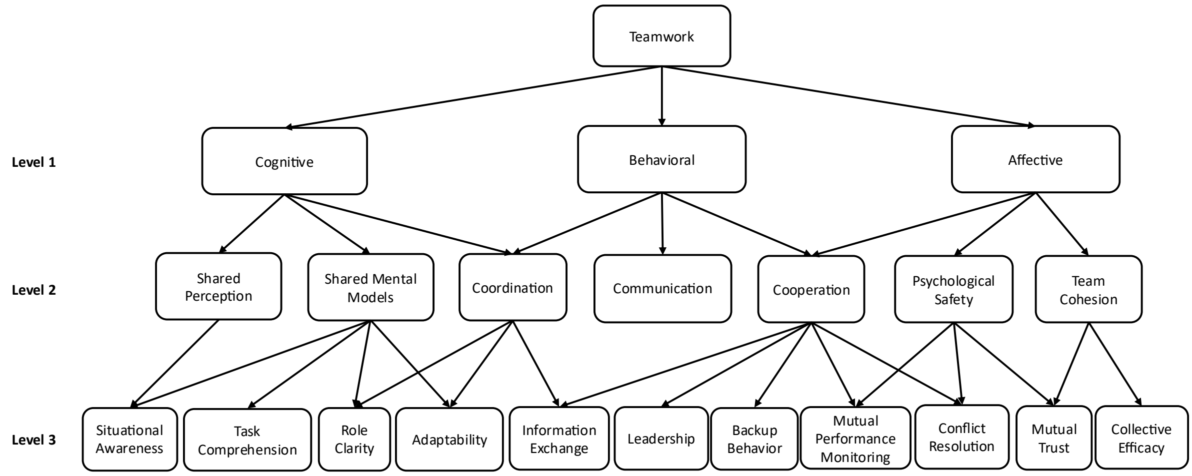
The increasing importance of teamwork has led organizations to develop training methods that help team members develop the necessary skills to become effective in complex problem-solving scenarios. Achieving competency in teamwork involves several parameters that go beyond individual learning but, traditional methods for evaluating teamwork often rely on time-consuming and intrusive interviews with team members and other manual analyses (Zachary et al., 2000). However, the increasing availability of sensors capable of unobtrusively monitoring team actions and behaviors during training, and new advanced AI and machine learning (ML)-based algorithms capable of analyzing complex tasks, makes it possible to evaluate teamwork more systematically. In this work, we develop a framework for evaluating teamwork competency using a comprehensive, hierarchical task interpretation model based on the well-known ABCs of teamwork (Kozlowski, 2018). In addition, we develop multimodal learning analytics methods to interpret trainee actions and derive performance metrics for their task-related activities using our hierarchical ABC framework (H-ABC). We show that measures computed using this framework match the evaluations from a human domain-expert instructor. Through a mixed-methods approach combining these data-driven evaluations with traditional instructor feedback, trainees can get rich feedback on their taskwork and teamwork performance.

Theoretical Framework

When analyzing teamwork, it is common to describe cognitive skills as a set of transferrable competencies that mediate performance and apply across multiple domains. In this paper, we adopt the ABC model of teamwork (Weaver et al., 2010; Kozlowski, 2018), which theorizes that teamwork is set of temporal processes that include: (1) *Affective processes*, i.e., the emotional, attitudinal, and motivational aspects of teamwork including concepts such as mutual trust, self and group efficacy, and team cohesion; (2) *Behavioral processes*, i.e., actions performed by the group and the resultant emergent states, such as communication, coordination, cooperation, regulatory mechanisms, and decision-making; (3) *Cognitive processes*, i.e., the knowledge structures held by the group, including concepts such as shared mental models, collective perception, and transactive memory.

We extend the traditional ABC taxonomy into a multi-level hierarchy, where low-level concepts link to *multiple* high-level processes, making the model more expressive. The hierarchical structure is generated by *cognitive task analysis* and review of relevant literature (Cooke et al., 2013; Kozlowski, 2018; Sottilare et al., 2018; Weaver et al., 2010; Zachary et al., 2000). Concepts at the higher levels represent more abstract teamwork processes and each deeper level represents more concrete and directly observable competencies. Figure 1 shows our extended hierarchical ABC model for teamwork. The concepts included in the hierarchy mediate teamwork performance across multiple domains, but many of these concepts are not directly observable or measurable. Instead, they must be interpreted in specific task environments using domain-specific analysis of component behaviors. In the next sections, we present our proposed approach utilizing multimodal data collected from the teamwork task environments to measure these competencies. To explain and validate these methods, we present the methods in the context of a case study of soldiers undergoing training drills.

Figure 1. *The complete hierarchical ABC (H-ABC) model to describe teamwork. Note that links from communication to every competency in level 3 were omitted to avoid overcrowding the figure.*



Case Study: Soldier Training Drills

Two infantry fire-teams of three and four soldiers each participated in the study. Each team performed training on the *Enter and Clear a Room* (ECR) dismounted battle drill that simulates operations in modern urban warfare. In the drill, the team enters a room to control the space and neutralize enemy combatants. Each room may contain enemy combatants, civilian non-combatants, physical obstacles, and weapons. The ECR training was conducted on the *Squad Advanced Marksmanship Trainer* (SAM-T), a mixed reality synthetic training environment consisting of three screens setup in a U-shaped arena. We collected soldier movement data, event data from the SAM-T simulation and screens, weapon sensor data, and biometric harness data, using the open source Generalized Intelligent Framework for Tutoring (GIFT) (Goldberg et al., 2021) to enable the synchronized capture of the data. For a complete overview of ECR, SAM-T, and GIFT, see Vatral, et al. (2021). Each training drill took approximately 2 minutes and each team performed between 20 and 30 drills. Between each drill, an instructor provided feedback and suggestions for improving performance. We recorded these conversations for subsequent analysis.

Analysis Methods: Multimodal Learning Analytics

Traditionally, two major techniques have been used for analyzing teamwork behaviors: *retrospective analysis* and *think-aloud protocols* (Zachary et al., 2000). Retrospective analysis involves interviews, where team members recall their behaviors and performance during the exercise. This analysis requires continual participation of domain experts, requiring significant overhead and cost, and recollections of the team may not be accurate. In addition, the time taken for interviews takes away time for further practice and training. In the think-aloud protocol, team members are asked to verbally introspect on their behavior and thoughts during an exercise. Vocalization of thought processes is highly intrusive to natural communication patterns and the added cognitive demands of expressing thought processes may lead to degradation of performance. However, the increasing availability of high-fidelity sensing devices, which can unobtrusively capture subjects' activities and cognitive processes, combined with advanced multimodal learning analytics (MMLA) provides a viable alternative for analyzing teamwork behaviors (Blikstein & Worsley, 2016; Starr et al., 2018). MMLA combines data from multiple sensors in the environment using analytical techniques derived from artificial intelligence and machine learning to analyze team performance. In this work, we use our H-ABC model as a theoretical framing to ground our MMLA analyses.

To start the analysis, we define a set of performance metrics, which describe domain-specific measures of how well teams achieve their goals. For our case study, we have generated 10 performance metrics for evaluating team performance. Space limitations prohibit discussing details of the metrics, but for a full review, see Vatral et al. (2021). To compute these performance metrics, we utilize AI and ML techniques to analyze the observable data generated by the sensors in the task environment. For this paper, we limit the analysis to video data, supplemented with some manual coding of communication patterns; however, a variety of other sensor data will be analyzed in future work. These computable performance metrics are then linked to the lowest layer to the H-ABC model in consultation with task domain experts and instructors. This enables us to establish computational links between these performance metrics and the level three teamwork competencies.

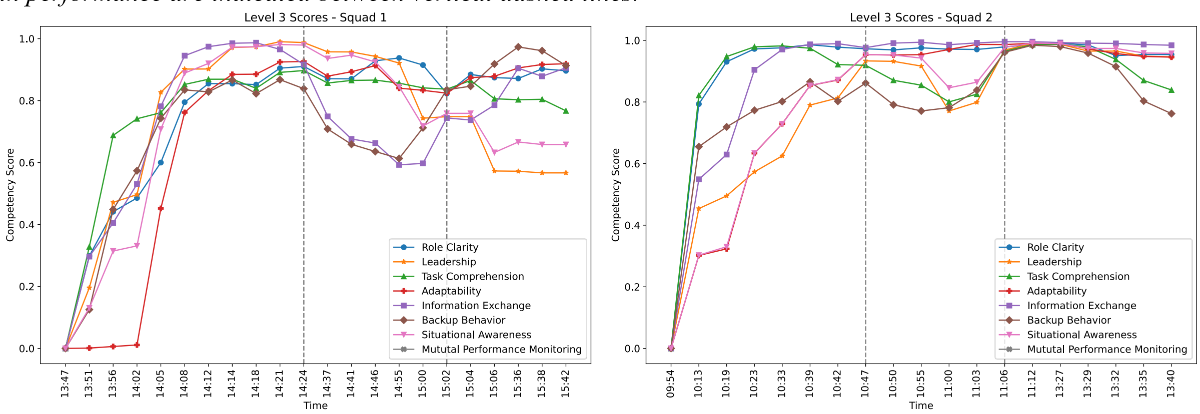
While these metrics do represent a direct observation of performance, their computation ignores the dynamic and temporal nature of teamwork. Scores on each training scenario are not independent events: teams learn as they continue to train. In previous work (Biswas et al., 2020), we utilized first-order differences in metric scores to compute *trends,* to evaluate changes in performance over time. However, trend measurements can be noisy; a *slip up* may result in performance scores that do not reflect the team's true competency. Conversely, a team may perform well in an exercise *by chance*, which again does not represent their true competency. To overcome these issues, we apply a modification of Bayesian knowledge tracing (BKT) methods (Corbett & Anderson, 1995). In our model, instead of the traditional binary outputs of BKT, we use the continuous-valued performance metrics, represented in the HMM by a Gaussian mixture model with two means: (1) a mean close to 1 to represent *satisfactory* execution of an activity; and (2) a mean close to 0 to represent *unsatisfactory* execution of the activity. Weighting the means in the mixture and their respective variances controls the distribution of slips and guesses. The competency score for a given performance metric is represented by the probability of being in the learned state of the BKT HMM. This approach is robust to temporal variations in team performance.

Given the temporal sequence of BKT performance metrics and the newly extended H-ABC model, multiple low-level performance concepts can be combined to evaluate higher-level teamwork competencies by following the links up through the hierarchy. For each competency, *c*, at level *n* of the ABC hierarchy, we compute the average of the competencies at level *n+1* that have links to competency *c.* In future work, we will explore more sophisticated methods for propagating the competency scores to higher level concepts in the hierarchy, but as a proof of concept, this simple averaging works well. By propagating the derived performance metrics up the hierarchy, we can generate a complete evaluation of teamwork at multiple levels of abstraction.

Results

After each training drill, we computed performance metrics for the two fire-teams that participated in our study using the BKT model to produce scores for the ten performance metrics. For each competency at level three, we averaged the linked performance metrics at each time step to produce a sequence of scores for the general teamwork concepts. Results of the level three competencies for each team are displayed in Figure 2. Both teams showed similar trends in their competencies. Each began with a period of rapid ascent in all their scores, representing the initial learning period where teams become familiar with the drill. Toward the middle of training, indicated by the regions between the two vertical dashed lines in Figure 2, we observed a drop in performance for both teams across several competencies. Team 1 dropped performance in *information exchange* and *backup behavior,* while team 2's performance dropped in *task comprehension*, *backup behavior, leadership,* and *situational awareness*.

Figure 2. *Level 3 competency scores for the two participating soldier teams. Areas of interest representing dips in performance are indicated between vertical dashed lines.*



Of particular interest is the relationship between the performance drops in the computed competency measures and the instructor feedback during these periods. First examining team 1, the performance drops begin with the exercise at 14:24. At the same time, instructor feedback and discussion increased. The issues indicated by the instructors and the subsequent discussions align well with the teamwork competency measure drops during this period. In the 14:24 scenario, *information exchange* decreased because team members failed to communicate vacating of their sectors of fire. *Backup behavior* failed because soldiers did not adjust their sectors of fire to aid their teammate, even when they heard rapid gunfire. For the 14:41 scenario, *information exchange* and *backup behavior* failed because the presence of an explosive device was not communicated soon enough. In the 14:46 exercise, *information exchange* failed, as soldiers did not call out their identifications of combatants. In the 14:55 exercise, *information exchange* and *backup behavior* failed as the team did not evaluate the health status of their injured team leader and did not take over the team leader's responsibilities. After this exercise, competencies rose again as the team completed more drills without issues and the updated model showed increasing competencies.

Team 2 showed similar results, with their performance measures and subsequent discussions, matching well with the evaluated competencies during the performance drop period. In the 10:47 exercise, both *task comprehension* and *backup behavior* failed as team members did not call their exits as protocol dictated. In the exercise at 11:00, *task comprehension* and *situational awareness* failed as the team did not search the neutralized entities, and *backup behavior* and *leadership* failed, as no other team members emerged to fulfill the team leader's roles and responsibilities after the leader was injured. Even though both teams had similar lengths of their respective poor performance periods, team 2 had more intermittent successes during this period, meaning the model's confidence that the team understood the competencies did not drop as much as it did for team 1.

Conclusions and Future Work

In this work, we presented the new H-ABC model for evaluating teamwork and showed how the model can utilize multimodal learning analytics for evaluation of teamwork competencies. We validated our approach using a case study with soldier teams training in a mixed-reality simulation environment and showed that evaluations generated by the H-ABC model supported by MMLA match evaluations given by domain-expert instructors. These results indicate a great potential that such model-based and data-driven evaluations can be used to supplement instructor feedback and improve team training. By providing data-driven evaluations alongside instructor feedback, teams will have more insight into their learning behaviors, both in an individual training instance and across time.

This work and the field of multimodal learning analytics is in its early stages. Additional data collection and analyses are necessary to develop this methodology. However, the initial findings presented here indicate great potential for automated analyses and supplementing instructor feedback with data-driven evaluations. The current study examined only two soldier teams training on a single drill. Future work will apply similar analysis to additional sensor modalities (e.g., log files from the simulation, speech, and physiological data), a greater number of teams, and to other domains, further validating the generalizability of the model and analysis methods.

References

Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238.

Biswas, G., Rajendran, R., Mohammed, N., Goldberg, B. S., Sottilare, R. A., Brawner, K., & Hoffman, M. (2020). Multilevel Learner Modeling in Training Environments for Complex Decision Making. *IEEE Transactions on Learning Technologies*, *13*(1), 172–185.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. Cognitive Science, 37(2), 255-285.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, *4*(4), 253–278.

Goldberg, B., Owens, K., Hellman, K., Robson, R., Blake-Plock, S., Hoffman, M., Gupton, K. (2021). Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy. In *Proceedings of the 2021 I/ITSEC*. Orlando, FL

Kozlowski, S. W. J. (2018). Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspectives on Psychological Science*, *13*(2), 205–212.

Sottilare, R. A., Shawn Burke, C., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing Adaptive Instruction for Teams: a Meta-Analysis. In *IJAIED* (Vol. 28).

Starr, E. L., Reilly, J. M., & Schneider, B. (2018). Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. In Kay, J. and Luckin, R. (Eds.) *13th ICLS*, Volume 1.

Vatral, C., Mohammed, N., Biswas, G., & Goldberg, B. S. (2021). Gift external assessment engine for analyzing individual and team performance for dismounted battle drills. In *9th Annual GIFTSym* (p. 109-127).

Weaver, S. J., Rosen, M. A., Salas, E., Baum, K. D., & King, H. B. (2010). Integrating the science of team training: Guidelines for continuing education. *Continuing Education in the Health Professions*, *30*(4), 208–220.

Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (2000). Building Cognitive Task Analyses and Models of a Decision-making Team in a Complex Real-Time Environment. In Chipman, Shalin, & Schraagen (Eds.), *Cognitive Task Analysis*. Erlbaum.

