Team Assessment Methods for Training Command Staff Decision Making

Randy Jensen[1], Evan Finnigan[1], Grace Teo[2], Gregory Goodwin[3]
Stottler Henke Associates, Inc.[1], Quantum Improvements Consulting, LLC[2],
U.S. Army Combat Capabilities Development Center (CCDC)[3]

INTRODUCTION

This paper shares an updated view of assessment and feedback methods under development for team training in an applied Army command-level decision making domain. The overarching objective is to work toward generalizable practices for assessing team performance and team processes in command-level training domains where team effectiveness centers on the dynamics of teamwork and associated cognitive skills. The initial example training application focus for this effort is on wargaming, a critical stage in the Army's military decision making process (MDMP). Wargaming is a collective activity where command staff perform course of action (COA) analysis by stepping through major events and considering actions, reactions, and counteractions, with representatives of multiple warfighting functional areas contributing perspectives critical to the synchronization of plans. A prototype is currently under development for this application, designed as a distributed team trainer. This is intended to create an environment where a team of participants can join from remote locations and step through a facilitated wargaming exercise with opportunities to exhibit teamwork behaviors in the process of collaboratively walking through a COA scenario. A previous GIFTSym presentation (Jensen et al., 2021) described the preliminary structural design of the distributed trainer and interactions across modules and GIFT (Generalized Intelligent Framework for Tutoring, Sottilare et al., 2012). The focus for this paper is on the design of team assessment methods for command staff wargaming in the prototype. Teo et al. (2021) presented a model for teamwork constructs applicable to wargaming performance assessment, which provides the theoretical underpinning for team dimensions to be utilized for assessment in the prototype application. The model is adapted from common constructs appearing in the literature (Johnston et al., 1998; Kozlowski et al., 2015; Marlow et al., 2018; Salas et al., 2005; Sottilare et al., 2018), and tailored to command-level teamwork in wargaming. Teamwork dimensions in the model include Leadership, Team Cognition, Information Exchange, Communication Quality, Supporting Behaviors, and Team Orientation.

This paper walks through practical design considerations for constructing assessment rules that are organized around the team dimensions in the model, and that can be applied in wargaming exercises under development in the current effort. A fundamental goal for this effort in a technology sense is to construct automated assessment capabilities that can facilitate instructors as much as possible in identifying markers of teamwork exhibited in a wargaming exercise. And yet this goal is also moderated by the need to include mechanisms for human observers and/or team participants themselves to contribute assessment inputs when feasible, to augment those generated automatically. Since there are channels of team interaction that the system is unable to parse (such as communications outside of the training environment), it is important to provide a means for assessments based on human understanding, when available. In complex team decision making domains, assessment results are ideally echoed to the training audience in a manner that stimulates a guided process of self-correction within the team. Thus, markers intended for presentation in debriefing come from two sources: automated assessments and manual inputs by human observers or participants. Both are tagged with relevant team dimensions, and information about noted positive or negative teamwork behaviors. Tagging facilitates post-processing to distill the aggregated set of markers for practical use in team reflection. This paper gives examples of assessment rules in the wargaming application, and preliminary design thoughts for how assessment logic

will be encoded in the GIFT Domain Module. Although this is work in progress, the intention is to share design and development approaches that may inform related efforts.

TEAM TRAINING APPLICATION

The team training application discussed throughout this paper is under development in an effort called Reusable Automated Assessment and Feedback for Teams (RAAFT), which is being conducted for the U.S. Army Combat Capabilities Development Command Soldier Center. With the focus of the RAAFT distributed training prototype on MDMP and wargaming, the development team observed instructor-led training at the Command and General Staff College (CGSC) at Fort Leavenworth, where students have the opportunity to practice wargaming in fully-developed scenarios. The curriculum starts with classroom training to introduce concepts before ultimately moving to culminating exercises with cohorts of students performing in command staff roles throughout the MDMP process. The initial prototype is designed for preliminary training, to work on the dynamics of teamwork in wargaming scenarios, while operating under a number of constraints to simplify exercises. As a browser-based application, it supports distributed training with participants in remote locations, but this is not a requirement and the trainees can also be co-located. Some elements of the training experience that are simplified for this instructional context include:

- Participants. The initial training application is structured for 6 students assigned to key warfighting functions in each exercise. In contrast, a full command staff performing COA analysis represents a wider range of warfighting functional areas, and typically involves greater numbers of participants.

- Observer / instructor roles. In addition to student participants, the training environment allocates roles for observers / instructors, providing the opportunity to monitor exercise events, and tools to add markers for teamwork (complementary to automated assessment mechanisms which also generate teamwork markers).

- Structured decision making. The COA analysis process is approximated in the training application with a structure involving predefined, scenario-oriented prompts. In real-world wargaming, MDMP practices provide a certain amount of structure for evaluating each phase of the COA, with open-ended analysis of the events involved, enemy reactions, possible counteractions, intelligence requirements, associated time and distance calculations, projected adjudication of engagements, and so on. However, in order to make it less open-ended to work effectively in a browser-based setting, the training environment uses structured prompts to guide participants through a constrained sequence of decisions. These prompts aim to be analogous to freeform wargaming decisions, with similar opportunities for teamwork.

- Overt teamwork actions. Input mechanisms are integrated into the trainer design to make it easy for participants to convey information or coordinate with other team members during an exercise. Examples include widgets to send pre-formatted messages, to facilitate not only the act of passing information, but also the training system's ability to track what's being conveyed to whom. Another example is a set of widgets used when reviewing decisions, to mark when team members agree with decisions or wish to discuss further. The initial training application does not attempt to process the content of freeform chat messages, or any of the natural communications that may happen between co-located participants.

- Decision engagement. Partly because of the distributed setting, one design concern was to keep participants engaged. So the exercise flow is designed to include all participants in all prompts, at

least for collecting initial responses. In real-world wargaming processes conducted in person, discussion of a particular topic or decision may be limited to the representatives of warfighting functional areas directly involved in the decision. For example, for a decision requiring a choice of suppression methods against an enemy target where options may include either artillery or close air support, the fire support and aviation leads may be the main contributors. However, all prompts in the trainer elicit responses from all participants. This approach was chosen not only to maintain engagement throughout the exercise, but also for a secondary benefit of yielding insights about shared thinking among the team with each prompt, as participants are asked to give opinions about decisions that don't directly involve their own functional areas.

- Team-driven exercise control. In order to mimic real-world command staff wargaming and also support distributed training, the environment is designed so that the team of participants manages their own progress through an exercise. The leader, in this case the Chief of Staff (COS), is given special controls to advance through "turns" and prompts in the exercise. This is intended to resemble the COS leadership role in real-world wargaming. Exercise controls are designed to be simple (e.g., one-click tools to advance) to minimize special familiarization for the COS in order to carry out an exercise. Also most cases of system-generated feedback are delivered via the COS, who has an opportunity to review before relaying to other staff team members.

These design constraints in the training environment provide background for the discussion of team assessment methods. In constructing automated assessment rules, first it is helpful to delineate constructs relating to team performance and team processes, to help organize behavioral indicators to be monitored by the system (Grand et al, 2013). **Team performance** relates to outcomes in terms of team tasks or status, whereas **team processes** relate to the dynamics of interactions or cognitive states within team members. In a wargaming context, team performance examples include planning decisions in the COA scenario (e.g., a choice of a route or position, unit tasking, timing decision, method of attack, etc.). Team processes can include any of the dynamics that lead to these outcomes (e.g., information sharing or communication in the process of deciding upon a route, etc.). Assessment rules for the wargaming trainer are designed to identify behavioral indicators, either in specific actions or sequences, that match conditions for a teamwork marker of interest. For example, among the exercise environment's mechanisms for overt teamwork actions, the design includes interface elements for team members to express agreement or a need to discuss further, as the team works toward collective decisions during wargaming. These inputs can be analyzed in sequence to recognize patterns of supporting behavior even when lacking knowledge of other modes of communication (e.g., unparsed chat messages). Similarly, actions in the exercise environment serve as a data source for other team process indicators of information sharing, leadership, and knowledge of roles. Measures like the *frequency* of messaging between certain roles associated with certain elements of COA decision making provide insight into team processes, even if freeform messages are not parsed for semantic content.

Team performance assessment rules are designed to look more directly at the content of decisions, which is relevant for dimensions such as team cognition. For example, although COA analysis frequently involves the consideration of competing options without a singular best answer, in some cases there may be decisions that reflect a lack of understanding or a lack of synchronization among the team. In such cases, if an explicitly suboptimal decision is reached, this can be an indicator of poor shared awareness of the situation or objectives. Assessment rules should capture behaviors at the granular individual level in order to support traceability to markers referencing dimensions at the team level. This is especially helps for individual trainees to be able to see how their own behaviors affect performance outcomes.

Example Training Vignette

In this section we step through a condensed example sequence of prompts and team inputs in a wargaming exercise. For illustration, it omits some team interactions that would occur in a team training event with 6 participants and one or more human observers. The sequence provides context for selected examples of assessment applied to both team processes and team performance.

The participating roles include the Chief of Staff (COS), with five staff leads for Intel, Maneuver (ground maneuver and plans), FSCOORD (fire support coordinator for indirect fires), Aviation, and Logistics (also called sustainment). The exercise is structured around the wargaming, or analysis, of a single course of action (COA), which proceeds through a series of exercise turns corresponding to major events, phases or decision points in the battle. The pattern for each exercise turn involves a series of prompts to first consider the next planned action, the enemy reaction, and the Blue counteraction. Each turn also involves at least one prompt for a primary decision about how the action will be carried out or synchronized across warfighting functional areas. Additional prompts may also be given to probe dimensions of teamwork relevant to the context of the turn at hand. All prompts are designed to conform to one of two types:

- ***Discussion*** prompts assess teamwork dimensions in context
    - Directed to all participants, for the purpose of prompting discussion and gathering inputs, with no explicit step for team review of answers. Participants do not see others' answers.

- ***Decision*** prompts involve COA decisions
    - Often multi-part prompts, for example with a decision and associated rationale.
    - Initially directed to all participants, with a designated lead role for the decision. After all participants respond, there is an opportunity for team review of the lead's answer.
    - After team review, and any revisions, the lead's answer is recorded as a COA decision.

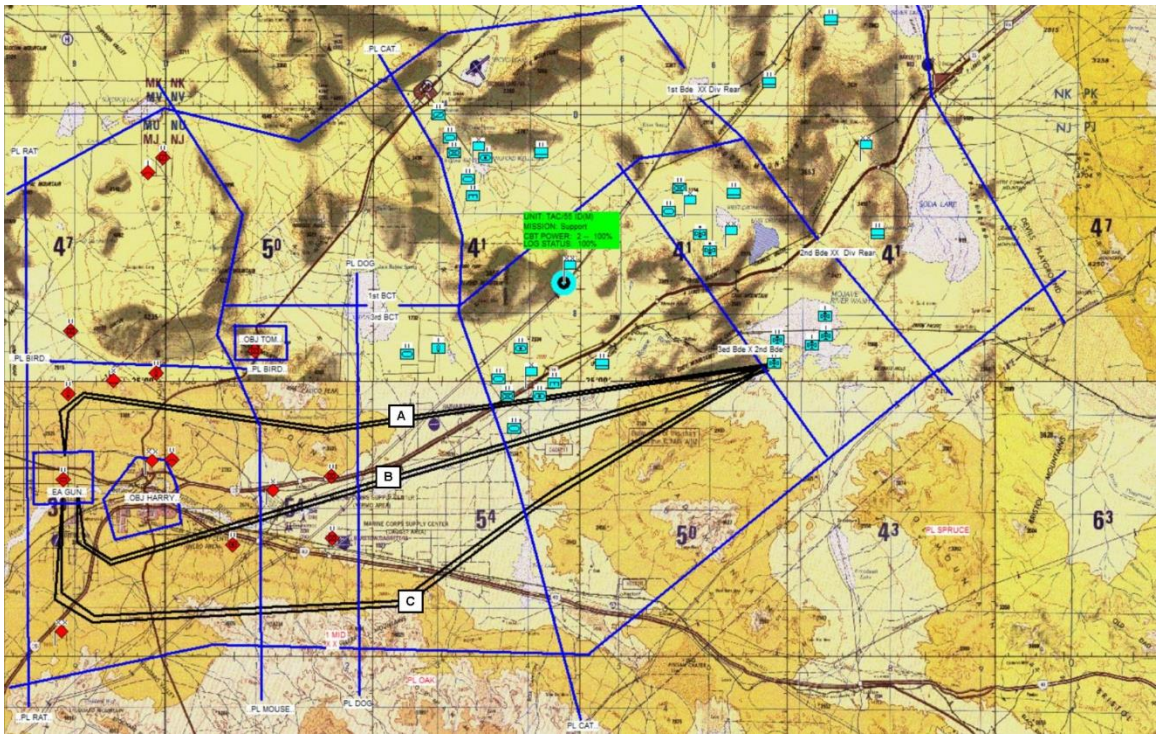

Figure 1. Tactical Map View for Choice of Helicopter Attack Route

Figure 1 above shows the tactical map presented at the start of a wargaming turn involving a helicopter attack. In this turn, the command staff reviews the helicopter attack plan, with prompts about action, reaction, and counteraction, followed by a COA decision to choose among three helicopter attack routes (A, B, or C as shown on map overlays). This is one of the ways that the distributed trainer is a simplification of real-world wargaming, which would not involve predefined choices for attack routes. Although terrain or other elements may ultimately limit the range of options for certain decisions, a real-world training audience would not have options explicitly provided, as they are here. At the start of this wargaming turn, the map is pre-configured to display the overlays with the three routes. The following figure steps through a condensed sequence of prompts and team inputs for this turn.

Turn: Plan helicopter attack			
1	Discussion Prompt	[Discuss Action – no decision lead] *What is the purpose of the attack helicopter mission in this COA?*	
	a)	[All]	[Response] *Destroy the Enemy reserve.*
2	Discussion Prompt	[Discuss Reaction – no decision lead] *What enemy defenses are a threat to the helicopter mission?*	
	a)	Intel	[Response] *air defenses at OBJ HARRY, ground units along route*
	b)	FSCOORD	[Response] *air defenses at OBJ HARRY*
	c)	COS	[Communication] *@Intel Please share your thoughts*
	d)	Intel	[Communication] *All routes are threatened by air defenses at OBJ HARRY*
	e)	[Others]	[Responses omitted]
3	Discussion Prompt	[Discuss Counteraction – no decision lead] *What action should we take if the attack helicopters cannot complete their mission?*	
	a)	FSCOORD	[Response] *Use available on-call CAS to attack the enemy reserve*
	b)	Aviation	[Response] *Task the Aviation Brigade to conduct a second attack*
	c)	Logistics	[Response] *Take no further action*
	d)	[Others]	[Responses omitted]
4	Decision Prompt	[COA decision – lead role is Aviation] *Select the best helicopter attack route and give rationale.*	
	a)	FSCOORD	[Communication] *Active fire line planned for RED 208 Arty Bn north of EA GUN*
	b)	FSCOORD	[Response] *Route: B* *Rationale*: *low fratricide risk, minimal conflict with other targets, short distance*
	c)	Aviation	[Response] *Route: C* *Rationale*: *low fratricide risk, minimal conflict with other targets*
	d)	[Others]	[Responses omitted]
	e)	COS	[Exercise Control - trigger review of Aviation inputs]
	f)	Maneuver	[Review] *Agree*
	g)	FSCOORD	[Review] *Discuss*
	h)	FSCOORD	[Communication] *@Aviation B and C are good, but B is shorter*
	i)	Aviation	[Communication] *B overflies too many enemy positions*
	j)	[All]	[Review] *Agree*
	k)	COS	[Exercise Control - trigger logging of Aviation inputs]

Figure 2. Example Prompts and Team Inputs for Choice of Helicopter Attack Route

In the example sequence above, there are three discussion prompts (for discussion of action, reaction, and counteraction), followed by a decision prompt. Sample participant inputs are shown with each prompt, labeled either by individual role (COS, Intel, FSCOORD, etc.) or as a group ([All]) when all have the same input. For this example, there are four kinds of input:

- [Response] inputs are the answers directly responding to a prompt. These are submitted using user interface mechanisms that vary with the prompt, such as multiple choice, checklists, and dropdown menus. For example, 4(b) and 4(c) involve a combination of choosing a route by multiple choice, and specifying rationale from checklists with general purpose objectives.

- [Communication] inputs are messages in the chat panel, created either as freeform text or using a dialog window for pre-formatted messages from templates.

- [Review] inputs capture indications from team members when reviewing a decision. These are simply either an Agree or Discuss value.

- [Exercise Control] inputs are actions taken by the COS to advance exercise states.

Examples of teamwork assessments in the following section refer to the prompts and team inputs in the sequence above.

ASSESSMENT METHODS

The objective of team assessment methods in the wargaming staff trainer is to produce markers of good and bad teamwork from an exercise, for the purpose of team review and self-correction in an after action review (AAR) discussion. For the examples of assessment methods in this section, we focus primarily on the conditions and logic that produce teamwork markers.

Assessment Examples

Referring to the sequence of prompts and team inputs in Figure 2 above, the following examples give a sampling of cases where assessments are made, the conditions they look for, and the team dimensions they relate to.

Response Correctness

The concept of the correctness of individual responses to prompts during the exercises is intended to play only a small role in assessments of teamwork. In MDMP in general, but more specifically in wargaming but also many other forms of tactical decision-making, there are no singular correct answers when it comes to tactical choices. But 1(a) is an example where a very simple prompt is expected to produce a straightforward answer ("*Destroy the Enemy reserve*."). As the staff team begins the discussion of the helicopter attack, this prompt is a simple spot-check about the purpose of the attack, which is roughly considered a situational awareness level one question. In this case, there really is only one correct answer, and it would be unusual if team responses are not all correct. So while many prompts may provide possible answers that are all considered acceptable, in this case the scenario markup includes information about which answers are considered unacceptable, and what team dimension is involved. The actual assessment mechanism is also very simple in this case, and it only produces a marker when unacceptable answers are received. In 1(a) no marker is created. Had the responses included unacceptable answers, then a marker would be generated and tagged for Team Cognition, relating to team performance (as opposed to process). However, because of the relative low priority on markers related to correctness, this would most likely not be highlighted for team discussion in AAR. Other cases where correctness comes into play can include compound answers to prompts, such as the rationale associated with a route (e.g., 4(b)), which can be evaluated for the validity of the association between the two.

Response Consistency

In 3(a), 3(b), and 3(c), three different staff members give three different responses to the same prompt, a "what-if" question about possible counteractions, as part of the initial structured discussion of the helicopter attack (to consider action, reaction, counteraction). In contrast to prompts involving concepts of correctness, none of these responses is necessarily incorrect, even if some might be considered better than others. However, the main problem with the fact that team members have three different ideas of the best counteraction in this case is that this is likely to lead to problems during execution. In fact a significant part of the purpose of wargaming is to synchronize plans. In both real-world wargaming as well as the approximation in this exercise, discussion between staff members is encouraged. When each prompt is directed to the training audience, they are welcome to discuss answers before submitting. So the differences in answers also suggest that there has not been enough communication (team process), in addition to the lack of an integrated mission plan (team performance). Assessment rules to generate teamwork markers relating to consistency are tagged with markup about the level of consistency needed, and the team dimensions involved – in this case Information Exchange and Team Cognition.

Leader Support

2(c) is a simple example of leader support during an exercise ("*@Intel Please share your thoughts*"), and often there will be many such examples. The team dimension of Leadership relates to the degree of guidance, direction, and coordination from those in team leader roles. In this case, the COS understands that the Intel officer is the primary source for information about likely enemy capabilities, actions, and tactics. The prompt itself asks about enemy defenses relating to the helicopter attack, with responses available in a checklist. So the COS asks Intel to share thoughts to help the staff, not only for the immediate purpose of responding to the prompt, but also to reach shared awareness about action, reaction, and counteraction. Commonly the Intel officer volunteers this kind of information anyway, as a matter of routine for any questions about projecting enemy actions. In this case, the COS uses a pre-formatted message to Intel, which allows the system to recognize the content of the message and its relevance to the current prompt, and generate a positive teamwork marker for Leadership. This is a simple example, but it should be noted that for more complex examples or also cases where the pre-formatted message dialog is not used, a human observer may also produce the same or similar marker. Generally the actions of the leader to support the team relate to measures of team processes, and are somewhat independent of the outcomes reflected in the answers collected for the prompts themselves (except for the fact that good team processes ideally produce better team performance outcomes).

Providing Relevant Information

In the previous example with the COS eliciting information from Intel, the information shared by the Intel role (2(d) – "*All routes are threatened by air defenses at OBJ HARRY"*) is useful and relevant. So this also relates to the team dimension of Information Sharing. However, it's not an optimal example. Intel's answers to the prompt included both the enemy air defense threats and also the concern about enemy ground units along any helicopter attack route to be considered. But the Intel officer did not share comments about the latter concern involving ground units. Another example of Information Sharing is in 4(a) ("*Active fire line planned for RED 208 Arty Bn north of EA GUN*"). Once the prompt initiates consideration of the helicopter attack routes, the FSCOORD proactively shares information to inform or remind all staff members about a planned active gun-target line. This involves indirect fires against the enemy at EA GUN at the time of the helicopter attack, so routes passing through this active fire line should be avoided (in this case, avoid route A, in Figure 1). The pre-formatted message dialog is constructed to make this kind of information sharing easy for participants, where role-specific information is available to simply select and send. So for example, the message in 4(a) is only available to the

FSCOORD to share. Not all such information is relevant at all times, so the assessment logic looks for cases of messages sent, along with markup about relevance. In this case, a positive teamwork marker is generated, which is related to team processes and involves the team dimensions of Information Sharing and Supporting Behavior. If the FSCOORD typed a similar message, or verbalized similar comments in a co-located or VTC setting, then the system has no knowledge of this kind of natural teamwork, but a similar marker can be created by a human observer or participant.

Participation in Decisions

Another assessment measure related to team processes in wargaming involves the participation of team members in the decision prompts. As discussed above, the decision prompts involve a team review sequence. These prompts are initially directed to all participants, but there is a designated lead role for the decision. In the example exercise turn above, involving the helicopter attack route decision, the Aviation staff lead is the designated lead for the prompt involving the actual choice of the route (prompt #4). After all participants respond, the next step is for the team to review the lead's answer and discuss any possible changes before then recording an answer as a decision for that turn in the COA. For each decision prompt, scenario markup includes information about which roles should be a part of the final decision, based on domain knowledge related to the decision itself. The exercise environment implements a simple formalization for process-oriented data capture in decision prompts, with user interface tools to express agreement or a desire to discuss a decision further.

In 4(g) above, the FSCOORD wishes to discuss the Aviation lead's choice of route C. In subsequent communications, the FSCOORD suggests that route B is shorter (while still avoiding the gun target line mentioned earlier), but the Aviation lead also points out the route B forces the helicopters to overfly too many enemy ground units. Ultimately this leads the FSCOORD (and all other team members) to agree with the choice. There are several opportunities for assessment rules to look for good and bad teamwork markers based on these inputs. If the team proceeds with a decision without receiving any input (*Agree* or *Discuss*) from key roles for the prompt, then this may be a shortcoming in Leadership and/or Supporting Behavior. If a discussion sequence occurs while reviewing the designated lead's input, and the decision changes as a result, this can be an excellent example of Supporting Behavior which can be identified even without parsing the contents of text messages in the chat window. If a team member seeks to discuss the lead input, but the team proceeds with the decision without further review, then this may potentially be an indicator to track for a cumulative effect on Team Orientation if it happens repeatedly (i.e., if team members are being disregarded). Each of these rules is designed to be implemented with relatively simple logic to process the discrete inputs from team members and produce teamwork markers.

GIFT Interoperability

Initial prototyping of the wargaming trainer started with a simple infrastructure where GIFT is used to manage participant profiles and logins with assigned roles within the staff team. From the GIFT lobby the exercise is launched and participants interact with the exercise environment in browsers, with exercise flow and data managed by the RAAFT server. In the previous paper describing the system design (Jensen et al., 2021), we outlined a plan to use the GIFT survey mechanism to deliver prompts throughout the exercise. This included a notional message passing infrastructure between the RAAFT server used to deliver training content to each training participant's browser and GIFT survey utilities, through the Gateway Module to the Tutor Module. However, as details for the nature of prompts evolved beyond the initial design, this made it more complex to preserve the mechanics for switching control from the server-driven exercise interface to GIFT surveys, and also to present the multi-part prompts needed in exercises. Especially with the need to coordinate team member responses within the exercise flow, the updated design plan shifts the implementation of prompts to be entirely within RAAFT client-server components.

By the same token, the RAAFT server also executes the logic for analyzing team inputs and identifying conditions to produce teamwork markers. Team member inputs in the browser client exercise environment are processed and characterized on the RAAFT server to generate teamwork markers, while markers coming from human inputs can also be created manually. Both are recorded on the RAAFT server exercise log, and both are also treated as state transition events in terms of the design for exercise data to be pushed to the GIFT Domain Module. Initially, the primary instructional strategies triggered by these transitions on the GIFT side is to record the teamwork markers for AAR. However, it remains an area of further investigation to determine if there are benefits for generalization, reuse, or authorability if more of the assessment logic can be performed in the GIFT Domain Module. Instead of limiting the outgoing data to the teamwork markers already identified at the RAAFT server, an alternative option to be considered is to create more data throughput with more discrete predicates for user input events in the trainer environment, to be processed by condition class logic in GIFT Domain Knowledge Files to generate markers.

Similarly, we plan to also investigate further about the application of instructional strategies during exercises, and the possible benefits of a more modular function with GIFT. Even operating under the constraints of the browser-based version of a wargaming exercise designed for the initial prototype, there are certain circumstances where immediate feedback or an adjustment in the direction of decision-making is necessary. For example, in order to reduce branching during COA analysis, scenarios are constructed to follow a planned sequence of COA events. If the exercise team makes an early decision that would require a significantly different path in subsequent phases of the battle with the COA under consideration, then some feedback to the training audience or redirection may be needed, beyond the simple assessment function of generating and logging a teamwork marker. In such cases, under the current design, logic based on scenario markup produces a message shared privately first with the team leader (COS), who can then relay to the rest of the team. Using the earlier example with the helicopter attack, if route A were to create serious continuity problems for the remainder of the exercise, then if the team chooses route A, the COS is given a message to relay explaining the problems with route A and the need to settle on a different route. This relay convention also helps preserves natural team relationships with the COS. Under the initial design, this logic for potentially controlling the scenario is implemented on the RAAFT server, but there may be conditions where it is more generalizable to move the logic to instructional strategies in the GIFT Domain Module, using more data throughput with predicates for exercise events.

CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

Many factors make automated and adaptive training methods complex for domains involving team decision-making. Often in team domains, the constructs for competencies and how they are used as the structure for assessment and feedback are different than in individual training use cases. This paper describes work in progress and examples of some of the assessment mechanisms for training command staff teams in wargaming exercises, with several unique factors that impact how assessments are initially designed and used. One characteristic of the initial training application is that it's not designed around a persistent competency model for specific teams assembled for an exercise. That is, the purpose of assessment and the review of teamwork markers generated in an exercise is limited to the scope of the exercise itself. Starting with this initial narrow purpose, the natural next step for future development is to expand the use case to situations where the same team composition engages in a series of exercises, where the goal is to see improvement among the team as a result of repeating the cycle of conducting exercises and team review.

Another unique factor is that the initial intention is not to produce granular scoring on competencies associated with specific tasks. In a broad sense, the competencies are linked to six dimensions of teamwork organized in a model as mentioned earlier. When teamwork markers are identified during an

exercise as a result of automated assessments and/or input from human observers or participants, they are tagged with relevant team dimensions in the model. But the purpose of the markers is more to spur discussion and team self-correction, and less to be used as tallies summed for scores associated with the team dimensions as discrete competencies. One reason for this is that instructors do not have an immediate need for system-generated scores; within the training curriculum the purpose of the exercise is more as practice and less as a scored event. Another reason is the fact that teamwork markers are produced from a combination of automated and human sources, which may skew formulas that would attempt to aggregate results. However, this is another area planned for further investigation, to explore post-processing rules that can analyze the total picture of teamwork markers produced from an exercise not only to yield well-founded high-level scoring outcomes, but also to facilitate the team AAR. It is expected that exercises will produce more teamwork markers than the training audience would want to directly review in an AAR. In order to facilitate team self-review discussion in AAR, the purpose of post-processing rules is to highlight trends and the most significant or meaningful specific markers from the exercise. AAR feedback is designed to allow the training audience to further explore and browse any of the teamwork markers from the exercise, but a distilled AAR highlighting key topics for team discussion is critical. Future research will continue to explore these areas, not only for the specific wargaming training application but also for more general methods that can be applied to other team decision-making domains.

ACKNOWLEDGMENTS

The research reported in this document was performed in connection with contract number W912CG-20-C-0014 with the U.S. Army. The views presented in this paper are those of the authors and should not be interpreted as presenting official positions, either expressed or implied, of the U.S. Government.

REFERENCES

Grand, J.A., Pearce, M., Rench, T.A., Chao, G.T., Fernandez, R., & Kozlowski, S.W. (2013). Going DEEP: guidelines for building simulation-based team assessments. BMJ Qual Saf. 2013 May;22(5):436-48. doi: 10.1136/bmjqs-2012-000957. Epub 2013 Jan 25. PMID: 23355693.

Jensen, R., Presnell, B., DeFalco, J., and Goodwin, G. (2021). Designing a Distributed Trainer Using GIFT for Team Tutoring in Command Level Decision Making and Coordination. *In Proceedings of the 9th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym9).* 176-185.

Johnston, J. H., Cannon-bowers, J. A., & Salas, E. (1998). Tactical Decision Making Under Stress (TADMUS): Mapping a program of research to a real world incident—The USS Vincennes.

Kozlowski, S. W., Grand, J. A., Baard, S. K., & Pearce, M. (2015). Teams, teamwork, and team effectiveness: Implications for human systems integration.

Marlow, S. L., Lacerenza, C. N., Paoletti, J., Burke, S., & Salas, E. (2018). Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes*, 144, 145–170. https://doi.org/10.1016/j.obhdp.2017.08.001

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? Small Group Research, 36(5), 555–599.

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. *International Journal of Artificial Intelligence in Education*, 28(2), 225–264. https://doi.org/10.1007/s40593-017-0146-z

Teo, G., Jensen, R., Johnston, J., DeFalco, J., & Goodwin, G. (2021). Measures for Assessing Command Staff Team Performance in Wargaming Training. *Proceedings of the Interservice / Industry Training, Simulation, and Education Conference (I/ITSEC 2021).*

ABOUT THE AUTHORS

Randy Jensen *has been a project manager for over 20 years at Stottler Henke Associates, Inc., in San Mateo, California, where he has led projects to develop intelligent tutoring systems for domains involving complex decision-making. Examples include trainers for air attack planning, unmanned vehicle command and control, small unit tactics, combined arms team exercises, information systems troubleshooting, satellite scheduling, and a current effort for division level command group wargaming. Randy has a B.S. with honors from Stanford University.*

Evan Finnigan *is a software engineer at Stottler Henke Associates Inc., where his work includes developing intelligent tutoring software and building interactive web-based user interfaces. Previously, Evan was a student researcher in a human robot interaction lab where he built web-based interfaces to control an assistive robot. Evan has a B.S and M.S. from the University of California at Berkeley.*

Grace Teo, Ph.D. *is a Senior Research Psychologist at Quantum Improvements Consulting. Grace's research involves understanding and improving human performance under various conditions and in different contexts such as working with different technologies, and in teams. Other research interests include assessments, decision making processes and measures, vigilance performance, human-robot teaming, automation, and individual differences.*

Gregory Goodwin *is a senior research scientist with the U. S. Army Combat Capability Development Command Soldier Center at the Simulation and Training Technology Center (CCDC-SC-STTC), in Orlando, Florida. For the last decade, he has worked for the Army researching ways to improve training methods and technologies. He holds a PhD in Psychology from Binghamton University and an M.A. in Psychology from Wake Forest University.*