Team Performance in Dynamic Settings: Evaluating Shared Team Mental Model Similarity & Accuracy

Prosper A. Kwei-Narh
Inland Norway University of Applied Sciences

Teamwork in dynamic conditions relies on accurate/similar shared task mental models. Consensus exists in extant literature that shared task mental model (STMM) accuracy, rather than STMM similarity is important for team performance. This study examines the conceptual and measurement issues underlying these earlier findings and tested the complementary role of both STMM accuracy and similarity to team performance in a real world simulation of dynamic teamwork among forty-four (44) 3-member teams. The discussion focused on the implications of these findings firstly to the measurement of shared task mental models, and to future conceptualization of shared task mental models.

Keywords: shared team mental model, paired ratings, team performance

INTRODUCTION

Within the rapidly changing teamwork context, team members' ability to anticipate the needs of their teammates and of the task is sometimes the difference between success and failure (Cannon-Bowers, Salas, & Converse, 1993). Teams can use shared team mental models (STMM); cognitive structures that enable team members to anticipate and adapt to their teammates and to coordinate their task (Cannon-Bowers et al., 1993; Orasanu & Salas, 1993)—to increase their effectiveness in dealing with unexpected challenges and to enhance their team performance. There exists two aspects of shared team mental models; a shared task-knowledge mental models and a shared team-knowledge mental model (Cannon-Bowers et al., 1993; Wildman et al., 2012). Shared task mental models focus on the knowledge structures necessary for the task performance such as overall task objectives, the tools necessary for task performance and knowledge of procedures for successful task activities. Shared team mental models reflects knowledge structures regarding the characteristics and qualities of one's teammates such as the roles they perform on the team (Cannon-Bowers et al., 1993; Orasanu & Salas, 1993). Both shared task and team mental models are important in relatively enduring teams (Thomson, Levine, & Messick, 1999), however for newly established teams, it is shared task mental models which emerges quickest and can be proximally linked to team performance (DeChurch & Mesmer-Magnus, 2010b). On the other hand, shared team mental model emerges over the course of time as teammates build upon the general interactions around teammates' roles, individual attributes, and task preferences over the task interactions. For the purpose of short-lived teams, it is more relevant to focus on shared task mental models, and this is the focus of the following discussion.

After a team evolves a STMM, we can evaluate the extent of STMM *similarity* and STMM *accuracy*. STMM similarity describes the degree of convergence in task relevant knowledge content and understanding (Cannon-Bowers et al., 1993; Rentsch & Klimoski, 2001). Team members may have developed an overlapping structural representation of the basis of the unique task knowledge requirement for their teammates. Likewise, STMM accuracy reflects the extent to which the STMM is correct according to expected standard of task operations or the mental model of a subject matter expert (e.g. Smith-Jentsch, Campbell, Milanovich, & Reynolds, 2001), in other words '*accuracy*' describes the correctness of STMM content (Mohammed, Ferzandi, & Hamilton, 2010; Mohammed, Klimoski, & Rentsch, 2000). STMM similarity and accuracy may correlate (e.g. Lim & Klein, 2006) but in some cases can be independent of each other (Mohammed et al., 2010). In essence, a STMM can either be accurate or similar, both accurate and similar, or neither accurate nor similar and thus have varying impact on team performance. In Lim and Klein's (2006) study of recruits from the Singaporean army, for instance, STMM accuracy and similarity correlated together and both had positive impact on team performance. Obviously, we will avoid a situation where STMM are neither accurate nor similar. However, STMM may be highly accurate but not similar when teammates understand their individual tasks but do not adequately coordinate their task communication to allow for the development of STMM similarity. This can happen in a multi-functional team without adequate coordination (e.g. Korb, Geißler, & Strauß, 2015).

Whereas numerous studies have consistently reported a significant positive relation between STMM accuracy and team performance (e.g. Cooke et al., 2003; Lim & Klein, 2006; Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005), the relation between STMM similarity and team performance have not had such unanimous empirical support (e.g. Betts & Hinsz, 2013; Sander, van Doorn, van der Pal, & Zijlstra, 2015). This is remarkable because STMM similarity should be very important under conditions in which there is a high need for coordination among team members with disparate roles, which pertains in most teamwork. Under these task conditions, STMMs similarity will enable team members in 'organizing and understanding (task relevant) phenomena' (Rentsch & Davenport, 2006, p. 406) in the same way. That is to say that when individuals each deal with their unique task roles, there is the tendency to develop mental models unique to the task the individual is playing. However during teamwork, the individual must not only have their unique mental models of their work but the team together must be able to recognize ways in which the various elements of their task must be approached. This suggests that without STMM similarity, team members are more likely to devote distinct attention to some aspect of the team task and ignore some other aspects (Korb et al., 2015; Levesque, Wilson, & Wholey, 2001), especially the aspect that impinges on the collective team. STMM similarity should also be important when the task performance depends not only on the core routines of the task but in understanding elements of the task environment which directly impinges on the task routines (Levy, 2005). These serves to indicate that STMM similarity should complement STMM accuracy in predicting team performance.

Two underlying issues may account for the perplexing findings regarding the relationship between STMM accuracy and team performance on one hand, and STMM similarity and team performance. This has to do with the measurement/elicitation of mental models to assess STMM accuracy; and secondly the conceptualization of STMM accuracy and similarity. Firstly, in regards to measurement/elicitation, different elicitation strategies—the process of capturing information about mental models—obtains different information about mental models (Resick, Murase, et al., 2010). As the knowledge content within our mental models exists as a distribution of relationships; causal or covariational (Mohammed et al., 2000), the elicitation strategy and whether or not we seek for STMM accuracy or similarity plays an important role. With respect to elicitation techniques, a meta-analysis of the mental model literature suggested that the mental model measurement strategy—i.e., elicitation of content to obtain accuracy and representation of structure to describe similarity—may capture unique aspects of cognition (DeChurch & Mesmer-Magnus, 2010b). That is, STMM accuracy captures a specific individual mental model content whereas STMM similarity capture the associative network of the mental model content, involving both interpretation and perceptual cognitions (Rentsch, Delise, & Hutchison, 2012). To illustrate, security operatives understand that to successfully prevent infiltration of territories by hostile agents, they must understand the characteristics of the hostile agents and be aware of their motivations. However, whereas one team member might be responsible for acquiring information on hostile agents, another team member might be responsible for eventually engaging these agents. Both team members will accurately recognize the central elements of their teamwork (and thus score high on accuracy), but when they have to consider their roles in relation to the general objectives (i.e. association of their roles to the overall objectives), they may seem to have lower STMM similarity. These observations are based on the fact that similarity judgments requires subjects to consider the whole range of relationships within the mental model content whereas accuracy considers the successful recall/recognition/awareness of mental model content. It means that by eliciting and representing information about STMM similarity for instance, we might be capturing broader knowledge scope and expectation regarding the underlying relational structure between the knowledge components that might not be the same for STMM accuracy.

Additionally, the presence of a quality criterion (SME-subject matter experts' judgment) in assessing STMM accuracy and not STMM similarity may account for the disparate findings. That is because STMM accuracy measurement involves a quality criterion, e.g. subject matter experts (SME) whereas similarity indices often do not include a quality criterion (for instance subject matter expert's similarity judgments: exception, Mathieu et al., 2005). Since SME represent information in a qualitatively different way from that of novices (Hoffman, Shadbolt, Burton, & Klein, 1995; Mathieu et al., 2005), we can directly link SME knowledge content and the structure of their knowledge content to relevant team outcome criteria, in this case performance. It is not that the STMM similarity in the team bears no relation to the team performance but rather the representation of the knowledge and the consideration of the relation between different knowledge content might be the main issue. Therefore, to ensure a clear understanding of the complementary roles of STMM accuracy and STMM similarity in team performance, it is important to create similar quality criteria for both STMM similarity and accuracy as an initial step.

This paper sets out to answer what is the complementary role of STMM accuracy and similarity to team performance in dynamic settings. In addition, the paper seeks to understand the utility of STMM similarity in predicting team performance when a quality criterion is part of the measurement index. Answering these questions will contribute to our theorizing of whether indeed STMM accuracy and similarity capture unique forms of cognition. In addition, it answers how both STMM accuracy and similarity complement each other in predicting team performance. Practically, this study should be useful in highlighting the factors that are important for assessing STMM accuracy and similarity and ways in which we can facilitate team knowledge frames to marshal STMM accuracy and similarity in enabling team performance. Knowing that STMM accuracy and similarity are important to the team performance, we have to consider how we structure knowledge content in order to make knowledge associations easier to form. This can enable training interventions designed to improve the mental models of team members, focusing on both the content of training and the knowledge organization and representation within the team.

THEORY

The Importance of STMM Accuracy and Similarity to Team Performance

A team is defined as a group of two or more people who must interact cooperatively and adaptively in pursuit of shared objectives (Cannon-Bowers et al., 1993). Teamwork—the interactions that enable the accomplishment of task objectives within a team setting (Salas, Shuffler, Thayer, Bedwell, & Lazzara, 2015)—varies according to the dynamic environment of the team (Sundstrom, de Meuse, & Futrell, 1990). These environments can range from high-risk settings such as the military, air-traffic controllers, and surgical teams to lower-risk settings such as a team responsible for a student project report (Sundstrom, 1999). In all these teams, there is the need for the team members to anticipate their teammate's needs, the dynamic requirement of their task, and unexpected scenarios in the team's task. Teams—especially those in high-risk environments such as action teams (Cannon-Bowers et al., 1993)—need some degree of STMMs in order to enhance their team performance.

Team performance reflects an evaluation of how well a team obtains their goals (DeChurch & Mesmer-Magnus, 2010a) by engaging in the appropriate behaviors and operations demanded by the task.

Objective performance is captured in this paper because it minimizes the problem of rater bias when subjective performance is used (DeChurch & Mesmer-Magnus, 2010a). Team performance depends on how knowledge that is critical to teamwork is organized, represented, and used within the team (DeChurch & Mesmer-Magnus, 2010a; Kozlowski & Ilgen, 2006). This suggest the role of STMMs—developed through a team member's interactions with their task tools and when teammate's share knowledge about their task. For instance, the lack of shared knowledge within a Navy crew resulted in a shipping disaster in Norway (Norwegian Official Reports, 2000).

STMM accuracy is a quality criterion that reflects the extent to which the team members' task knowledge differs from/reflect the accepted standard of the task performance (Ellis, Bell, Ployhart, Hollenbeck, & Ilgen, 2005). High accuracy is the degree to which the knowledge content of the subject matches the knowledge content in the defined criterion. For instance, Lim and Klein (2006) used subject matter experts and assessment center ratings to compute the STMM accuracy of army officers. Participants and subject matter experts judged the relatedness of pairs of statements on a Likert scale. STMM accuracy was the degree to which teams deviate from SME's rating, where lower deviation reflects higher accuracy. In another study, Resick and colleagues (2010) had participants perform a search and capture operation and used a paired-rating of critical tasks within the simulation to obtain STMM. The STMM accuracy was the match of the team to a team of expert's mental model. STMM accuracy presupposes that each member of the team has all the information needed to complete the task (Hollenbeck, Colquitt, Ilgen, LePine, & Hedlund, 1998). This implies that team members with less STMM accuracy may reflect cognitive ability issues (e.g. Edwards, Day, Arthur, & Bell, 2006) or the need for extra training (e.g. Cooke et al., 2003; Smith-Jentsch et al., 2001). When we know a team's level of STMM accuracy, we will thus predict the possible task performance potential of a team and can usefully diagnose the training needs of a poor performing team.

In that direction, STMM accuracy is a useful predictor of team performance in various settings. For instance, Cooke and colleagues (2003) found a positive predictive ability of taskwork positional accuracy on team performance. In Resick and colleagues (2010) study described above, STMM accuracy moderately predicted goal accomplishment but significantly predicted perceived coordination. A field study conducted by Mathieu and colleagues used paired ratings to examine the effect of task mental model quality (or accuracy) on team performance on a synthetic team task. Results showed a positive effect of task mental model accuracy on team performance. Combining these and other studies (see DeChurch & Mesmer-Magnus, 2010a, 2010b) leads to the hypothesis that:

Hypothesis 1: *STMM accuracy will predict team performance.*

STMM Similarity and Team Performance

STMM similarity reflects the extent to which STMM structure overlaps among team members (Mohammed et al., 2000), such that teammates can use similar knowledge structures for organizing and understanding task-relevant phenomena (Rentsch & Davenport, 2006). STMM similarity focuses on the internal representation and relation within knowledge content. STMM similarity indices reveal the extent to which different team members possess such an overlapping mental model representation—in some cases agreement. Team members can interpret internal and external events using the same related meanings (Rentsch et al., 2012). STMM similarity is strictly a group level variable which further constrains the individual teammate's understanding of the task (Betts & Hinsz, 2013). This is because STMM similarity enables teammates to interpret, organize, and integrate task knowledge in the same way (Rentsch et al., 2012).

Similarity indices reflect the underlying assumption teammates possess of the relation between constructs (in this case task-relevant knowledge) using the considerations described by Mohammed and colleagues (2000)—causal, co-occurring, dependent, or contingent. For instance, assume in a three-member team, Member A judges two concepts to be similar because he/she thinks there is a causal relation among the concepts, whereas Member B and C judges the two concepts to be related because they occur together (co-vary). We may obtain a high mental model similarity index that does not reflect the individual teammates' underlying assumption regarding the relation between the concepts (e.g. Healey, Vuori, & Hodgkinson, 2015). In that sense, mental model similarity may differentially influence team outcomes (e.g. Mathieu, Rapp, Maynard, & Mangos, 2009; Sander, van Doorn, van der Pal, & Zijlstra, 2015). Such a dynamism in shared mental model similarity has led Sander and colleagues (2015) to question the importance of shared mental model similarity in predicting team outcomes. Consider the case in which a random set of two teams were asked to judge the relatedness of a number of task constructs, which is the usual practice in STMM similarity assessments (DeChurch & Mesmer-Magnus, 2010b). Whereas all the teammates in one team may consider the constructs to have weak causal linkage and therefore rate them as unimportant, another team may consider all the constructs to have strong causal linkage and rate them as important. In both cases, the ratings of STMM similarity will be high but the team's behavior may be different because of each team's underlying evaluation of the relation within the mental models. Therefore, STMM similarity is important for diagnosing not only the structural organization of knowledge but also the likelihood that the team will engage in well-coordinated behavior.

A number of studies have reported the positive influence of STMM similarity on team performance. Kellermanns, Floyd, Pearson, and Spencer (2008) used content analysis to examine resource allocation priorities in a field study using university faculty. Their study revealed a significant effect of STMM similarity on decision quality. Mathieu and colleagues (2005) also examined STMM similarity in an air flight simulation study where participants rated the perceived relation between task-relevant knowledge attributes. Their study revealed a significant effect of STMM similarity on team performance. In a field study using army recruits, Lim and Klein (2006) examined STMMs using paired rating of the team's procedures, equipment, and task. Their study revealed a significant effect of task mental model similarity on team performance. These studies and theorizing leads to the formulation of the hypothesis:

Hypothesis 2: *STMM similarity will predict team performance.*

Generating an Account of Why TMM Similarity and Accuracy Differentially Relate to Team Performance

The nature of taskwork in dynamic conditions provides the need for STMMs that deals with the core task procedures, task strategies—that is STMM accuracy—and in addition, a STMM similarity that enables different teammates to collectively agree on the components, scenarios, and task strategies. Thus STMM accuracy is important for the core content of the task whereas the STMM similarity enables agreement about the how different aspect of the task are important in order to ensure effective performance within dynamic settings. Therefore, STMM accuracy and similarity are conceptualized as tapping unique cognitive content yet complementarily important for team performance.

Considering accuracy and similarity are properties of the same phenomenon—in this case STMM—it is an important observation that they could be capturing unique aspect of cognition (Hoffman et al., 1995) and will therefore have different implications for team performance (DeChurch & Mesmer-Magnus, 2010b). However, in dynamic settings where teammates are performing multiple roles, both STMM accuracy and similarity are important in predicting team performance. We first have to consider the reason why STMM accuracy and similarity will be capturing unique knowledge and yet why they are together essential in team performance within dynamic settings.

There are empirical, conceptual, and measurement issues that point to why STMM accuracy and similarity might be capturing unique cognition and why the focus have been on STMM accuracy. Firstly, similarity depends on shared features of the team task, for example similar role in the team (Korb et al., 2015) and on background characteristics such as cognitive ability (Edwards et al., 2006) and the use of similar tools (Bolstad, Schneider, Graham, & Gonzalez, 2004). It means that during the teamwork, teammate's who already have the background for integrating new knowledge rapidly develops a similar mental model (Allen & O'Neill, 2015) and may remain stable over the course of the team interaction (Cooke et al., 2003).

An additional explanation for the unique relation of task mental model similarity and accuracy on team performance is conceptual. This involves the nature of structural knowledge representation such that knowledge elicitation techniques may only capture a snapshot of a larger knowledge base responsible for task performance (Rouse & Morris, 1986). According to DeChurch and Mesmer-Magnus (2010b, p. 3), "similarity ratings typically prompt respondents to think in terms of the degree of association between distinct components of their team or task. In this way, they (i.e. similarity ratings) capture associative networks of knowledge" within the task mental models. On the other hand, when we elicit mental models of an individual in order to compute their accuracy, we are tapping directly to a narrow and limited knowledge component considered relevant for the task. What we obtain is an 'approximation' of knowledge content considered relevant for the task performance (Rouse & Morris, 1986). We often do not access a wider range of mental representations of the task as would have been the case if the elicitation is to compute task mental model similarity.

Another difference between STMM accuracy and similarity could also be that similarity measurement compels us to compare both the content, the relationship between the content, and the ordering/ranking of the content simultaneously. That means that even if teammates have similar mental model content but they rank order these contents in different ways, an elicited similarity index may not be the same as an elicited content. Rentsch and Davenport (2006), for instance illustrated a knowledge structure that considers the task strategies for scoring and handling opponents in a sports team. Although the knowledge content is the same, the ordering and prioritization of the content will depend on the skill level/task understanding of the individual teammate and the particular role on the team. Thus the STMM similarity or accuracy measures will capture unique contents about the task mental models in the team.

Considering these conceptual and measurement issues, STMM accuracy is likely to be a stronger predictor of team performance than STMM similarity because mental model elicitation methods may not be strong enough to fully represent the nature and hierarchy of associations in mental models necessary for team performance (when using the similarity rating approach). In addition, the associative network may contain elements that are related together but each individual team member may have given more consideration to different elements of that relation, i.e. considering causal or covariation within the elements (Healey, Vuori, & Hodgkinson, 2015). To assess whether this prediction holds, I include a quality criterion—similarity of the team to the similarity indices involving subject matter experts.

Hypothesis 3: *STMM accuracy and STMM similarity will complement each other in predicting team performance better than STMM accuracy/similarity alone.*

METHOD

The hypotheses for this study were tested in a novel environment which demands STMM. In addition, we need a task whose performance can be objectively measured and where we can identify the 'best' or optimal task strategies (Mathieu et al., 2005). This will allow us to compare the effect of STMM accuracy and similarity (with a quality criterion included) to the team performance.

Participants, Procedure, and Design

One hundred and thirty two (132) business students from the largest business school in Norway participated in the study. The sample was randomly assigned to 44 teams, each with 3 members, and data were collected over three scenarios providing 369 observations. Biographical data were collected using Qualtrics (Qualtrics, Provo, UT) a web-based data collection tool. Fifty three percent (53%) of the sample was male, the average age was 21.24 (min = 18, max = 38) and three people did not provide any information about their gender or age.

The simulation commenced with a 20-minute presentation on the general purpose of the study and the ethical considerations including informing participants of their rights to withdraw from the experiment and data anonymity during analyses. Included in the presentation was an in-depth description of the simulation, and the roles as members of a three-person team. Participants did not know the identity of their teammates until they were going to sit at their cubicles in the lab. This is to make the teammate's interaction reflect newly established teams and an organic mode for STMM development. A measure of familiarity with teammates was included to check for any previous professional or private exchanges among teammates. Teammates did not sit adjacent each other but rather in different parts of the lab. Each computer is located within separate partition of the lab, so they only get a brief acquaintance to each other just before they occupied their positions. All team communication is via email.

The Simulation Task

To simulate specialized teamwork within dynamic settings, I used Mindlab, which is a dynamic distributed decision-making simulation modelled along the work of Hollenbeck and colleagues (2002) but modified to make the setting more relevant and familiar to the research participants. This task is a team-based micro-world simulation (Brehmer & Dörner, 1993) in a setting with specialized roles and high interdependence between individual team members. Team members had to coordinate each other's work to protect oilrigs in the North Sea of Norway from potential terrorist attacks. Teams were required to detect, identify, and eliminate potential enemy vessels as contained in pre-programmed mission orders. To initiate movements and other user-defined actions, the players uses right and left click mouse combinations, and this was common for all team members.

The simulation involved team members occupying specialized roles identified as Orion, Patrol, and Frigate. These roles can perform the following functions respectively; search, detect, and attack objects that threatens sensitive locations. Orion is a maritime surveillance aircraft, which specializes in detecting objects within a 200km radius. Orion can select to patrol within a specific geographic area (See Figure 3 in Appendix for a screenshot of the graphical user interface of the simulation environment). As Orion navigates within the simulation environments, its radar sensors detect objects. Information about the description of these detected objects appears as message pop-ups on the screen. Orion can decide to follow an object and must maintain the object within its sensor range, ensuring the object is visible and allowing the teammates to conduct further intelligence/interrogative actions on the object.

Patrol is a boat with slower movement and comparatively weaker sensor capabilities than Orion. It specializes in 'information search', which is correctly classifying an object either as a security threat or as a friendly vessel that can be allowed to proceed on its course. Information search is carried out in the action routine of point and click system, Patrol can approach a vessel and interrogate the vessel for information to determine whether the vessel is threatening or not. Message pop-up notifications provide a real time update on the vessels and intra-team communication.

The third team role, Frigate, represents a navy ship specialized to engage vessels determined to be threatening to the rigs. Frigate can use its weapons to neutralize threatening objects within a 40km radius.

At the beginning of the study, there was a training period of 20 minutes where the participants experienced the functionality of the game. Thereafter, I conducted the simulation over three scenarios, with some variations in the starting location of the vessels over successive scenarios in order to minimize learning effects. Scenarios lasted for 20 minutes during which I captured data on team communication and task performance. At the end of each scenario, Qualtrics was used to administer a survey containing STMM items. The third scenario contained additional scales measuring task load, experience playing simulation or computer games.

Measures

Team Performance

The performance measure was made up of three components: attack score, speed, and accuracy of positioning at the end of the scenario. Attack points are scores obtained by Frigate, and this was assigned to all the team members. The speed and accuracy of positioning were unique to team members. The team performance was the aggregate of this score; average is 57.08 (SD = 31.06; range = -18.82–118.19).

Shared Task Mental Models (STMM) Accuracy

STMMs seek to describe, explain, and predict situations (Cannon-Bowers et al., 1993; Rouse & Morris, 1986) and I developed a number of questions that was presented to participants. STMM within this setting is the extent to which teammates understood the task they performed, predict possible scenarios given the dynamics of their task environment, and could describe the actual dynamics of the tasks they performed. To develop the questionnaires, I used 'theoretical modeling' (Rouse & Morris, 1986), where I relied on theory and data to describe the most important content of task knowledge in our setting. I used the questionnaire approach to elicit mental models (DeChurch & Mesmer-Magnus, 2010b).

I developed eleven (11) multiple-choice questions after a task analysis and SME discussions. These questions tapped task relevant information needed for optimal team performance (e.g. what type of activities were required to eliminate hostile vessels), and team members' awareness of the operating environment (e.g. the location of enemy vessels). Correct responses obtained a score of one (1) point. Wrong responses were scored zero. Questions with ordinal or interval scales could be more or less correct, coded from zero to one [0-1] with 0.25 intervals. A question on the geographical location of an object was coded as 1 for exact correct location, 0.75 if close to correct location, and 0 if far removed from the correct location. Some questions required participants to have knowledge of the number of objects their team attacked or detected.

After this scoring, I submitted the 11 items to a partial least squares analysis to assess the nature of the items as a formative measure (Chin, 1998; Hair, Ringle, & Sarstedt, 2011) using SmartPLS version 3.2 software (Ringle, Wende, & Becker, 2015). I assessed reliability by demonstrating that the items had no collinearity and the indicator weights are significant (Sarstedt, Ringle, Smith, Reams, & Hair Jr, 2014). Firstly, non-collinearity was demonstrated by "running a multiple regression of each indicator of the formatively measured construct on all the other measurement items of the same construct" (Sarstedt et al., 2014, p. 109). I use the R^2 value of the regression to compute the collinearity scores, using the VIF estimates. All eleven indicators demonstrated VIF values ranging from 1.02 to 1.28 (mean VIF =1.08), thus satisfying the non-collinearity requirement. Secondly, I demonstrated the indicator significance of the measures using the bootstrapping procedure with 5000 resamples. I choose to have individual sign changes, and the confidence interval estimation was achieved by means of a two-tailed studentized bootstrap. The bootstrapping method assesses the significance and relevance of the formative indicators to the construct. Poorly weighted and insignificant indicators were removed from the measurement (Chin, 1998). After this decision point, eight items were considered as meeting the requirement for inclusion into the final STMM accuracy scale (M = 5.7, SD = 1.05). I kept these scores at the individual level without summarily adding everything together to the team level because of the advantages of multilevel latent modelling (described in detail) which I used in the analyses.

STMM Similarity

Measures of STMM similarity have traditionally used a diverse range of indices such as the closeness of individual networks to each other (DeChurch & Mesmer-Magnus, 2010b). I chose the R_{wg} index (James, Demaree, & Wolf, 1984; Levesque et al., 2001; Lindell, Brandt, & Whitney, 1999), where the lowest score is 0 and the maximum score is 1. R_{wg} was used because it describes the extent of within- group agreement on the features of the task. Comparing team members with differentiated roles suggests assessing the level of perceptual agreement; for instance, Levesque and colleagues (2001) used R_{wg} indices to capture the level of agreement among software development teams.

Knowledge content areas relevant for the simulation after a careful task analysis are given as Attack, Identify, Search, and Follow (see Appendix for their explanation). This yielded a 10*9 matrix and I used the upper triangular portion for ratings (45 pairs). Thereafter, the paired rating strategy was used (e.g. Edwards et al., 2006) to elicit the STMMs. This involves participants rating the importance of the pairs of actions to the team performance on a nine-point Likert scale, ranging from 1 = not at all related to 9 = highly related.

After the participants rated the 45 pairs, three subject matter experts (SMEs) also rated the pairs from the most important to the least important (Cronbach's α = 0.69). These SMEs ratings reflect an attempt to ensure the similarity ratings are treated on the same criterion as accuracy scores. This serves as a measure of the quality of the STMM similarity which is a needed improvement on existing measures (Mohammed et al., 2010). I adopted a convention to choose SMEs ratings with the highest possible values (i.e. 7-9) since those ratings capture the most distinct and important pairs (e.g.Eisenhardt, 1989). That is by using extreme scores, I create a profile where teammates display the same underlying agreement on the paired ratings (in terms of whether there is underlying covariation or correlation between paired terms). Fourteen new pairs were extracted from these SMEs ratings.

Two different agreements indices resulted. One is the degree of agreement across the team for each paired ratings. The other is the degree of agreement for all paired ratings across the whole team. The latter would have been arrived at when we take the mean across all ratings for each team. Although this was according to tradition, recent advances in data analyses has provided us with powerful tools like multilevel modelling which can generate group level estimates without directly aggregating the individual ratings (see detail below). I chose the multilevel modeling approach and therefore did not perform a direct aggregation of the paired ratings. The agreement indices for all pairs within the whole team is the STMM similarity. The teams' r_{wg} was computed on these fourteen new pairs of ratings to form STMM similarity.

Strategic and Computer Game Experience

Some of the actions and maneuvering during task performance can be influenced by participants' prior experience. I developed a nine-item measure of experience playing computer games. The participants responded to this item on a 7-point Likert scale. The items have high internal consistency (Cronbach's $\alpha = 0.88$). This was therefore aggregated to the group level (M=1.88, SD=0.42) as a measure of experience.

Data Analytic Approach

I used a multilevel approach to test hypotheses related to this study. Multilevel analyses is preferred in cases where observations are non-independent (Bauer, Preacher, & Gil, 2006) caused in this case by the repeated measurements and individuals nested within teams. To enable the partitioning of variance into individual and group level variance, it is suggested to use the intra-class correlation estimates (Muthén, 1997). These ICC estimates indicates the degree of nestedness of the variables due to individuals belonging to the same team. According to LeBreton and James (2008, p. 823), ICC (1) values can be interpreted as an "effect size estimate" revealing the extent to which judges' ratings were affected by their membership within the team. Estimated intra class correlations are shared team mental model accuracy (ICC (1) = 0.43) and shared team mental model similarity (ICC (1) = 0.57). I used the multi-item option of the interrater agreement index (LeBreton & Senter, 2008), since I treated observations on each scenario as separate observations of the same construct. Estimates of coefficients are assessed significant at $p < 0.05$.

Hypothesis 1 will be supported when there is a significant positive relation between STMM accuracy and team performance, controlling for experience. Hypothesis 2 will be supported when there is a significant positive relation between STMM similarity and team performance, controlling for team experience. Hypothesis 3 will be supported if dominance analysis (Luo & Azen, 2013) indicates that the predictors can predict team level variance as well as team level scores using the $\boldsymbol{R^2}$ estimates of Raudenbush and Bryk (2002) and Snijders and Bosker (1994). Dominance analysis is "is concerned with rank ordering the predictors in terms of relative importance by comparing the additional contributions the predictors make to variance reproduced (or explained) by all possible subset models (consisting of subsets of the predictors)" (Luo & Azen, 2013, p. 5).

RESULTS

Data Structure

Table 1 presents the summary of descriptive statistics, where experience was positively correlated with STMM accuracy ($r = 0.14$, $p<0.01$) and team performance ($r = 0.24$, $p<0.01$), but not STMM similarity ($r = 0.03$, p = n.s.). There was a negligibly and weak correlation between STMM accuracy and similarity ($r = 0.07$, p = n.s.), a rudimentary indication that the measures capture unique knowledge. STMM accuracy correlated significantly with team performance ($r = 0.68$, $p<0.01$), and a significant positive relation existed between STMM similarity and team performance ($r = 0.22$, $p<0.01$).

Tests of Hypotheses

A series of two-level random intercept models were fitted with individuals nested within teams using Mplus (Muthén & Muthén, 1998-2012). To allow for a meaningful interpretation of the intercepts, STMM accuracy was grand mean centered (Enders & Tofighi, 2007) whiles STMM similarity was not centered because it was a team level variable. Grand mean centering yields scores that are correlated with variables at both lower and upper levels (Enders & Tofighi, 2007) and adjusts for any individual level differences on the predictor variables (Heck & Thomas, 2015).

An initial model with no predictors was estimated to assess the distribution of variances within and between levels (Snijders & Bosker, 1999). Table 2 below presents the results of the multilevel random intercept estimation using Mplus. No individual level predictors were included in the model, so the between level variances remained unchanged. On the other hand, the between-level variance reduced from 331 to 277.24 when the control variable, experience playing computer games was included. Prior experience was significantly related to performance (β = 17.24, p <.05). STMM accuracy also significantly predicted team performance (β = 19.06, p <0.01) thus supporting hypothesis 1. In the subsequent model (Model 3 in Table 2), I estimated a separate regression equation for STMM similarity, controlling for experience. The effect of STMM similarity on team performance was significant (β = 80.94, p <0.01), thus, hypothesis 2 was supported. To test hypothesis 3, I used dominance analysis as summarized below.

Relative Importance of the Predictors Using Dominance Analysis

I used dominance analysis (Azen & Budescu, 2003; Luo & Azen, 2013) using two measures of $\boldsymbol{R^2}$—variance explained in the team level scores (Snijders & Bosker, 1994) and the variance explained within the team (Raudenbush & Bryk, 2002). According to Luo and Azen (2013), when we are interested in explaining the variance of the random cluster effects, Raudenbush and Bryk's (2002) $\boldsymbol{R^2}$ should be used. It is recommended to use Snijders and Bosker (1994) $\boldsymbol{R^2}$ when we are interested in determining the relative importance of the level-2 predictors in predicting cluster mean scores. Because the estimated models did not include any level 1 predictor, both $\boldsymbol{R^2}$ estimates were the same but their interpretations are different. A change in the between level $\boldsymbol{R^2}$ value using Raudenbush and Bryk (2002) formula indicates the amount of additional team level variance explained by the addition of the predictor. On the other hand, a change in the between level $\boldsymbol{R^2}$ value using Snijders and Bosker's (1994) formula estimates the amount of additional variance in cluster means that the predictor variable explains (Luo & Azen, 2013).

Our $\boldsymbol{R^2}$ estimates indicated that STMM accuracy uniquely explains 57% in its model whereas STMM similarity uniquely explains 14% in its model, and the total model with all predictors explained 74% of variance (note also that since I estimated R^2 using Snijders and Bosker (1994) formula, this value represents the predicted group scores of the model), thus supporting hypothesis 3a. For hypothesis 3b, a dominance analysis will compare variance explained (and scores predicted) by different subsets of the model predictors (Luo & Azen, 2013). This dominance analysis (Table 3) indicated that STMM accuracy dominates STMM similarity. Hypothesis 3b was thus supported.

TABLE 1
DESCRIPTIVES AND ZERO-ORDER CORRELATIONS FOR STUDY VARIABLES

	Mean	***SD***	***1***	***2***	***3***
Measure of experience	1.88	0.43			
STMM accuracy	5.52	1.02	0.14**		
STMM similarity	0.73	0.09	0.03	0.07	
Team performance	57.08	31.06	0.24**	0.68**	0.22**

N=132; *p < .05. **p < .01,

TABLE 2
PREDICTING TEAM PERFORMANCE SCORES

	Model O		*Model 1*		*Model 2*		*Model 3*		*Model 4*	
	Unstd. Coeff.	***SE***	***Unstd. Coeff.***	***SE***	***Unstd. Coeff.***	***SE***	***Unstd. Coeff.***	***SE***	***Unstd. Coeff.***	***SE***
Intercept (γ_{00})	57.08	3.02	57.08	2.81	57.08	2.20	57.08	2.64	57.08	2.02
Experience			17.24*	5.37	10.70*	4.97	16.75**	5.16	10.37*	4.36
STMM accuracy					19.06**	3.51			18.69**	3.07
STMM similarity							80.94**	27.77	75.32**	23.31
Level 1 variance	623.23	108.49	623.39	108.52	632.27	108.47	632.40	108.52	623.11	108.42
σ^2_u ***(level 2 variance/residual variance)***	331.17	69.38	277.24	70.20	123.27	40.10	233.36	61.99	85.91	33.85
R&B R_2^2			0.16		0.63		0.30		0.74	
S&B R_2^2			0.16	0.10	0.62*	0.12	0.29*	0.12	0.74**	0.10
AIC	3760.407		3756.109		4744.182		2898.644		3880.676	

Note: *$p < .05$. **$p < .01$, *** p<.001, [a] p<.10 , ***Unstd Coeff***: unstandardized coefficient estimate. ***R&B***: Raudenbush and Bryk; ***S&B***: Snijders and Bosker. Note also that R^2 should be understood as explained model variance, not total variance.

TABLE 3
DOMINANCE ANALYSIS TO ESTIMATE THE IMPORTANCE OF STMM ACCURACY AND SIMILARITY IN PREDICTING VARIANCE AND SCORES IN TEAM PERFORMANCE

		Additional contribution of		
Subset Model	***Model R^2***	**Experience**	**TMM Accuracy**	**TMM Similarity**
Experience	0.16			
STMM accuracy	0.57	0.41		
STMM similarity	0.14	0.02	0.43	
Exp., STMM accuracy	0.63	0.47		0.49
Exp., STMM similarity	0.29	0.13	0.28	
STMM accuracy, STMM similarity	0.69	0.53		
ALL predictors	0.74	0.58		

Note: Exp.: is experience. STMM: Shared task mental model.

DISCUSSION

The aim of this study was to examine and evaluate the complementary role of STMM accuracy and similarity within dynamic team settings. In addition, I examined which of STMM accuracy and similarity acts as a better predictor of team performance scores and variations around the scores. This study underscored the findings that STMM accuracy and similarity each uniquely and significantly predicted team performance scores. The results indicated that STMM accuracy—capturing unique task knowledge content—and STMM similarity—capturing the associative network of task knowledge are important to predict team performance scores. I detail below the relevance of these findings.

It is a robust finding that STMM accuracy predicts team performance (DeChurch & Mesmer-Magnus, 2010a), and this study added further evidence. Indeed, there was a strong correlation between STMM accuracy and team performance, and STMM accuracy significantly predicted team scores. The results also indicated the role of STMM similarity to explaining team performance as indicated by the significant R^2 scores. When both STMM accuracy and similarity are considered together, the initially significant residual variance became insignificant, indicating that both aspects of STMM needs to be considered simultaneously in such a dynamic setting as was the focus of the study. The results indicate that teams who can evaluate and agree on the most important aspect of their task can quickly respond to their task needs and engage in the desired task behaviors. In the following, I examine the implications of these findings to shared mental model theory and the practice to improve shared task mental models.

Implications of Findings to Theory

The implication of this study to the shared team mental model literature is as below. Firstly, there is a direct implication to measuring shared mental model similarity. Previous studies that have examined and found no support for the relation between STMM similarity and team performance may have overlooked an important aspect of cognitive representation during teamwork involving multiple roles. When individuals are performing different roles, it is highly likely that their cognitive representation may relate more to their individual roles rather than the general task objectives of their team. It means that even if they have the same cognitive representations of task elements, the nature of the linkages among these representations will vary. It means that the network measures may not be appropriate for measuring the team's task mental model similarity. Rather, the extent to which the team agrees on the relation between task-relevant knowledge is important. The agreement indices do not focus on the strength of the relation within the task knowledge content but the recognition that those elements of the task are important. Since action teams perform in a dynamic environment, agreement on the necessary element of the task helps teammates to collectively respond to and anticipate the effect of this task element in a similar manner.

Another important issue is the consideration of what it means to elicit shared mental model representations of teammates, versus eliciting shared mental model representation of teammates compared with a subject matter expert. When we elicit the associative network of knowledge within a team, the level of expertise in the team as well as teammate's expectation of what the researcher wants affects the elicited content (Rouse & Morris, 1986). We know that even on the same task, novices and experts represent information in very different ways. The combination of these factors means that we need a quality criterion that aligns both the expectation among the team members and the representation of the mental models content. This is very important to consider because research which questioned the importance of mental model similarity to team outcomes (e.g. Sander et al., 2015) may have underestimated the impact of the mental model measurement strategy on the elicited mental models (Mohammed & Hamilton, 2012).

The distinction between accuracy and similarity is long standing (DeChurch & Mesmer-Magnus, 2010b; Mohammed et al., 2000). This conceptual distinction may have stalled researchers from exploring how both STMM accuracy and similarity can aid team performance in dynamic team settings. In these team settings, the interdependency between team members means it is not enough that teammates understand their individual tasks but that teammates share agreement on and evaluate their task strategies and procedures in the same way. This paper, unlike previous studies which found no significant effect between STMM similarity and team performance (Edwards et al., 2006) has shifted the focus back on the consideration that without a similar way of conceptualizing and processing knowledge, the team performance may be impeded. This sets the stage to examine models in which STMM accuracy may be considered as the most important element of the task. This study indicates that in dynamic conditions, it is not enough to understand the task but that the team must also possess agreement on the most important element of their task to direct their focus and response. This joint consideration—of STMM accuracy and similarity—was the best predictor of team performance scores.

Implications of Findings for Practice

The findings have important implications for diagnosing performance deficiencies in emergency response teams especially in cases where intervention can be aided by experts mental models (Smith-Jentsch et al., 2001). Emergency response teams often have members playing different roles and therefore, they may represent information differently. However, there are important aspects of the task upon which they have to recognize and agree. When training points teammates towards these task objectives, team members are in a better position to engage in the behaviors that will enhance their performance (Smith-Jentsch et al., 2001). For optimal knowledge organization in order to enhance team performance, we need to pay attention to the underlying ways in which the knowledge elements are interrelated. For instance, when the knowledge organization involves considering how knowledge elements are causally related, we may obtain better agreement within the team. These results indicate that although we regularly use experts to evaluate the content of knowledge, the ability of teams to recognize and agree on experts' knowledge will predict their team performance.

Limitations and Proposal for Future Studies

One of the major limitations of this study is that estimating multilevel relationships with fewer than 50 groups can be problematic. To compensate for this situation, I included measurement over three-time situations that enlarged the dataset (i.e. 132 observations on 44 teams). This should enable stability in the estimation of standardized errors. The conclusions drawn from this study should bear this limitation of sample size in mind.

As a laboratory simulation, this study lacks generalizability to other settings. However, the central elements of the study are directly applicable to emergency teams where specialization and disparate pieces of information are the norm. The expectation is such that we can find similar patterns of results within settings where team members play differentiated roles but must engage in mutually dependent adaptive performance. In such settings such as rescue operations and surgical teams, teammates may receive disparate information but their action impinges on the rest of the team. The STMM accuracy and similarity is necessary to predict team performance.

Previous studies have examined the STMM similarity and accuracy using the same sample of items. The advantage of that approach is that we can directly compare the influence of STMM accuracy and similarity to each other. The challenge with that approach is that there could be multiple mental models and by limiting the range of considerations, these approaches may have underestimated both the prevalence and need for STMMs within their settings. I, therefore, elected to broaden the scope of the STMM considered in this setting, to capture both the content and the organization of a broad set of variables within the study setting. The obvious advantage is that the models have large predictive power (74%), but the disadvantage is that it may not reflect existing approaches to capturing both STMM similarity and accuracy in the same study. Nevertheless, the growing literature seems to suggest that using multiple metrics is valuable in providing a broader understanding of team performance in most settings (DeChurch & Mesmer-Magnus, 2010b).

CONCLUSION

This paper answered the call for research which "explore multiple methods for operationalizing mental models within the same sample of teams" (DeChurch & Mesmer-Magnus, 2010b, p. 9).This paper contributed in that direction by showing that both indices of STMMs capture unique aspects of cognitions. I also assessed dominance as the predictive capability of different subsets of the models in predicting team performance scores and variations between teams on their performance. The main message is that in dynamic settings, team members need to have both an understanding of their task and agree on the most important content to organize and understand task phenomena.

REFERENCES

Allen, N. J., & O'Neill, T. A. (2015). The Trajectory of Emergence of Shared Group-Level Constructs. *Small Group Research*, 46(3), 352-390. doi:10.1177/1046496415584973

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8(2), 129-148. doi:10.1037/1082-989X.8.2.129

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 11(2), 142-163. doi:10.1037/1082-989X.11.2.142, 10.1037

Betts, K. R., & Hinsz, V. B. (2013). Strong shared representations promote schema-consistent memory errors in groups. *Group Processes & Intergroup Relations*, 16(6), 734-751. doi:10.1177/1368430213486206

Bolstad, C., Schneider, M., Graham, J., & Gonzalez, C. (2004). Collaborative Tools and Shared Mental Models. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(3), 610-613. doi:10.1177/154193120404800371

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2–3), 171-184. doi:http://dx.doi.org/10.1016/0747-5632(93)90005-D

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 221-246). New Jersey: Lawrence Erlbaum.

Chin, W. W. (1998). The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-336). Mahwah, N.J.: Lawrence Erlbaum.

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, 7(3), 179-199. doi:10.1037/1089-2699.7.3.179

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010a). The Cognitive Underpinnings of Effective Teamwork: A Meta-Analysis. *Journal of Applied Psychology*, 95(1), 32-53. doi:10.1037/a0017328

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010b). Measuring Shared Team Mental Models: A Meta-Analysis. *Group Dynamics-Theory Research and Practice*, 14(1), 1-14. doi:10.1037/a0017455

Edwards, B. D., Day, E. A., Arthur, W., Jr., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, 91(3), 727-736. doi:10.1037/0021-9010.91.3.727

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review,* 14(4), 532-550.

Ellis, A. P., Bell, B. S., Ployhart, R. E., Hollenbeck, J. R., & Ilgen, D. R. (2005). An evaluation of generic teamwork skills training with action teams: Effects on cognitive and skill-based outcomes. *Personnel Psychology*, 58(3), 641-672.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi:10.1037/1082-989X.12.2.121, 10.1037

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2), 139-152. doi:10.2753/MTP1069-6679190202

Healey, M. P., Vuori, T., & Hodgkinson, G. P. (2015). When teams agree while disagreeing: Reflexion and reflection in shared cognition. *Academy of Management Review*, 40(3), 399-422. doi:10.5465/amr.2013.0154

Hoffman, R. R., Shadbolt, N. R., Burton, A. M., & Klein, G. (1995). Eliciting Knowledge from Experts: A Methodological Analysis. *Organizational Behavior and Human Decision Processes*, 62(2), 129-158. doi:http://dx.doi.org/10.1006/obhd.1995.1039

Hollenbeck, J. R., Colquitt, J. A., Ilgen, D. R., LePine, J. A., & Hedlund, J. (1998). Accuracy decomposition and team decision making: Testing theoretical boundary conditions. *Journal of Applied Psychology*, 83(3), 494-500. doi:10.1037/0021-9010.83.3.494

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

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98. doi:10.1037/0021-9010.69.1.85

Kellermanns, F. W., Floyd, S. W., Pearson, A. W., & Spencer, B. (2008). The contingent effect of constructive confrontation on the relationship between shared mental models and decision quality. *Journal of Organizational Behavior*, 29(1), 119-137.

Korb, W., Geißler, N., & Strauß, G. (2015). Solving challenges in inter- and trans-disciplinary working teams: Lessons from the surgical technology field. *Artificial Intelligence in Medicine*, 63(3), 209-219. doi:http://dx.doi.org/10.1016/j.artmed.2015.02.001

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. doi:10.1111/j.1529-1006.2006.00030.x

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.
Levy, O. (2005). The influence of top management team attention patterns on global strategic posture of firms. *Journal of Organizational Behavior*, 26(7), 797-819. doi:10.1002/job.340
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.
Lindell, M. K., Brandt, C. J., & Whitney, D. J. (1999). A Revised Index of Interrater Agreement for Multi-Item Ratings of a Single Target. *Applied Psychological Measurement*, 23(2), 127-135. doi:10.1177/01466219922031257
Luo, W., & Azen, R. (2013). Determining Predictor Importance in Hierarchical Linear Models Using Dominance Analysis. *Journal of Educational and Behavioral Statistics*, 38(1), 3-31. doi:10.3102/1076998612458319
Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: equifinality and normative comparisons. *Journal of Organizational Behavior*, 26(1), 37-56. doi:10.1002/job.296
Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor No More: A 15-Year Review of the Team Mental Model Construct. *Journal of Management*, 36(4), 876-910. doi:10.1177/0149206309356804
Mohammed, S., & Hamilton, K. (2012). Studying team cognition: The good, the bad, and the practical. In A. Hollingshead & M. S. Poole (Eds.), *Research Methods for Studying Groups* (pp. 132-153). New York: Taylor and Francis.
Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The Measurement of Team Mental Models: We Have No Shared Schema. *Organizational Research Methods*, 3(2), 123.
Muthén, B. O. (1997). Latent variable modeling with longitudinal and multilevel data. In A. E. Raftery (Ed.), *Sociological methodology* (Vol. 27, pp. 453-480). Boston: Blackwell Publishers.
Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, R. Calderwood, & C. Zsambok (Eds.), *Decision making in action: models and methods* (pp. 327-345). Norwood, N.J.: Ablex Pub.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : applications and data analysis methods* (2nd ed. ed., Vol. 1). Thousand Oaks: Sage Publications.
Rentsch, J. R., & Davenport, S. W. (2006). Sporting a new view: Team member schema similarity in sports. *International Journal of Sport and Exercise Psychology*, 4(4), 401-421. doi:10.1080/1612197X.2006.9671805
Rentsch, J. R., Delise, L. A., & Hutchison, S. (2012). Cognitive similarity configurations in teams : in search of the Team MindMeld TM. In E. Salas, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: cross-disciplinary perspectives and approaches* (pp. 241-266). New York: Routledge.
Rentsch, J. R., & Klimoski, R. J. (2001). Why do 'great minds' think alike?: antecedents of team member schema agreement. *Journal of Organizational Behavior*, 22(2), 107-120.
Reports, N. O. (2000). *The grounding of fast coastal passenger MS Sleipner grounding 26 November 1999*. In M. o. J. a. t. Police (Series Ed.) N. O. Reports (Ed.) (pp. 31).
Resick, C. J., Mitchelson, J. K., Dickson, M. W., Allison, L. K., & Clark, M. A. (2010). Team Composition, Cognition, and Effectiveness: Examining Mental Model Similarity and Accuracy. *Group Dynamics-Theory Research and Practice*, 14(2), 174-191. doi:10.1037/a0018444
Resick, C. J., Murase, T., Bedwell, W. L., Sanz, E., Jiménez, M., & DeChurch, L. A. (2010). Mental model metrics and team adaptability: A multi-facet multi-method examination. *Group Dynamics: Theory, Research, and Practice*, 14(4), 332-349. doi:10.1037/a0018822

Ringle, C. M., Wende, S., & Becker, J-M. (2015). *SmartPLS (Version 3.0)*. Hamburg: SmartPLS. Retrieved from www.smartpls.com

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(3), 349-363. doi:10.1037/0033-2909.100.3.349

Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., & Lazzara, E. H. (2015). Understanding and Improving Teamwork in Organizations: A Scientifically Based Practical Guide. *Human Resource Management*, 54(4), 599-622. doi:10.1002/hrm.21628

Sander, P. C., van Doorn, R. R. A., van der Pal, J., & Zijlstra, F. R. H. (2015). Team adaptation to an unforeseen system failure: Limits of the potential aids of shared knowledge and standardized communication. *European Journal of Work and Organizational Psychology*, 24(5), 796-811. doi:10.1080/1359432X.2015.1006199

Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F., Jr. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. *Journal of Family Business Strategy*, 5(1), 105-115. doi:http://dx.doi.org/10.1016/j.jfbs.2014.01.002

Smith-Jentsch, K. A., Campbell, G. E., Milanovich, D. M., & Reynolds, A. M. (2001). Measuring Teamwork Mental Models to Support Training Needs Assessment, Development, and Evaluation: Two Empirical Studies. *Journal of Organizational Behavior*, 22(2), 179-194.

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled Variance in Two-Level Models. *Sociological Methods & Research*, 22(3), 342-363. doi:10.1177/0049124194022003004

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: Sage.

Sundstrom, E., de Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, 45(2), 120-133. doi:10.1037/0003-066X.45.2.120

Sundstrom, E. D. (1999). The challenges of supporting work team effectiveness. In E. D. Sundstrom (Ed.), *Supporting work team effectiveness: best management practices for fostering high performance* (pp. 3-23). San Francisco, Calif: Jossey-Bass.

Thompson, L. L., Levine, J. M., & Messick, D. M. (1999). *Shared cognition in organizations: the management of knowledge*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Wildman, J. L., Thayer, A. L., Pavlas, D., Salas, E., Stewart, J. E., & Howse, W. R. (2012). Team Knowledge Research: Emerging Trends and Critical Needs. *Human Factors*, 54(1), 84-111. doi:10.1177/0018720811425365

APPENDIX A

FIGURE A1
EMAIL INTERFACE.

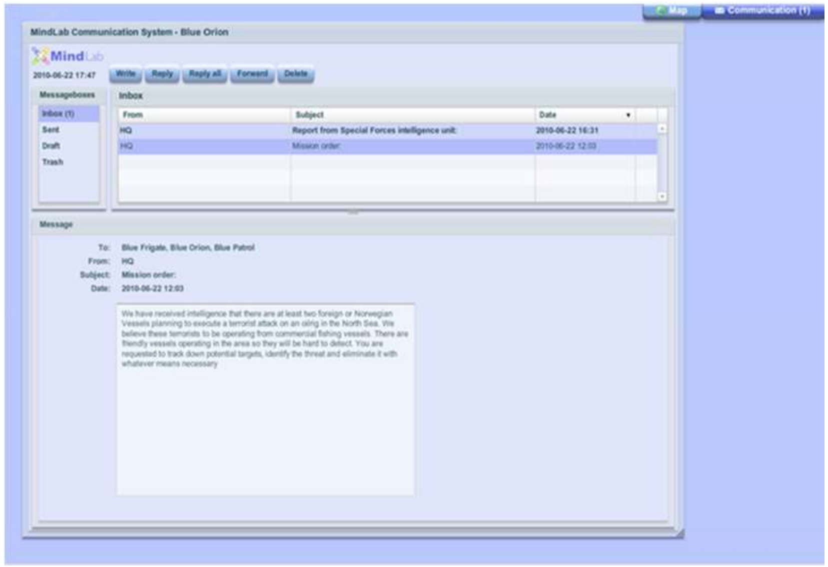
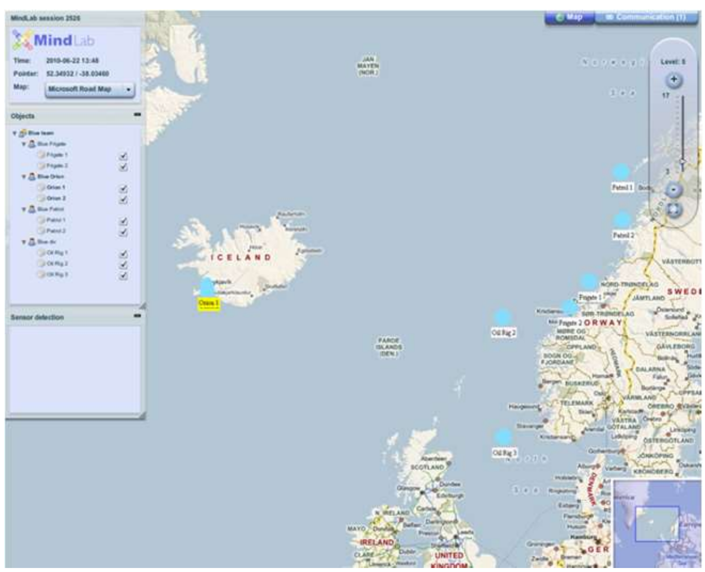


FIGURE A2
MAP INTERFACE



APPENDIX B

TABLE A1
OVERVIEW CAPACITY PER ROLE IN THE TEAM

	Orion	Patrol	Frigate
Sensor range (km)	200	50	40
Max speed (km/h)	500	120	80
Detection capacity	High	Medium to low	Low
Information search capacity	-	High	Low
Ability to attack	-	-	High

APPENDIX C

Instructions for paired ratings of STMM similarity

You are to make judgments about the "relatedness" of pairs of statements describing actions on the game you just played. There are several ways you might think about relatedness of the terms being judged. For instance, two terms might be related because they describe the behaviors that are necessary to perform the task. For this task, think about the statement as they relate to playing the game. On the columns, you will find the full description of each term:

PLEASE FOLLOW THE FOLLOWING STEPS CAREFULLY
Read through all the definitions of the dimension listed on the left hand side before entering any information
Insert a number into the grid that ranges from "1= not at all related" to "9 = highly related".
9: Extremely related: These actions are necessary together for the team to attack a terrorist
1: Not related: The two actions are not together necessary for you team performance
4: Somewhat related: These actions are important but does not have to be done together for your team to perform well.
It is easier to complete the boxes at the extreme ends of the scale FIRST (i.e. Extremely Related and Not Related), then complete those in the middle.

	Identify	Search	Follow	Write & Send	Maneuver	Intel	Inform	Estimate	Surveillance
Attack *Intercept enemy target from reaching restricted area*									
Identify Identify object as friendly or enemy									
Search Conduct search for information on target									
Follow Follow target identified as enemy									
Write & Send write message to team member									
Maneuver The opportunity to use your vessels to undertake the right strategy for the task									
Intel Forward intel messages to team-members									
Inform Inform about the progress of the task and enemies									
Estimate Estimate location of your terrorist based on knowledge of map									
Surveilance Patroling the sensitive territories									

APPENDIX D

MENTAL MODEL QUESTIONNAIRES FOR COMPUTING ACCURACY

How many vessels did your team detect during the scenario?
What was the position of Oil Rig 4?
How many enemy vessels did your team detect during the scenario?
Which vessel did your team identify as enemy vessels?
Where were the enemy vessels located?
Which of the oil rigs was a target for terrorist attacks in the scenario?
From what you know, what is the likely outcome of a terrorist attack?
What kinds of operations are necessary for your team to successfully prevent a terrorist attack?

APPENDIX E

EXPERIENCE PLAYING COMPUTER GAMES

Please indicate how you experienced the following aspects of the simulation, on a scale ranging from 1 = no experience degree to 7 = extensive experience.

	1	2	3	4	5	6	7
Please indicate your experience playing strategy computer games,	☐	☐	☐	☐	☐	☐	☐
To what extent have you received training in crisis management (civilian or military)?	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Big Accidents)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Natural Disaster)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Border crossing)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Illegal fishing)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Terrorist attack)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Military crisis)	☐	☐	☐	☐	☐	☐	☐
How will you rank your expertise in crisis management with respect to the following situations and incidents? (Civilian crisis)	☐	☐	☐	☐	☐	☐	☐