Measuring teamwork mental models
to support training needs assessment,
development, and evaluation:
two empirical studies

KIMBERLY A. SMITH-JENTSCH*, GWENDOLYN E. CAMPBELL,
DANA M. MILANOVICH, ANGELIQUE M. REYNOLDS
†Naval Air Warfare Center Training Systems Division, Orlando, U.S.A.

Summary
The present paper reports data from two studies that utilized a card sorting approach to measuring mental model similarity in naturalistic training environments. Results from the first study indicated that higher ranking navy personnel held mental models of teamwork that were more similar to an empirically derived model of expert team performance than lower ranking personnel. Furthermore, comparisons of mental model similarity within groups of high and low ranking trainees and within groups of high and low experience trainees indicated greater similarity between those of higher rank and between those with greater experience. The second study tested the effects of a computer-based training (CBT) strategy that was designed to develop teamwork mental models that were more similar to the ‘expert model’ described in Study 1. Using the same card sorting approach, positive training effects were demonstrated on similarity to the expert model, similarity to other trainees, and consistency. Copyright © 2001 John Wiley & Sons, Ltd.

Introduction

In recent years, organizations have increasingly adopted team-based work structures to better utilize expertise, minimize the impact of increasing workload on one individual, and maximize the use of increasingly more sophisticated technology. This change in organizational structure has had a significant impact on both selection and training requirements, and has opened a new area of research. In particular, the recent literature on teams and team training has begun to identify the cognitive underpinnings of effective teamwork. It has been argued that when teammates hold similar team-related knowledge and knowledge structures, they are better able to anticipate the actions and information needs of their teammates, and to respond effectively (e.g., Blickensderfer et al., 1997 – Paper presented

*a Correspondence to: Kimberly A. Smith-Jentsch, Naval Air Warfare Center Training Systems Division (NAWCTSD), Air 4961, 12350 Research Parkway, Orlando, FL 32826-3275, U.S.A.

†The views expressed here are those of the authors and do not reflect the official position of the organization with which they are affiliated.

Copyright © 2001 John Wiley & Sons, Ltd.
at the 12th SIOP Conference, St Louis, MO; Cannon-Bowers et al., 1993; Kleinman and Serfaty, 1989; Orasanu, 1990 – Paper presented at the 34th meeting of the Human Factors and Ergonomics Society, Orlando, FL; Rouse et al., 1992; Smith-Jentsch et al., 1998b).

For the purpose of this research, we focused on one particular type of knowledge structure: mental models of teamwork. A mental model of teamwork was defined as an individual’s understanding of the components of teamwork that are critical for effective team performance, as well as the relationships between those components. For example, one’s mental model about teamwork may include the understanding that proper phraseology, brevity, and clarity are all positive aspects of effective communication that contribute to team success. Mental models of teamwork are expected to influence the manner in which individuals organize concrete observations of team behavior in their minds, assess the quality of those observations, and attribute underlying root causes to performance trends.

As defined here, mental models of teamwork are ‘teammate-generic’ (Cannon-Bowers et al., 1995) in that they pertain to the principles of effective and ineffective teamwork that transcend team membership. In this way, our definition of teamwork mental models is similar to that used by Rentsch et al. (1994) to describe ‘teamwork schemas’ and by Mathieu et al. (2000) to describe ‘team mental models’. It has been argued that because teamwork mental models guide the manner in which individuals perform their tasks and interact with one another, team members who hold similar mental models of teamwork are better able to coordinate with one another and thus achieve superior performance outcomes. This notion has been empirically supported in at least one published study (Mathieu et al., 2000). Additionally, teammates with similar mental models of teamwork are expected to be better able to engage in team self-analysis and participative goal setting (e.g., Smith-Jentsch et al., 2000).

While it appears to be important that teammates share similar teamwork knowledge, one would expect that the benefits of similar knowledge are dependent upon the degree to which this knowledge is accurate. Teamwork mental models can be considered accurate if they mirror actual relationships between specific teamwork behaviors and effective team performance in a particular team task environment. If team members share similar but inaccurate mental models about teamwork, they may select and strive towards goals that do not necessarily improve their performance. However, researchers have only just begun to measure and to parse the contribution of mental model similarity and accuracy to effective team performance (Mathieu et al., manuscript submitted).

The studies reported in this paper employed a relatively novel approach to measuring the similarity among individuals’ teamwork mental models and between their mental models and an expert model of teamwork (accuracy) for a specific team type. In Study 1, relationships among these measures and two indicators of domain-specific experience were examined in support of a training needs analysis. In Study 2, the same mental model measures were employed to evaluate the effectiveness of a computer-based training (CBT) tool designed to prepare individuals for team training opportunities.

**Experience-related Differences in Mental Model Accuracy**

Researchers have begun to examine individual characteristics that help to explain differences in team member knowledge and knowledge structures (Rentsch et al., 1994; Rentsch and Klimoski, 2001). One of the aims of such research is to target training strategies toward those most at risk for holding team knowledge that is inaccurate, and/or dissimilar to (and potentially in conflict with) that held by other teammates. For example, differences in teamwork knowledge are expected to result from individuals’ previous team experiences. As one acquires team-related experience, he/she is expected to develop an understanding about the components of effective teamwork by observing what works, what does not work, and the manner in which various teamwork behaviors co-vary across a variety of
situations. Thus, findings from the literature on expert–novice differences (e.g., Borko and Livingston, 1989; Chi and Koeske, 1982) would suggest that those with a great deal of team experience should hold teamwork knowledge that is abstract, consistent, and multilevel. In contrast, individuals who have little experience in team settings are expected to have simplistic and/or inconsistent teamwork knowledge representations that are organized around superficial cues.

In support of these notions, Rentsch et al. (1994) found that participants who reported high team experience tended to use fewer categories or dimensions to describe teamwork, used more abstract definitions, and represented their knowledge more consistently across two measurement methods than those reporting low team experience. These researchers suggested that training for low experience teammates could be developed to guide them in developing knowledge and knowledge structures that are more similar to those held by high experience teammates. This implies that high experience individuals hold teamwork knowledge that more accurately reflects the nature in which effective teamwork operates. However, this hypothesis was not empirically tested.

Previous research has suggested that the nature of effective teamwork will vary by team type (Sundstrom et al., 1990). This would imply that the accuracy of an individual’s teamwork mental model must be assessed relative to the actual components of effective teamwork within a particular team task domain. Furthermore, it is more likely that domain-specific team experience rather than general team experience would be related to the accuracy of teamwork mental models for any given team task domain. Thus, while Rentsch et al. (1994) investigated team experience across any and all team tasks (e.g., football team, organizational task force) and linked it to ‘core teamwork knowledge,’ they noted that the effects of team experience on teamwork knowledge for specific team types requires additional investigation.

The present research sought to extend the findings from previous research on teamwork knowledge by examining relationships between two domain-specific experience variables (i.e., length of time in the navy, and navy rank) and the degree to which an individual’s teamwork mental model was accurate for a particular team type and similar to the mental models held by others. Rank and length of time within a specific organization may be highly correlated in some cases. However, in many cases certain individuals enter a new organization or team at a mid- or high-level rank due to their education or special expertise. Regardless of the length of experience they have in a particular environment, higher ranking teammates generally serve leadership functions that require them to attend to more global issues such as evaluating and managing team processes. This is true within the military as well as in many business organizations. Thus, higher ranking individuals may develop more accurate mental models of teamwork than lower-ranking individuals due to the type of experience that comes with their team role, although they may have the same or lesser amounts of experience working in a particular type of team.

As an indicator of mental model accuracy in the present research, we assessed similarity between trainees’ mental models of teamwork and an empirically derived model of effective teamwork for navy command and control teams. This ‘expert model of teamwork’ was defined through the analysis of behavioral ratings assigned to teams of navy personnel who performed a series of simulated command and control scenarios (Smith-Jentsch et al., 1998a). A factor analysis of the ratings revealed that 11 teamwork behaviors clustered within four higher-order dimensions. Moreover, teams that had previously performed together for a significant period of time were found to receive significantly higher ratings on the four dimensions than did newly formed teams. As such, the expert model was deemed to accurately reflect the nature of effective teamwork for navy command and control teams. Our first two hypotheses stated:

**Hypothesis 1.** Similarity between an individual’s teamwork mental model and the expert model (i.e., mental model accuracy) would be positively associated with the length of time that he had been in the navy.

**Hypothesis 2.** Similarity between an individual’s teamwork mental model and the expert model (i.e., mental model accuracy) would be positively associated with that individual’s navy rank.
Experience-related Differences in Mental Model Similarity

Given that a single or relatively homogeneous set of accurate mental models exist in a particular task environment, greater similarity among individuals should increase with experience as they become more accurate. Moreover, the fact that low experience individuals tend to represent their teamwork knowledge less consistently (Rentsch et al., 1994) is likely to lead to greater measurement error and consequently lower similarity scores among less experienced individuals than is found among individuals with high team experience.

In this case, one would not necessarily expect to see a linear relationship between similarity of experience and similarity of teamwork mental models. Instead, those with similar low levels of experience and/or rank would be expected to report a dissimilar set of inaccurate mental models while those with similar high levels of experience and/or high rank would be expected to share similar set of accurate mental models. It follows that our third and fourth hypotheses stated:

**Hypothesis 3.** Teamwork mental models will be more similar among individuals with a high degree of experience in the navy (i.e., length of time in service) than among those with a low degree of experience.

**Hypothesis 4.** Teamwork mental models will be more similar among high-ranking individuals than among low ranking individuals.

Study 1: Investigating Training Needs

**Participants and procedure**

One hundred and seventy-six male personnel from a navy submarine community participated in Study 1. These participants were drawn from 15 submarine attack centre teams. Participants completed a demographic questionnaire and a measure of teamwork mental models on two consecutive days.

**Measures**

**Demographic data**
Participants completed a questionnaire that asked them to report their age, the number of months that they had been in the navy, and their current rank in the navy.

**Mental models of teamwork**
As mentioned previously, the present study adopted an expert model of teamwork that was derived through the analysis of performance ratings collected from navy command and control teams (Smith-Jentsch et al., 1998a). This model consisted of four dimensions defined by 11 component behaviors: information exchange (i.e., passing information, providing big picture summaries, seeking information from all available sources), communication (i.e., proper phraseology, brevity, clarity, completeness of standard reports), supporting behavior (i.e., error correction, back-up/assistance), and leadership (i.e., providing guidance, stating priorities).

A card sorting task was used to assess each participant’s mental model of teamwork. Each card listed a concrete example of either effective or ineffective teamwork that could occur in a submarine.
attack centre. These examples were generated by a group of subject matter experts. Three concrete examples of each of the 11 component behaviors within the expert model were generated. These examples were then printed on 33 numbered index cards. Participants were instructed to sort the examples into categories of teamwork that were meaningful to them and to label each of their piles. For example, card no. 4 contained the phrase ‘Sonar announced, ‘Weapon in the water’ instead of ‘Torpedo in the water’’ printed on it and card no. 6 had the phrase ‘The WCC announced, ‘Misfire’ instead of ‘Check fire.’’ These two examples are both representative of the teamwork component behavior ‘Phraseology.’

Participants’ data were then entered into a BASIC computer program that created a 33 × 33 matrix representing all possible combinations of the 33 index cards. A ‘1’ was placed in each cell where the corresponding cards were placed together in a single category. For example, if the 33 numbered cards had been sorted into four separate categories and one category contained only the no. 4 and no. 6 cards listed above, this particular card combination would be represented by a ‘1’ in the (4, 6) cell. All other cells in the rows and columns of the matrix headed by cards no. 4 and no. 6 would have a ‘0.’ Participants’ similarity to one another and to the expert model (accuracy) can be computed based on these matrices using the Phi coefficient, which represents the Pearson correlation coefficient between two dichotomous variables.

In order to obtain an expert matrix from which to score the accuracy of participants’ mental models, three researchers sorted the examples into piles that would be consistent with the expert model of teamwork. For instance, ‘The Fire Control Commander directed all members to focus on developing solutions for Sierra 4’ was considered a concrete example of ‘Stating priorities’ which is a component behavior of ‘Leadership’ within the expert model. The average correlation among the matrices produced by the three researchers was 0.93. After discussing discrepancies in how the cards were sorted, the three researchers came to consensus on a card sort that was considered to be consistent with the expert model. The matrix based on this card sort served as the expert matrix from which participants’ accuracy was scored.

Results

Demographic measures

Participants reported an average time in service of 81 months ($N = 153, SD = 63$) and an average age of 26 years ($N = 152, SD = 6$). Ranks were reported by titles (e.g., Seaman Apprentice, Senior Chief, Ensign, Lieutenant Commander). Rank titles were ordered sequentially (from Seaman Recruit to Captain) and assigned a numeric score. The average rank score was 6 ($N = 154, SD = 3$), which corresponds to a rank of E6 (First Class Petty Officer). As can be seen in Table 1, all three demographic measures (military rank, age, and time in service) were significantly positively correlated.

Similarity to expert model (accuracy)

A Phi correlation coefficient between each participant’s card sort matrix and the expert card sort matrix was calculated to assess the accuracy of each participant’s mental model of teamwork. On average, participants’ similarity ($N = 176$) to the expert model was 0.346 ($SD = 0.166$). Pearson’s product moment correlations between these accuracy scores and the three demographic measures were calculated. With respect to this set of variables, only the correlation between similarity to the expert model
of teamwork and navy rank was significant (see Table 1). Therefore, Hypothesis 1 was not supported.

In support of Hypothesis 2, those holding higher ranks held mental models of teamwork that were more similar to the expert model than those holding lower ranks.

### Similarity among participants

In order to investigate differences in similarity among high and low experience groups, we sought to split our sample into approximately the top and bottom 25 per cent on time in service and on navy rank. As explained above, the similarity between each pair of trainees within a group could be calculated using a $\Phi$ correlation coefficient. The goal was to generate an overall measure of similarity within each group. Calculating an average similarity score that included each trainee’s similarity to every other trainee is statistically problematic, because it mixes within and between subject data and thus violates the assumption of independence of observations. Therefore, we wrote a program in MatLab (1996) that applied a variation of bootstrapping similar to the approximate randomization technique, to estimate an empirical sampling distribution from our data (Lee and Rodgers, 1998; Rasmussen, 1989).

Specifically, for each group, we (1) randomly selected pairs of participants, without replacement, until we had exhausted the pool; (2) tallied their pair-wise similarity scores; and (3) calculated an average similarity score. Thus, each participant’s data were only used once in producing the average. However, because the particular (randomly selected) pairs of participants could skew this average, we replaced all of the participants and repeated the sampling procedure 100 times. This allowed us to calculate 100 overall group similarity scores for each of the two groups of participants. The mean of these 100 numbers provides a more stable estimate of the actual average within group similarity than any of the original scores.

### Time in service-based groups

There were 41 participants with 127 or more months in service (27 per cent of 153 participants) and 40 participants with 34 or fewer months in service (26 per cent of 153 participants). These two groups were chosen to represent participants with ‘high’ and ‘low’ experience, respectively. The average age of those in the high time in service group was 34.1 years ($SD = 3.9$) and the average age of the low time in service group was 21.4 years ($SD = 2.0$).

Our next step was to statistically compare the similarity among participants in the high time in service group to the similarity among participants in the low time in service group. Several different techniques have been proposed and evaluated using Monte Carlo studies for performing inferential statistical analyses on bootstrapped data (e.g., Lee and Rodgers, 1998; Rasmussen, 1989). The approach taken in this research was to test our hypotheses using one-tailed independent $t$-tests for each of the 100 bootstrapped samples, and then apply a chi square goodness-of-fit test to determine if the number of significant results was more than would be expected by chance (5 out of 100). The average
similarity score across the 100 resamples for the high time in service group was 0.278 (SD = 0.024), and the average similarity of the low time in service group was 0.203 (SD = 0.027). In support of Hypothesis 3, a comparison of similarity among high and low time in service groups across the 100 resamples produced 18 significant t-tests, which was significantly more than would be expected by chance, $\chi^2(1) = 35.58, p < 0.01$.

**Rank-based groups**

There were 41 officers (27 per cent of all participants) and 38 enlisted men in ranks E1–E5 (25 per cent of all participants). These two groups were chosen to represent participants with ‘high’ and ‘low’ ranks, respectively. The average age of the high ranking group was 28.2 years (SD = 4.0) and 26.7 years (SD = 3.8) for the low ranking group.

Following the same procedure described previously, each group was sampled 100 times to determine an empirical distribution of similarity scores. Across all 100 resamples, the average similarity among participants in the high rank group was 0.367 (SD = 0.028) and the average similarity among participants in the low rank group was 0.286 (SD = 0.025). There were 16 significant t-tests out of 100, which was significantly more than would be expected by chance, $\chi^2(1) = 25.47, p < 0.01$. Thus, in support of Hypothesis 4, higher ranking individuals held mental models of teamwork that were significantly more similar to one another than did lower ranking individuals.

**Discussion: Study 1**

As hypothesized, greater similarity among individuals was found within high ranking relative to low ranking groups and within high time in service as compared to low time in service groups. Additionally, those with higher rank held teamwork mental models that were more similar to the expert model. In contrast, length of time in service was not significantly related to mental model accuracy. For practical purposes, results from Study 1 suggested that it was appropriate to design training that guided navy personnel in adopting the expert model as a common frame of reference for thinking about teamwork. Teams within a command and control environment vary widely in rank. Thus, since rank was associated with both accuracy and similarity among individuals, results from Study 1 suggested that a self-paced strategy incorporating individualized feedback was appropriate for preparing teammates to participate in training strategies involving team self-correction and goal setting.

**Study 2: Training Mental Models of Teamwork**

The purpose of Study 2 was to examine the effectiveness of a CBT tool designed to guide individuals in adopting the expert model of teamwork. Furthermore, we sought to investigate the validity of the card sorting methodology for predicting individuals’ ability to apply the expert model. This is important given that assumptions regarding individuals’ need for training, as well as assumptions regarding the effectiveness of the CBT, were being based on scores of mental model accuracy using this approach. Thus, our fifth hypothesis stated:

*Hypothesis 5.* Accuracy scores on the measure of teamwork mental models will be significantly correlated with participants’ ability to generate their own concrete examples of situations that illustrate components of the expert model.
On the basis of results from Study 1, the CBT tool was already in use by such individuals in a variety of field settings as part of an interactive workshop for intact teams (Smith-Jentsch et al., 1999). Study 2 was set up so as to test the CBT tool alone as a strategy for preparing individuals to perform or train with a new or existing team. If the CBT were truly effective at influencing mental models of teamwork, and not simply producing rote memorization of declarative knowledge, one would expect trainees to demonstrate the ability to generalize and discriminate among concepts within the expert model regardless of superficial features specific to a particular task domain. Thus, participants in Study 2 trained using examples from task domains that were different from the task domain in which they typically worked. Furthermore, the concrete examples within the card sorting measure were drawn from a task domain that was different from those presented in the CBT. Our remaining hypotheses stated:

**Hypothesis 6.** The CBT tool will produce mental models of teamwork that are more similar to the expert model.

**Hypothesis 7.** Groups of trainees will hold mental models of teamwork that are more similar to one another after training than before training.

**Hypothesis 8.** Individuals will represent their teamwork mental models more consistently across two administrations of the card sorting measure after training than before training with the CBT.

**Participants and procedures**

As a matter of face validity, we encountered resistance to having members of navy command and control teams train using concrete examples that were not specific to their particular task domain. Thus, participants in Study 2 were 42 civilian government employees (e.g., engineers, trainers, research psychologists) involved in the development, acquisition, and implementation of military training systems. Twenty-nine of the participants were male, while 13 were female. These individuals are frequently called upon to form short-term teams on an ‘as needed’ basis, and to apply their technical expertise toward solving training problems within a number of navy task domains (e.g., aviation, shipboard, subsurface). Thus, the face validity of using concrete examples from a variety of task domains to learn the expert model was not a problem in this field environment.

**Experimental conditions**

Participants were randomly assigned to one of three experimental conditions: pretest/post-test treatment group ($N = 21$); double pretest control group ($N = 11$); and double post-test treatment group ($N = 10$). Participants were scheduled such that card sort administrations were always separated by one day. Additionally, training on the CBT and post-training card sorts were always separated by one day. The self-generated free response measure was always administered on the same day as the last card sort administration.

**Analysis plan**

The double pretest and double post-test groups were included to rule out the possibility that trainees’ similarity to the expert model (i.e., accuracy) may have improved simply as a result of taking the card sorting measure twice. Specifically, one would hope to see significant increases in similarity to the expert model between the first and second administration of the card sort only for the one group that received training between these two administrations (i.e., pretest/post-test treatment group). Additionally, those in the double pretest and double post-test groups were used to test Hypothesis 8 which stated that trainees would represent their teamwork mental models more consistently after training than before training with the CBT.
Computer-based training

All participants received one-hour of training using the CBT tool. Participants received this training individually and thus, progression through teamwork examples was self-paced. The CBT programme presented participants with concrete examples of effective and ineffective teamwork from three military domains and allowed them to practice categorizing examples within the expert model. As each example appeared on the screen, participants were asked to click on one of four buttons representing the four higher order teamwork dimensions (e.g., supporting behavior). Once a higher order dimension had been selected, a drop down menu containing the component behaviors associated with that dimension appeared (e.g., error correction, backup/assistance). Next, participants selected one of the component behaviors. If the correct behavior had been selected, participants automatically received another teamwork example to classify.

When an incorrect response was selected, the CBT programme presented the trainee with the correct response and clarified the distinction between two. For example, if a trainee misclassified an example of passing information before having to be asked (a component of information exchange) as an example of completeness of reports (a component of communication), the trainee would receive the following feedback: ‘The term passing information before being asked refers to instances where a team member anticipates another team member’s need for information and passes that information to him/her without having to be asked. This involves knowing when and to whom a piece of data should be passed. In contrast, the term completeness of report refers to following standard procedures that indicate which pieces of information are to be included in a particular type of report and in what order.’

The CBT was populated with examples that had been generated by subject matter experts in the areas of combat systems, aviation, and seamanship. All participants first began classifying examples from a combat systems domain. Once they completed all examples from this domain, they proceeded to train on examples from an aviation domain. Trainees that completed examples from both of these domains prior to the end of the training period (i.e., one hour) were presented with examples from a seamanship domain. At the completion of training, all participants had classified examples from at least the first two task domains.

Measures

To test our hypotheses, two types of measures were collected from the research participants, mental models of teamwork and self-generated teamwork examples.

Mental models of teamwork

Following the technique used in Study 1, mental models of teamwork were assessed via the card sorting task. Participants in this study sorted teamwork examples specific to a damage control task domain. This domain was not represented in the CBT. In addition to calculating participants’ similarity to other trainees and similarity to the expert model, the current study also examined the consistency with which participants represented their mental models of teamwork across two administrations of the card sorting task. Consistency of mental models was scored for those in the double pretest and double post-test conditions by correlating each participant’s card sort data/matrices collected at time 1 with that collected at time 2.

Self-generated examples

Participants were provided with definitions of each component behavior included in the expert model and asked to generate their own concrete examples of those components. Two research assistants that
were blind to experimental condition later provided independent assessments regarding the degree to which the teamwork examples generated by participants accurately reflected components of the expert model. Scores ranged from 0 (example is not at all representative of this component behavior) to 2 (example is very representative of this component behavior). Inter-rater reliability estimated by Pearson product moment correlation was 0.85. After making their independent assessments, the raters met to form a consensus. The consensus ratings were used in all subsequent analyses.

Total scores were calculated by summing the scores from each of the 11 component behaviors for each participant, with a possible range of 0 to 22. The averages and standard deviations of these total self-generated free response scores, by treatment condition, can be found in Table 2.

### Results

**Predictive validity of accuracy scores**

In support of Hypothesis 5, accuracy scores on the second card sort administration were significantly correlated with scores on the self-generated free response task, \( r = 0.390, p < 0.05 \) (\( N = 30 \)). As predicted, participants whose mental models were more similar to the expert model were better able to generate new examples that represented components within the expert model of teamwork.

**Similarity to the expert model**

The similarity between each participant’s mental model of teamwork and the expert model of teamwork was calculated for each of the two card sorts that he or she completed. As in Study 1, similarity to the expert model was taken to indicate mental model accuracy. Group accuracy means and standard deviations may be found in Table 2. A \( 3 \times 2 \) mixed model ANOVA was conducted, with three levels of the between-subjects variable or experimental condition (i.e., pretest/post-test treatment, double pretest, double post-test) and two levels of the repeated measure, card sort administration. Results indicated a significant main effect for both card sort administration, \( F(1,39) = 35.031, p < 0.01 \), and experimental condition, \( F(2,39) = 7.84, p < 0.01 \). We did not investigate these main effects in any more detail, because there was also a significant interaction between experimental condition and...
and card sort administration, $F(2,39) = 30.374, p < 0.01$. In support of Hypothesis 6, a follow-up test of simple effects found that accuracy scores improved significantly on the second administration of the card sort for the pretest/post-test treatment group, $F(1,39) = 79.927, p < 0.05$. In contrast, no significant difference existed in trainees’ similarity to the expert model of teamwork between the first and the second administration for either the double pretest group, $F(1,39) = 0.388, p > 0.05$, or for the double post-test group, $F(1,39) = 3.762, p > 0.05$. Thus, participants’ mental models did not become more similar to the expert model of teamwork simply by virtue of sorting the identical cards a second time.

**Similarity among trainees’ mental models of teamwork**

Following the procedures described in the previous study, similarity matrices were created for each card sort administration within each experimental condition, and sampled 100 times each to approximate the empirical distributions of overall group similarity scores. The one important distinction in this study is that the matrices within each group were not sampled independently. This modification was used to preserve the dependent nature of the data. If the similarity between participant no. 4 and participant no. 7 was randomly selected from the 1st matrix, then the similarity between those same two participants was also selected from the 2nd matrix.

For the pretest/post-test treatment group, across all 100 resamples, the average similarity scores among trainees on the first and second administrations of the card sort were 0.341 ($SD = 0.033$) and 0.525 ($SD = 0.026$), respectively. For the double pretest group, across all 100 resamples, the average similarity scores among trainees on the first and second administrations of the card sort were 0.281 ($SD = 0.042$) and 0.260 ($SD = 0.046$), respectively. For the double post-test group, across all 100 resamples, the average similarity scores among trainees on the first and second administrations of the card sort were 0.365 ($SD = 0.048$) and 0.391 ($SD = 0.048$), respectively.

Mixed model, $3 \times 2$ ANOVAs were calculated for each of the 100 resamples. As with the data related to trainees’ similarity to the expert model of teamwork, there were three levels of the between-subjects variable (experimental condition) and two levels of the repeated measure. The number of significant effects of each type (main effect of condition, main effect of measure administration and interaction) were tallied and a chi square test was used to determine if the number of significant effects was significantly different from what would be expected by chance. The main effect of the repeated measures variable, card sort administration, was significant in 28 (out of 100) resamples, which is significantly more than would be expected by chance, $\chi^2(1) = 111.4, p < 0.01$. The main effect of the treatment condition was significant in two (out of 100) resamples, which is not significantly more than what would be expected by chance, $\chi^2(1) = 1.9, p > 0.05$. The interaction was significant in 41 (out of 100) resamples, which is significantly more than would be expected by chance, $\chi^2(1) = 272.8, p < 0.01$.

We followed up on the interaction effect by performing simple effect tests on each of the 100 resamples, tallying the number of significant simple effects at each level of the treatment, and using a chi square test to determine if the number of simple effects was significantly different from what would be expected by chance. In support of Hypothesis 7, post-training similarity scores were significantly greater than pretraining similarity scores among trainees in the pretest posttest treatment group on 78 out of 100 resamples, $\chi^2(1) = 1121.9, p < 0.01$. There were no cases of a significant effect of card sort administration for the double pretest group, and only one case of a significant effect of card sort administration for the double post-test group. No statistical test was necessary to determine that these findings were not significantly higher than what would be expected by chance. Thus, participants’ similarity to other trainees did not appear to improve by virtue of repeated measurement alone.
Consistency
The consistency with which a given trainee represented his/her teamwork mental model was computed by correlating the matrices produced by his/her first and second card sort administration. This was done only for those within the double pretest and double post-test groups. These participants completed the same card sort on two successive days without any intervention in between the two test administrations. Consistency of mental models for the double pretest (N = 11) and double post-test groups (N = 10) were 0.543 and 0.619, respectively.

We had hypothesized that participants would demonstrate greater consistency after training than before training. In order to test this hypothesis, we applied a Fisher’s r-to-z transformation and statistically compared the two groups’ average scores. In support of Hypothesis 8, participants in the double post-test group demonstrated significantly greater consistency in their mental models than did participants in the double pretest group, z = 1.863, p < 0.05 (one-tailed).

Discussion: Study 2

Results from Study 2 indicated that, after one-hour of training with the CBT tool, trainees were more consistent in representing their teamwork mental models as assessed by the card sort measure. Moreover, trainees’ mental models became more similar to one another and to the expert model. It is important to note that trainees did not interact with one another during training. Thus, greater similarity among trainees was produced by guiding individuals toward adopting a pre-specified expert model of teamwork. These results suggest that, in situations where an accurate model of teamwork can be prespecified and used as a framework for training, a CBT such as the one described here could be used to prepare individuals to perform with future teammates prior to joining a new team. Certain groups of individuals, such as low ranking teammates in the present example, may be specifically targeted for such preparatory training. Moreover, a methodology such as the card sorting technique employed in the present research may be used to preassess teammates’ need for this type of training on a case-by-case basis.

Furthermore, on the basis of Study 2 findings, the CBT tool is currently being used to prepare individual members of existing teams to participate in a post-exercise debriefing strategy involving team self-correction and goal setting within the framework of the expert model (Smith-Jentsch et al., manuscript in preparation). This use of the CBT is supported by the fact that accuracy scores predicted individuals’ ability to generate novel examples of teamwork that illustrated components of the expert model. The self-generated example task had been designed to mirror the cognitive processes involved in using the expert model as a framework from which to critique a team’s performance.

Summary

The findings reported here add to a growing body of research on shared team knowledge and knowledge structures. A unique aspect of the present research involved the nature in which the accuracy of participants’ teamwork mental models was assessed. Typically, researchers who have sought to assess mental model accuracy, or quality, have adopted the mental model(s) of those with high domain-specific experience as the criteria against which to evaluate participants’ mental models. This
would assume, for example, that those with greater domain-specific team experience hold cognitive representations of the components of teamwork and the relationships among those components that more accurately reflect the nature in which effective teamwork operates for a given team type. This assumption, however, has not been empirically tested previously. Furthermore, it is unclear which indicators of domain-specific experience may be most closely associated with mental model accuracy.

In the present research, we tested the hypotheses that those with higher rank (H1) and greater time in navy service (H2) would hold mental models that more accurately reflected how teamwork behaviors co-vary in a navy command and control environment. The expert model used as a criterion against which accuracy was assessed had been defined in a previous study through the analysis of behavioral ratings assigned to navy command and control teams. Thus, this model was taken to reflect the actual manner in which teamwork behaviors were related. Rank, but not time in service, was positively associated with similarity to this expert model.

A second key finding was that greater similarity existed among the teamwork mental models held by high-ranking individuals relative to low ranking individuals, and by navy personnel with high time in service relative to those with low time in service. Based on these findings, a strategy that incorporated individualized feedback in order to guide trainees with a variety of less accurate teamwork mental models toward a common understanding of the expert model was developed. As described in Study 2, the resulting CBT tool increased mental model accuracy (H6) and similarity (H7) among trainees, as well as the consistency with which individuals represented their teamwork knowledge (H8).

**Implications for training needs assessment, development, and evaluation**

In training situations like the one described here where a single or relatively homogeneous set of accurate mental models can be defined, training is likely to increase mental model accuracy and similarity among teammates in tandem. However, researchers have begun to examine situations where multiple and distinct accurate, or high quality, models and multiple and distinct inaccurate, or low quality, models can be defined (Mathieu et al., Manuscript submitted). Below we consider the following four possibilities and their implications for training. We hope these ideas stimulate theoretical discussion and empirical research in the future.

**Heterogeneous accurate and inaccurate models**

In an environment where both multiple heterogeneous accurate and inaccurate models exist, equally low mental model similarity may be found within experienced and inexperienced groups of individuals prior to training. In assessing the training needs for such teams one would have to determine whether there is a benefit to having teammates hold both accurate and similar mental models of teamwork or whether those holding dissimilar, but accurate mental models, can coordinate effectively. In the latter case, it may be more effective to focus on training individuals to more consistently apply their teamwork knowledge, and on ensuring that teammates are aware of one another’s teamwork mental models. In support of this notion, at least one previous study found that it was not the degree of similarity among teammates’ ‘teamwork schemas’ that was associated with performance indicators but their ability to accurately predict one another’s teamwork schemas (e.g., Jenkins and Rentsch, 1995 – Paper presented at the 10th SIOP Conference, Orlando, FL).

**Heterogeneous accurate models, homogeneous inaccurate models**

In other team task environments, there may be multiple and diverse ways of thinking about teamwork that are each accurate, but a relatively small number of common misconceptions or inaccurate mental models. In this case, prior to training, greater similarity may be found among inexperienced
individuals relative to groups of experienced individuals. As in the example described above, similarity among teammates may or may not be required for effective performance so long as teammates hold high quality, or accurate, mental models. Thus, trainees may hold more accurate mental models after training, however similarity among trainees may actually decrease.

**Homogeneous accurate models, homogeneous inaccurate models**

For other team types, a single or relatively small number of homogeneous accurate mental models of teamwork and a single or relatively small number of homogeneous inaccurate teamwork mental models may exist. In this case, groups of inexperienced individuals may demonstrate equally high mental model similarity to one another as compared to groups of experienced individuals prior to training, although the accuracy of these mental models should differ. The progression from novice to expert in such an environment may follow a relatively predictable path from a common ineffective way of viewing teamwork to a common effective one. This means that training strategies designed to accelerate expertise could explicitly address predictable misconceptions and guide trainees toward the development of a single model. Inexperienced teammates could be trained together since they are likely to progress along the same developmental path. While these individuals would be expected to show greater accuracy after training they would not necessarily be expected to show greater similarity to one another than they did before training. They would simply go from sharing similar inaccurate mental models to sharing similar accurate ones.

**Homogeneous accurate models, heterogeneous inaccurate models**

Finally, in some team environments, experienced individuals may hold a single or small number of homogenous accurate models, however inexperienced teammates may hold a large number of dissimilar inaccurate models. In this case, high experience should be positively associated with greater accuracy, and groups of high experience individuals should show greater similarity to one another than groups of inexperienced individuals. This was the pattern of results found in Study 1 with respect to rank. In this type of environment, self-paced strategies incorporating individualized feedback designed to guide individuals, whose pre-training mental models may vary widely, in adopting a pre-specified expert model may be most effective at fostering both mental model accuracy and similarity among trainees.

**Study limitations and directions for future research**

We would like to end by noting a number of limitations and directions for future research. First, it may be that a heterogeneous set of accurate mental models is more likely to exist for certain aspects of team performance than for others. For example, it may be common that multiple and diverse task strategies are effective for achieving superior team performance in a given environment. Therefore, teammates may hold highly dissimilar but equally accurate or ‘high quality’ task mental models. On the other hand, the nature of effective teamwork may be relatively straightforward and thus a single or homogeneous set of accurate teamwork mental models may be more commonly found. The present research focused exclusively on teamwork mental models. Moreover, accuracy was assessed against a single empirically-derived expert model. We cannot rule out the possibility that additional accurate models existed. However, the fact that rank was positively correlated with similarity to the expert model suggests that if multiple accurate models existed they were unlikely to be highly dissimilar. Future research is needed to investigate the degree to which homogeneous and heterogeneous accurate models exist for thinking about one’s task, equipment, and teammates, in addition to teamwork principles.

Second, future research should examine the relative importance of various team experience indicators in a variety of team environments. The present study measured length of time in the navy and navy
rank. While these variables were related to mental model similarity, they did not tap specific team experience directly as one may serve in a variety of team types throughout a navy career. Finally, much remains to be learned about the instructional features that are most effective at increasing the similarity and accuracy of teammates’ mental models. In an environment where a single expert model cannot be specified a priori, it is expected that guided interaction among teammates will be critical to achieving mental model similarity (e.g., Blickensderfer et al., 1997 – Paper presented at the 12th Conference, St Louis, MO; Smith-Jentsch et al., 2000 – manuscript in preparation).

Author biographies

Gwendolyn E. Campbell is a Senior Research Psychologist in the Science and Technology Division of the Naval Air Warfare Center Training Systems Division (NAWCTSD), Orlando, Florida. Dr Campbell holds a BA in Mathematics from Youngstown State University, Youngstown, Ohio, and MA and PhD degrees in Experimental Psychology from the University of South Florida, Tampa, Florida. Her research interests include cognitive theories of learning and instruction, cognitive modelling, and the measurement of structured knowledge.

Dana Milanovich is a Research Psychologist at the Naval Air Warfare Center Training Systems Division. She received her MS degree in Industrial/Organizational Psychology from the University of Central Florida in 1998.

Angelique M. Reynolds is a research Psychologist at the Naval Air Warfare Center Training Systems Division. She received a BA in Psychology from San Diego State University in 1983 and an MA in General/Experimental Psychology from San Diego State University in 1987.

Kimberly Smith-Jentsch is a Senior Research Psychologist at the Naval Air Warfare Center Training Systems Division. Dr Smith-Jentsch received a doctorate degree in Industrial/Organizational Psychology from the University of South Florida in 1994. Her previous research has focused on issues related to human performance measurement, training needs assessment, development, and evaluation, teamwork processes, team and organizational climate.

Acknowledgements

We would like to thank Kristin Kostowicz, Alicia Smeltzer, and Melissa Straughan for their assistance in coding the data reported in this manuscript, and Florian Jentsch for his advice on data analyses and comments on previous drafts of this manuscript.

References


