

**Investigating The Diagnosticity Of A Method For Measuring
Teamwork Mental Models**

Kimberly A. Smith-Jentsch
Patrick Rosopa
Alicia D. Sanchez

University of Central Florida

Lizzette Lima

University of South Florida

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 2003		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Investigating The Diagnosticity Of A Method For Measuring Teamwork Mental Models				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Central Florida				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 16	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

A number of studies have documented relationships between mental model similarity among teammates and performance (e.g., Mathieu, Heffner, Goodwin, Salas & Cannon-Bowers, 2000; Rentsch, Heffner & Duffy, 1994). On the basis of these findings, it has been argued that mental model measurement could be used to diagnose and remedy knowledge deficiencies in applied training settings. In order to be useful for this purpose, mental model measures need to do more than just predict performance. Trainers and trainees need mental model measures that diagnose specific underlying knowledge deficiencies that lead a particular individual to be dissimilar from their teammates or from some predefined expert model. However, few previous studies have demonstrated the diagnosticity of mental model measures.

The purpose of the present study was to investigate the diagnosticity of metrics designed to parse the contribution of conciseness, consistency, and grouping strategy in predicting mental model similarity. The ultimate goal of the effort was to support a training-related application of the measure whereby the measure would be used to tailor feedback and other instructional strategies that address the specific knowledge deficiencies of individual team members. Specifically, we were interested in diagnosing the root causes of dissimilarity between an individual's teamwork mental model and (1) an expert model, as well as (2) dissimilarity within participant dyads.

The particular measure used here to assess teamwork mental models has been used in a number of previous studies. Using this measure, results indicated that higher-ranking teammates held teamwork mental models that were

more similar to an expert model. Furthermore, those whose mental models were more similar to the expert model were better able to generate concrete examples that were consistent with the expert model.

In terms of similarity among individual participants, a group of individuals with high team experience held more similar mental models than a group of low team experience. Finally, teammate similarity scores were significantly related to team performance. Thus, previous findings have indicated that both similarity to the expert model and similarity among participants on this measure are related in expected ways to indicators of experience and performance. However, to be useful for purposes of developmental feedback and/or remedial training, this measure (like others) must produce metrics that help to diagnose the root causes of dissimilarity. The following sections describe the theoretical foundation for hypotheses regarding which additional metrics should be related to dyad similarity and similarity to the expert model.

Similarity to an expert model

Previous research on expert-novice differences has suggested that expert knowledge structures differ from novice structures in at least three ways. First, experts tend to represent their knowledge more concisely (Rentsch et al., 1994). Second, experts tend to be more consistent in representing what they know (e.g., Borko & Livingston, 1989). Third, experts tend to organize their knowledge in terms of more abstract underlying processes in contrast to concrete superficial features (e.g., Borko & Livingston, 1989, Glaser & Chi, 1988). Accordingly, we hypothesized that metrics of conciseness, consistency, and grouping strategy

(abstract or concrete) would contribute uniquely to the prediction of mental model similarity to an expert model.

Dyad similarity

Previous research has demonstrated that those with less team experience are both less consistent in representing their knowledge (Rentsch et al) and less similar to one another than those with greater team experience (Smith-Jentsch, Campbell, Milanovich, & Reynolds, 2001). From a basic psychometric perspective, low reliability of a measure limits the size of the correlation between that measure and any other variable. Thus, one reason why less experienced individuals tend to show less similarity with one another may simply be the fact that measures of their mental models are less reliable or consistent. Accordingly, we hypothesized that individuals who are less consistent in representing their teamwork mental models would also demonstrate lower similarity on average with other participants. Finally, given that those with high team experience tend to focus on abstract processes whereas more novice individuals focus on concrete features of a task, it was hypothesized that differences in grouping strategy (abstract or concrete) would also explain variance in mental model similarity among dyads.

Method

Participants

Participants were 107 students enrolled in introductory psychology classes at a large southeastern university. Participants were given extra credit points for participating in the study. The majority of the participants were female ($N = 71$).

The average age of the participants was 19.55 years ($SD = 3.12$). 71% of the subjects were freshmen, 3.7% were sophomores, 3.7% were juniors, 14% were seniors, and 7.5% were classified as other.

Procedure

First, participants completed a demographic survey. Next, participants completed a card-sorting task designed to measure their mental models of teamwork twice with a 10-minute distracter video between. Finally, participants received an experimental debrief.

Two versions of the card sorting task were used for data collection; a manual and computerized version. As reported earlier, these two versions of the task did not result in differences across participants on the metrics of interest in this study (e.g., similarity to the expert, consistency) (Harper, Jentsch, Van Duyne, Smith-Jentsch, & Sanchez, 2002). Thus, hypotheses were tested using the combined participant sample.

Measures

Demographic Data. Participants completed a form that asked them to report their gender, age, and class standing (i.e., freshman, sophomore, junior, senior, other).

Mental Models of Teamwork. The card-sorting method used to assess mental models of teamwork in the present study has been utilized previously in studies involving tasks from a variety of team environments (e.g., aviation, damage control/fire fighting) (Smith-Jentsch, Campbell, Milanovich, & Reynolds, 2001). This method requires participants to sort concrete examples of teamwork from a particular team environment into piles that represent categories of

teamwork that are meaningful to them. In the manual version, these examples are printed on index cards. In the computerized version, participants use a drag-and-drop feature to sort the examples electronically. Given the population used in the present study (college students), examples of teamwork in a restaurant setting were used. For instance, one example read “A cook noticed that an order was being sent out with French fries rather than mashed potatoes and pointed that out to the other cook.” Participants then label their piles in a way that explains why they grouped certain examples together (e.g., leadership, communication issues). The resulting card sort data are then scored in terms of the following metrics that are described below: Similarity to the Expert Model, Dyad Similarity, Mental Model Consistency, Conciseness, Grouping Strategy, and Similarity of Grouping Strategy.

Similarity to the Expert Model. The expert model employed in this research consisted of eleven teamwork components clustered within four behavioral dimensions. This model has been shown previously to discriminate experienced and inexperienced teams (Smith-Jentsch, Johnston, & Payne, 1998) and to predict performance outcomes (SIOP paper). The four behavioral dimensions are: Information Exchange (i.e., passing information, providing big picture summaries, seeking information from all available sources), Communication Delivery (i.e., proper phraseology, brevity, clarity, completeness of standard reports), Supporting Behavior (i.e., error correction, back-up/assistance), and Leadership/Initiative (i.e., providing guidance, stating priorities).

In order to assess similarity to this expert model using our card sorting method, three researchers familiar with the expert model independently sorted the restaurant examples into piles that would be consistent with the expert model. These three researchers were in perfect agreement (100%) in sorting the cards. A BASIC computer program was then used to create a string of zeros and ones that account for all possible combinations of the 33 examples, whether or not they were placed in the same pile by the group of three researchers. The same procedure was conducted to derive data strings for each participant card sort. Similarity to the expert model was then computed by correlating each participant card sort data string with the expert data string using the *Phi* coefficient (which represents the Pearson correlation coefficient between two dichotomous variables). This same procedure was used to derive scores of similarity to the expert at Time 1 and Time 2.

Dyad Similarity. Dyad similarity was computed by correlating each participant's card sort data string with every other participant's data string in the same manner that it was correlated with the expert data string. This resulted in 5671 dyad similarity scores at time one and time two.

Mental Model Consistency. Mental model consistency was measured by correlating each participant's Time 1 and Time 2 card sort data strings. Thus, unlike the remaining scores, mental model consistency could only be computed once. This score indicated how similarly the same participant sorted the identical teamwork examples when asked to do so twice. Participants were unaware that they would be asked to repeat the card sort until after the distracter video to

avoid their attempting to memorize the contents of their piles. Moreover, participants were told at the start of the second card sort that they were not being asked to replicate the first sort but rather to look at the examples again and to sort them in whatever manner they felt best represented their view of teamwork.

Conciseness. Conciseness was scored by computing the average number of words per label used by each participant to describe their piles. This was done for both Time 1 and Time 2 cardsorts.

Grouping Strategy. Two independent raters were trained to code the grouping strategy participants appeared to use in order to separate examples within their card sort by examining the labels they used to describe their piles. Specifically, we were interested in whether participants seemed to be grouping examples on the basis of abstract underlying processes (e.g., leadership, communication) or concrete features (e.g., managers, servers, cooks). Raters were trained to consider the participants' labels together and to determine what was being used to discriminate among piles. This was done rather than scoring each label individually and averaging those scores because it was deemed a better indicator of grouping strategy.

For example, if the labels were to be scored individually and averaged, a participant who used the labels "coordination among hostesses" "coordination among waitresses" and "coordination among cooks" would likely receive a score that indicated they grouped based on abstract processes. When considered as a group however, it appears clear that while each label contains the word

coordination, which is an abstract process, such a participant was clearly grouping examples based on a concrete features (e.g., team position).

Thus, two raters assigned a single grouping strategy code (abstract or concrete) by examining each participant's labels as a group at Time 1 and again at Time 2. Inter-rater agreement estimated using Cohen's Kappa, which adjusts for agreement expected due to chance, was 76%. Next, the two raters came to consensus on any scores for which they disagreed. Thus, grouping strategy scores used in all subsequent analyses represented judgments that were ultimately agreed upon by both raters.

Similarity of Grouping Strategy. A score for similarity of grouping strategy was assigned to each participant dyad (N=5671) at Time 1 and Time 2. If both participants in a dyad adopted the same grouping strategy, regardless of which strategy that was, they were assigned a '1'. If one participant in a dyad grouped their examples based on abstract features and the other on concrete features they were assigned a '0'.

Results

Table 1 lists the means, standard deviations and correlations for all study variables at the individual level of analysis (i.e., similarity to expert, grouping strategy, consistency, conciseness). Table 2 lists the means, standard deviations and correlations for all variables computed at the dyad level of analysis (i.e., dyad similarity, similarity of grouping strategy). As shown in this table, scores for the various metrics were moderately correlated across the two administrations (.34 - .58). Dependent T-tests computed on each measure indicated that no significant change was seen from Time 1 to Time 2.

Two sets of analyses were computed to test hypotheses. First, multiple regression analyses were computed to examine potential predictors of similarity to the expert model at Time 1 and Time 2. Second, multiple regression analyses were computed to examine potential predictors of dyad similarity at Time 1 and Time 2.

Predictors of Similarity to the Expert Model. Multiple regression analyses were computed separately for Time 1 and Time 2, with similarity to the expert mental model as the criterion variable and mental model consistency, conciseness, and grouping strategy as the predictor variables. Using the enter method, a significant overall model emerged for similarity to the expert at Time 1 ($F_{3, 103} = 8.287, p < .001, \text{Adjusted } R^2 = .171$) and at Time 2 ($F_{3, 103} = 5.334, p < .01, \text{Adjusted } R^2 = .109$).

In partial support of Hypothesis 1, participants who used more concise labels to describe their teamwork categories at Time 1, $\beta = -.179, p < .05$ (one-tailed), held mental models that were more similar to the expert. However, this relationship was not found using Time 2 data. In full support of Hypothesis 2, participants who were more consistent in sorting their cards held mental models that were more similar to the expert. This was true both at Time 1 $\beta = .292, p < .01$, and at Time 2, $\beta = .247, p = .01$. In full support of Hypothesis 3, participants who adopted a grouping strategy based on abstract underlying processes held mental models that were more similar to the expert at both Time 1, $\beta = .400, p < .001$ and Time 2, $\beta = .324, p = .001$.

Predictors of dyad similarity. Multiple regression analyses were computed separately for Time 1 and Time 2, with dyad similarity of mental models as the criterion variable. Consistency scores for each participant in a dyad and a single score representing dyad similarity of grouping strategy were the predictor variables. Using the enter method, a significant overall model emerged for similarity to the expert at Time 1, ($F_{3, 5667} = 335.316$, $p < .001$, Adjusted $R^2 = .150$) and at Time 2 ($F_{3, 5667} = 315.327$, $p < .001$, Adjusted $R^2 = .143$).

Hypotheses 4 stated that the consistency with which each participant in a dyad sorted their cards across the two administrations would contribute unique variance in predicting dyad similarity scores. In full support of this hypothesis, consistency scores for both dyad members were significant predictors of mental model similarity within the dyad both at Time 1 (member 1, member 2) and Time 2 (member 1, member 2). Finally, hypothesis 5 stated that dyads in which both members adopted the same grouping strategy would hold more similar mental models. This hypothesis was fully supported both at Time 1 and Time 2.

Conclusions

Results from this study are consistent with previous research on expert-novice differences in knowledge structures. Specifically, those who received greater similarity to expert scores tended to be more concise in describing their teamwork categories (time 1 only), to be more consistent in grouping teamwork examples across time 1 and time 2, and to adopt a grouping strategy based on abstract processes (time1 and 2).

Additionally, findings related to dyad similarity were also in the expected direction. Consistent with psychometric theory, the consistency (or test-retest reliability) of each person in a dyad contributed unique variance to the prediction of dyad mental model similarity. In other words, dyad similarity was limited in part by the consistency with which either person could express what they know about teamwork using the card-sorting task. This finding helps to explain one reason why low experience individuals are not only less similar to an expert model but also in disagreement with one another. It also suggests that differences in test-retest reliability across measures used to assess mental model similarity could explain mixed findings across studies. Finally, dyads where both participants adopted the same grouping strategy, regardless of which strategy that was, tended to hold more similar mental models. However, the effect size for this finding was very small.

Together, results from this study suggest that feedback provided to trainees based on their grouping strategy, conciseness, and consistency may be useful for increasing mental model similarity and similarity to an expert model with the ultimate goal of improving team performance. Future research is underway to investigate various feedback strategies for doing so.

References

Borko, H. & Livingston, C. (1989). Cognition and improvisation: Difference in mathematics instruction by expert and novice teachers. *American Educational Research Journal*, 26, 473-498.

Glaser, R., & Chi, M. T. H. (1988). *Overview*. In M.T.H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. Xv-xxviii). Hillsdale, NJ: Erlbaum.

Rentsch, J.R., Heffner, T.S., & Duffy, L.T. (1994b). What you know is what you get from experience: Team experience related to teamwork schemas. *Group and Organization Management*, 19(4), 450-474.

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, 179-194.

Table 1

Inter-correlations among variables at the individual level of analysis (N = 107)

Variable	1	2	3	4	5	6	7
1. Similarity to Expert Model at Time 1	-	.569**	.160	-.076	-.039	.336**	.175
2. Similarity to Expert Model at Time 2		-	.183	-.060	-.007	.196*	.274**
3. Consistency			-	.302**	.129	-.195*	-.170
4. Conciseness at Time 1				-	.510**	.038	.061
5. Conciseness at Time 2					-	.015	.107
6. Grouping Strategy at Time 1						-	.581**
7. Grouping Strategy at Time 2							-

*Note. *p < .05. **p < .01.*

Table 2

Inter-correlations among variables at the dyadic level of analysis (N = 5671)

Variable	1	2	3	4	5	6
1. Consistency of Member 1	-	-.007	.300**	.215**	.053**	.031*
2. Consistency of Member 2		-	.226**	.225**	.000	.017
3. Dyad MM Similarity at Time 1			-	.477**	-.079**	-.027*
4. Dyad MM Similarity at Time 2				-	-.016	-.031*
5. Grouping Strategy Similarity at Time 1					-	.340**
6. Grouping Strategy Similarity at Time 2						-

*Note. *p < .05. **p < .01.*

Table 3

Summary of Simultaneous Regression Analysis for Variables Predicting Similarity to the Expert Model at Time 1 (N = 107)

Variable	<i>B</i>	<i>SE B</i>	β
Consistency	.112	.037	.292**
Conciseness	-.008	.004	-.179*
Grouping Strategy	.082	.019	.400**

*Note. Adjusted R² = .171. *p < .05. **p < .01.*

Table 4

Summary of Simultaneous Regression Analysis for Variables Predicting Similarity to the Expert Model at Time 2 (N = 107)

Variable	<i>B</i>	<i>SE B</i>	β
Consistency	.096	.036	.247**
Conciseness	-.004	.005	-.073
Grouping Strategy	.067	.019	.324**

*Note. Adjusted R² = .109. *p < .05. **p < .01.*

Table 5

Summary of Simultaneous Regression Analysis for Variables Predicting Dyad Similarity at Time 1 (N = 5671)

Variable	<i>B</i>	<i>SE B</i>	β
Consistency of Member 1	.190	.008	.306**
Consistency of Member 2	.153	.008	.228**
Similarity of Grouping Strategy at Time 1	-.032	.004	-.095**

*Note. Adjusted R² = .150. **p < .01.*

Table 6

Summary of Simultaneous Regression Analysis for Variables Predicting Dyad Similarity at Time 2 (N = 5671)

Variable	<i>B</i>	<i>SE B</i>	β
Consistency of Member 1	.187	.008	.302**
Consistency of Member 2	.153	.008	.229**
Similarity of Grouping Strategy at Time 2	.012	.004	.036**

*Note. Adjusted R² = .099. **p < .01.*