Examining the convergent validity of shared mental model measures

Sarah Ross · Natalie Allen

Published online: 5 April 2012
© Psychonomic Society, Inc. 2012

Abstract Despite widespread interest in shared mental models (SMMs) within teams and groups and an extensive empirical literature examining SMM–performance links, very little is known about the convergent validity of commonly used measures of SMMs. In this study, two-person teams ($n = 96$) engaged in a complex flight task and completed three SMM measures: concept mapping, paired ratings, and causal mapping. Task-based sharedness scores were compared across the measures. Analyses were conducted in two ways: using SMMs of actual team members ($n = 96$ pairs) and using the SMMs of pairs of participants who worked separately but for whom similarity indices were calculated after the study ($n = 18,240$). The purpose of the latter pairs, coined *pseudo-partners*, was to create a sample with considerable power to test the convergent validity of the SMM measures. The results call into some question the convergent validity of these task-based SMM measures.

Keywords Shared mental models · Teams · Groups

Evaluating shared mental model measures

Over the past several years, considerable attention has been given to the possibility that *shared mental models* (SMMs) enhance team performance (e.g., Marks, Sabella, Burke, & Zaccaro, 2002; Marks, Zaccaro, & Mathieu, 2000; Mathieu, Heffner, Goodwin, Cannon-Bowers, & Salas, 2005; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000; Rouse,

Cannon-Bowers, & Salas, 1992). The term *mental model* refers to organized knowledge structures, or sets of concepts and the associations among them (Langan-Fox, Code, & Langfield-Smith, 2000; Smith-Jentsch, Campbell, Milanovich, & Reynolds, 2001). Thus, mental models are defined in terms of both *content knowledge* and, importantly, the *structure* (or *organization*) of that content knowledge. Mental models “help people to describe, explain, and predict events in their environment” (Mathieu et al., 2000, p. 274) and are described as being “shared” among individuals, to the degree that there is overlap of these knowledge organizations (Orasanu & Salas, 1993).

Theoretically, of course, one can have a mental model about any type of content. Researchers have examined mental models about content as diverse as group counseling interventions (Kivlighan & Kivlighan, 2009), statistics (Lavigne, Salkind, & Yan, 2008), and teamwork (Lim & Klein, 2006). Within the workgroup and team literatures, most empirical attention has focused on teamwork mental models and task-related mental models. Team members who have highly shared mental models about *teamwork* have similar ideas about the characteristics of the people they are working with and the nature, purpose, and patterns of their interactions. Team members who have highly shared mental models about *tasks* have similar ideas about procedures/strategies, technology/equipment, and the constraints and opportunities inherent in the team’s task (Cannon-Bowers, Salas, & Converse, 1993). Essentially, the members of high SMM teams (whether task, teamwork, or other knowledge content) are “on the same page,” and being on the same page allows members to anticipate what others in the team are going to do (Orasanu & Salas, 1993). When team members know what their fellow teammates are going to do, it is argued, they should be able to coordinate their actions well. When effective coordination exists, teams should perform better at their tasks (Marks et al., 2002; Mathieu et al., 2000; Rouse et al., 1992).

S. Ross (✉) · N. Allen
Department of Psychology, The University of Western Ontario,
London, Ontario, Canada
e-mail: sarah_ross@rogers.com

Shared mental models have received a great deal of attention during the past several years. Evidence of this is seen in numerous journal articles (e.g., Cooke, Salas, Cannon-Bowers, & Stout, 2000; Edwards, Day, Arthur, & Bell, 2006; Klimoski & Mohammed, 1994; Marks et al., 2000; Mathieu et al., 2000; Orasanu & Salas, 1993), meta-analyses (DeChurch & Mesmer-Magnus, 2010a, b), special issues in journals (*CoDesign*; Badke-Schaub, Lauche, & Neumann, 2007, and *Journal of Organizational Behavior*; Salas & Cannon-Bowers, 2001), and textbook chapters (e.g., Muchinsky, 2009; Spector, 2006) dedicated to the topic of SMMs. In addition, SMMs have been examined in numerous contexts, such as the military (Lim & Klein, 2006), medicine (Gillespie, Chaboyer, Longbottom, & Wallis, 2010), software development (Levesque, Wilson, & Wholey, 2001), and nuclear power (Waller, Gupta, & Giambatista, 2004). Finally, the SMM construct has made its way into the practitioner literature and is beginning to lay the groundwork for advice regarding various assessments and interventions aimed at improving team performance (e.g., Haig, Sutton, & Whittington, 2006; Johnson, Sikorski, Mendenhall, Khalil, & Lee, 2010).

Despite this growing body of research on SMMs and excellent overviews by Langan-Fox, Mohammed, and their colleagues (Langan-Fox et al., 2000; Mohammed, Ferzandi, & Hamilton, 2010; Mohammed, Klimoski, & Rentsch, 2000), very little empirical research has focused specifically on the *evaluation* of SMM measurement. In particular, only a few attempts have been made to empirically examine various SMM measures in relation to one another (Banks & Millward, 2007; Dorsey, Campbell, Foster, & Miles, 1999; Resick et al., 2010). This is somewhat disconcerting, most notably because it appears that different SMM measures are used somewhat interchangeably across studies. The goal of the present research, therefore, is to help fill an important gap by examining the intercorrelations among and, hence, the convergent validity of three SMM measurement techniques that receive much of the attention in the empirical literature and reviews of SMM measurement: concept mapping, paired ratings, and causal mapping measures. We begin by summarizing these approaches to SMM measurement within teams.

The measurement of shared mental models within teams

SMMs are measured in various ways, reflecting, perhaps, the complexity of the construct. Despite the complexity of the construct and its measures, the measurement process generally involves two stages: concept generation and sharedness calculation.

Concept generation In the first stage, researchers need to extract from participants the concepts, and relationships among these concepts, that are important to the knowledge domain in question. This can be done in various ways. For

example, the researcher might provide participants with a list of concepts and ask them to rate the similarity of each possible pair of concepts (e.g., Mathieu et al., 2005; Stout, Cannon-Bowers, Salas, & Milanovich, 1999). Alternatively, the researcher might ask participants to respond to open-ended questions about the knowledge domain and then code the responses to determine participants' perceived relationships among concepts (e.g., Carley, 1997).

Sharedness calculation Once participants have provided information that can be used to describe their own mental models, the researcher can use this information to determine the sharedness or similarity among group members' mental models; herein, we use the term *sharedness calculation* to refer to this stage. Various analytic techniques can be used to calculate similarity (or sharedness). For example, Mathieu and his colleagues (Mathieu et al., 2005; Mathieu et al., 2000) used the quadratic assignment procedure correlation to assess sharedness of paired ratings, and Langfield-Smith and Wirth (1992) used the distance ratio formula to assess the sharedness of causal maps.

Of the numerous ways that SMMs can be measured, we chose to examine three: paired ratings, concept mapping, and causal mapping. As was noted earlier, the selection of these techniques was based on their use in the empirical literature and the attention paid to each in reviews of SMM measurement. The terms *paired ratings*, *concept mapping*, and *causal mapping* make reference to the concept generation stage of the SMM measurement process, during which participants rate the relationships among concepts. Note, however, that each of these terms also has associated with it a separate technique for determining similarity (the sharedness calculation stage). Below, these three measures of SMMs are explained in more detail.

Paired ratings measure One of the most common methods of extracting mental model information is paired ratings, a procedure in which participants must rate the relationships between each and every pair of a set of generated concepts considered integral to the task (Edwards et al., 2006; Langan-Fox et al., 2000; Mathieu et al., 2005; Mathieu et al., 2000; Stout et al., 1999). Often, this is done by presenting participants with a list of concepts along the top and side of a matrix and asking them to provide a numerical assessment of the "relatedness," or similarity (e.g., 1 = *very unrelated*; 7 = *very related*), of the concepts in each pair in the matrix. Once one participant in a team has provided these similarity ratings for each pair of concepts, the researcher can compare them with the other team member's ratings.

Some researchers make similarity comparisons among paired ratings matrices using UCINET 6, a program that allows a researcher to compare the similarity of two matrices using a quadratic analysis procedure (QAP; Krackhardt,

1987, 1988). The QAP is useful for determining significance for data sets whose observations are interdependent. It is important to note, however, that the value obtained for a Pearson correlation after the QAP has been used is the same as the value obtained for a Pearson correlation. Still, researchers report using the “QAP correlation,” rather than a simple Pearson correlation, as a method for testing convergence of paired ratings data (e.g., Heffner, Mathieu, & Cannon-Bowers, 1998; Mathieu et al., 2000). Other researchers determine sharedness using a similarity index in a program called Pathfinder, which creates a network structure from a set of paired ratings (Schvaneveldt, 1990). Lim and Klein (2006) compared similarity scores generated with QAP correlations and Pathfinder similarity indices to show that the QAP correlation and Pathfinder similarity index are very similar to one another. Thus, due to its ease of use and in keeping with the seminal work by Mathieu and his colleagues on SMMs (Mathieu et al., 2000), we chose to use a Pearson correlation to measure similarity of paired ratings.

Concept mapping measure Concept mapping is a simple technique in which participants organize each of a set of concepts into a hierarchical order or a “map” of their knowledge. Participants order the set of concepts in a sequence, representing their organization of the concepts. Similarity between the group members’ maps is determined by the number of links held in common between the group members (Marks et al., 2000; Mohammed et al., 2000). Concept mapping has been mentioned as a possible measurement technique in reviews by Mohammed and colleagues (Mohammed et al., 2010; Mohammed et al., 2000) and has been used in empirical research conducted by Marks et al. (2000), Minionis (1995), and Ellis (2006).

Causal mapping measure In causal mapping, as in the paired ratings measure, pairs of concepts are rated by participants. However, whereas the paired ratings technique asks for similarity ratings from *unrelated* to *highly related*, the causal mapping technique asks participants to provide information about the direction of these relationships as well. For example, a causal mapping measure would ask participants whether pairs of concepts influence one another, whether they do so positively or negatively, and whether they are weakly, moderately, or strongly related (Langan-Fox et al., 2000; Langfield-Smith & Wirth, 1992).

Convergence among members’ causal mapping ratings may be determined via an algorithm called the *distance ratio formula*. Essentially, the distance ratio formula calculates the summed absolute value of the difference between each value in one person’s matrix and the corresponding value in his or her partner’s matrix and divides this number by the maximum possible difference between the two matrices (Langfield-Smith & Wirth, 1992). The maximum possible

difference between two matrices depends on the number of concepts in a matrix and the number of potential values any cell in the matrix could take. When the matrices are very similar, the distance ratio formula will be close to 0; when the matrices are very different, the distance ratio formula will be close to 1. The causal mapping technique is highlighted in reviews of SMMs and has been used in empirical research as well (Ambrosini & Bowman, 2005; Langfield-Smith & Wirth, 1992; Tegarden, Tegarden, & Sheetz, 2009).

Summary of shared mental model measures

The measures described above and used in this study were selected for several reasons. First, each measure makes use of a standard, researcher-determined set of concepts in which participants rate the relationships between the concepts. Other mental model measures (e.g., content analysis of interviews) rely on unique, participant-determined sets of concepts, which make comparisons among measures very difficult. Second, each of these three measures is easily quantifiable. Whereas measures such as content analysis of interviews rely on subjective judgments of the researcher, the three measures used in this study generate numerical indices of sharedness. Finally, these measures were chosen because of the compelling case made for each one in empirical research or reviews of SMM research. We recognize that other measures might make for a useful comparison as well (e.g., card sorting); however, to avoid tiring participants, we chose to limit the number of mental model measures provided to participants.

As was mentioned above, little empirical work has been conducted to compare these SMM measures with one another. Indeed, to our knowledge, relations among SMM measures have been assessed in very few studies (Banks & Millward, 2007; Dorsey et al., 1999; Resick et al., 2010). Banks and Millward measured participants’ mental models (using a paired ratings measure) and procedural knowledge (using the proportion of procedures that both team members reported), and participants supplied their own concepts for the procedural knowledge measure. Their nonsignificant correlations suggested that the measures were assessing different constructs. Dorsey et al. examined the accuracy of participants’ mental models (as compared against experts), using paired ratings and concept mapping techniques. They reported that, although positive correlations were found between scores on different measures, the correlations were not high enough to suggest convergent validity. Resick and colleagues compared three methods of assessing mental models: paired ratings, priority rankings, and importance ratings. They reported little evidence of convergent validity using different methods of SMM measurement.

Although each of these studies raises doubts about the convergent validity of SMM measures, none of these studies

were specifically designed to examine the issue with respect to the three familiar approaches that we examine here. In Banks and Millward's (2007) study, participants supplied their own concepts for the procedural knowledge measure, making comparisons across measures difficult. Dorsey and colleagues (1999) compared participant mental models against expert mental models; thus, accuracy was examined instead of similarity. Finally, Resick and colleagues (2010) used measures different from those used in the present study (e.g., importance ratings). In particular, their measures focused on the importance of key decisions, rather than relationships between key tasks. The unique contribution of the present study, therefore, lies in the fact that it examines commonly used SMM measurement strategies that assess task-based sharedness with respect to the same—and thus, comparable—concepts. We believe that this is the best means of conducting meaningful comparisons across SMM measures in order to assess their convergent validity.

Given this situation, and the fact that measures are often used—and interpreted—as if they are interchangeable, it is not surprising that several researchers have highlighted the need to examine them from a convergent validity perspective (DeChurch & Mesmer-Magnus, 2010b; Mathieu, Rapp, Maynard, & Mangos, 2010; Mohammed et al., 2010; Smith-Jentsch, 2009). As Mohammed and colleagues (2010) noted, “there remains a need to use diverse methods, holding both cognitive content and the sample constant. Multiple researchers have noted that it will be difficult, if not impossible, to advance our understanding of [team mental model] measurement without comparing and contrasting multiple measurement methods within the same study” (p. 891). We responded to this challenge by examining the convergent validity of the paired ratings, concept mapping, and causal mapping measures. In addition, to understand whether our results were due to the rating technique or to the way in which similarity was calculated, we used the Pearson correlation and the distance ratio formula to assess similarity for both the paired ratings and causal mapping measures.

Method

Participants

Participants were 192 university students (100 females, 92 males) who participated in this study for credit in an introductory psychology course. On average, they were 18.87 years old ($SD = 2.48$ years). Participants were assigned to dyads ($n = 96$) on the basis of the experimental session for which they registered. Participants reported playing an average of 14.33 h ($SD = 33.36$) of videogames and/or computer games per month; however, only 1 participant reported having experience with the particular computer game used in this study.

Team task

Cannon-Bowers and colleagues (Cannon-Bowers et al., 1993; Mathieu et al., 2000) suggested that SMMs are important in dynamic, flexible, low-communication settings in which team members must rely on their SMMs (rather than verbal communication) to anticipate their teammates' responses. To imitate these dynamic, flexible, low-communication environments, researchers have employed tank (Marks et al., 2000), aircraft (Cooke et al., 2003; Mathieu et al., 2000; Stout et al., 1999), and fire-fighting (Rasker, Post, & Schraagen, 2000) computer simulations. In addition to imitating tank, aircraft, and fire-fighting tasks, the chosen computer games also provide an opportunity to simulate many of the following characteristics: interdependence, need for coordination, challenging goals, and multiple objectives.

Because many of the computer programs used in these past studies were incompatible with current computer systems, we chose to use a newer game. In making this choice, we ensured that the game required task behavior and included task characteristics that typified those games used in earlier SMM research. Specifically, the task used in the present study was a fast-paced, dynamic, challenging flight simulation game in which players balanced objectives of shooting enemy planes while, in turn, avoiding being killed themselves (*IL-2 Sturmovik: Forgotten Battles*; Ubisoft, 2003). A multiplayer version of the game allowed participants to play cooperatively, against the computer, in an effort to shoot enemy planes, avoid shooting teammates, and avoid crashing into the ground. The actions involved in carrying out these subtasks are very similar to those required in earlier SMM research. Moreover, teammates held similar roles and were dependent on one another to achieve their goals; thus, as in earlier SMM research, it was important that players monitored each others' behavior and coordinated their own behavior in accordance with that of the other player. The game was played through a local area network by two players on computers separated by a partition; participants were allowed to communicate with their partners across the partition. Each player was provided with a joystick, a keyboard, and a mouse.

Procedure

First, participants were given a tutorial of the basic mouse, keyboard, and joystick controls required in playing the game. In addition, participants were given information on basic game strategies, such as what altitudes to fly at in order to find the most enemy planes. Participants were allowed to ask questions at this time.

Participants were then allowed to practice the game on their own. After practicing the game, participants were asked to play a mission individually, against the computer,

for 10 min or until they crashed the plane, whichever came first. Participants were told that, during this 10-min mission, their performance would be scored by the computer and that their score would be based on (1) not crashing the plane and (2) shooting down as many of the 16 enemy planes as possible.

Once participants completed the individual mission, they practiced the multiplayer mission. As with the individual mission, the strategies important to the multiplayer mission were explained to participants. The same controls were used in both the individual and multiplayer missions. Participants were told that they might communicate with each other if they wished. Once participants were aware of the multiplayer strategies, they had an opportunity to practice the multiplayer mission with their partners; participants could ask any questions at this time. After practicing the multiplayer game, participants completed the SMM measures (in the order of concept mapping, paired ratings, and then causal mapping), described below, and then played a 10-min multiplayer mission with their partners and against the computer. One problem with administering measures in the same order for all participants is that fatigue effects may have influenced scores on the causal mapping measure. However, we chose this order on the basis of the belief that participants would be more likely to persevere with the more tedious causal mapping measure after having completed the simpler measures. Participants were not allowed to speak to one another while completing the SMM measures.

Measures

Participants were asked to complete three SMM measures: paired ratings, concept mapping, and causal mapping measures. Each of the three rating techniques required participants to make judgments about eight concepts key to playing the flight simulator task. The eight concepts were the following: *avoid enemy fire*, *control direction and speed*, *control landing and flying*, *destroy your target*, *find the enemy*, *protect your teammate*, *react to damage to your aircraft*, and *watch for enemy planes*. Consistent with other SMM research (see Mathieu et al., 2000), these concepts were generated by three subject matter experts (SMEs) and by examination of technical documentation associated with the game. SMEs were individuals who had significant experience with the computer game used in this study. They were psychology and business students who were recruited by the first author on the basis of their interest and enthusiasm for computer games. In determining their suitability as SMEs, we relied on self-reports that they frequently played this computer game, in addition to their evident knowledge of, and enthusiasm for, the game.

Researchers differ as to whether concepts should be generated by SMEs or by the participants themselves. One benefit of having participants provide their own concepts is

that the participants are not restricted by the concepts that SMEs think are important and, as such, the mental models that they generate may be more “realistic.” A disadvantage of having participants provide their own concepts is that they may arrive at different concepts, which makes comparisons across participants difficult. Smith-Jentsch (2009) noted the value of comparing multiple methods of eliciting team cognition while holding content constant. Because the focus in this study was on comparisons across participants and across measures, the same set of concepts was used by each participant, for each measure.

Paired ratings measure Participants were provided with a matrix consisting of the eight concepts mentioned earlier. In the boxes formed by the intersection of concept pairs, participants were asked to rate the pairs on a 7-point scale from 1 (*unrelated*) to 7 (*highly related*). This rating system is consistent with that used by Stout and colleagues (1999).

Convergence on the paired ratings measure was defined as the Pearson correlation between the paired ratings in one member’s matrix and the same paired ratings in the other dyad member’s matrix. Therefore, paired ratings convergences varied from -1 (strong, negative relation) to $+1$ (strong, positive correlation), like any Pearson correlation.

Concept mapping measure Participants were provided with the list of eight concepts and were asked to put them in sequential order on the basis of the time sequence necessary to complete the mission. Convergence on the concept mapping measure was defined as a count of the number of links between concepts that were similar across dyad partners. For example, if one dyad partner placed concepts in order from *control landing/flying* → *avoid enemy fire* → *destroy enemy planes* → *control direction/speed*, while the other dyad partner placed concepts in order from *avoid enemy fire* → *destroy enemy planes* → *control landing/flying* → *control direction/speed*, the dyad would share one link; that is, both had a link between *avoid enemy fire* and *destroy enemy planes*. Note from this example that the links had to be in the same order in both maps (e.g., a link from *destroy enemy planes* to *avoid enemy fire* in one member’s map and from *avoid enemy fire* to *destroy enemy planes* in the other member’s map would not count as a shared link). The linked pairs of concepts did not have to be in exactly the same place in the sequence, however, in order to count as a shared link. Thus, in the example above, the first dyad member placed *avoid enemy fire* and *destroy enemy planes* as the second and third concepts in the sequence, while the other dyad member placed *avoid enemy fire* and *destroy enemy planes* as the first and second concepts in the sequence; this counted as a shared link. A maximum score of 7 would be obtained when dyad partners shared the sequential placement of each of the eight concepts, whereas a minimum score of 0 would be obtained when dyad

partners shared sequential placement of none of the eight concepts.

Causal mapping measure Participants were provided with a matrix consisting of the same eight concepts used in both the paired ratings and concept mapping measures. Although these concepts were arranged in a matrix similar to that used in the paired ratings measure, participants rated these pairs of concepts differently for the causal mapping measure. Instead of responding with simple relatedness ratings, participants were asked to fill in the boxes with numbers from -3 (*A leads to B decreasing greatly*) to 0 (*A leads to B not changing*) to $+3$ (*A leads to B increasing greatly*) in the causal mapping measure, in which “A” and “B” were the two intersecting concepts in question in the matrix. Similarity on the causal mapping measure was calculated using the distance ratio formula (Langfield-Smith & Wirth, 1992), or the summed absolute difference between dyad members’ ratings for each concept pair, divided by the maximum possible difference between the members’ matrices. In the present case, the maximum possible difference for any one matrix cell is 6 —that is, where one dyad member rated the relationship between a pair as -3 , while the other dyad member rated the relationship between a pair of concepts as $+3$. For the whole matrix, the maximum distance is 6×56 cells, or 336 units. The difference between dyad members’ matrices could range from 0 (complete similarity) to 1 (complete dissimilarity).

Analyses

We examined the intercorrelations among the SMM measures in two sets of analyses. The first was based on our sample of 96 pairs of research participants. This sample size compares favorably with that of past research on SMMs (e.g., Marks et al., 2002; Mathieu et al., 2000). Nonetheless, to enhance our power, we also conducted analyses on “pseudo-partners” (dyads whose members did not work together on the task). In considering the logic of this approach, it is critical to remember that each of the SMM assessment techniques relies on the relation between ratings from two individuals. If the three SMM techniques do assess the same construct (degree of “sharedness”), any given pair of participants (real partners or pseudo-partners) should produce similar scores on these techniques. These analyses are described separately below.

Real-partners analysis The real-partners analysis compared the sharedness scores on the three measures for each of the 96 pairs of participants who *did* work together on the task. This analysis was conducted to examine sharedness scores across SMM measures, using data from interacting pairs.

Pseudo-partners analysis The pseudo-partners analysis involved the creation of $18,240$ pseudo-partner pairs. To do

this, we paired each participant with every other participant *except* the person with whom he or she played the game. For each pair of pseudo-partners, three sharedness scores were calculated (one each for the concept mapping, paired ratings, and causal mapping measures); these three SMM scores were then correlated.

In addition, we wished to determine whether the correlations between measures were due to the rating technique per se or to the way in which similarity (sharedness) was calculated. Recall that the paired ratings measure and the causal mapping measure both made use of a set of concepts organized in a grid. Because it is not entirely clear why the paired ratings data needed to be analyzed using a *correlation* and the causal mapping data needed to be analyzed using the *distance ratio formula*, we opted to conduct additional analyses examining these two measurement techniques. This allowed us to determine whether the particular results we observed were due to something about the concept generation technique (i.e., paired ratings and causal mapping) per se or to the way in which sharedness was calculated (i.e., Pearson correlation and distance ratio formula). Specifically, we conducted two additional analyses in which we applied the distance ratio formula to the paired ratings data and applied the Pearson correlation to the causal mapping data.

Results

Table 1 shows similar patterns of results for both the real-partners analysis and the pseudo-partners analysis. Overall, in both sets of analyses, relations among the measures were much lower than would be expected if the measures assess the same construct. In both analyses, the strongest relations were between the paired ratings and causal mapping measures [real partners, $r(76) = -.26$, $p < .05$; pseudo partners, $r(14,971) = -.16$, $p < .001$]. Keep in mind that lower scores obtained with the distance ratio formula and higher scores obtained with the Pearson correlation indicate greater similarity; hence, a negative correlation is expected between scores on the two measures. Note that, although the pseudo-partner correlation was significant, Cohen’s rules of thumb suggest that correlations of the magnitude indicated above actually represent small to medium effects (Cohen, 1988). According to Cohen, an r of $.10$ or less is considered a small effect size, while an r of $.30$ or less is considered a medium effect size. The only other significant correlation was between the concept mapping and paired ratings measures for the pseudo-pairs, $r(15,317) = .05$, $p < .001$. However, in the real-partners analysis, there was no significant relation between scores on the concept mapping measure and scores on either the paired rating or causal mapping measures. Figure 1, a series of scatterplots, illustrates that the low correlations are not a result of nonlinear relationships.

Table 1 Correlations between sharedness calculations on concept mapping, paired ratings, and causal mapping measures for real partners (below the diagonal) and pseudo-partners (above the diagonal)

	1	2	3	4	5
1. Concept mapping		.05**	-.03**	.02*	-.01
2. Paired ratings (correlation)	.12		-.72**	.14**	-.16**
3. Paired ratings (distance ratio)	.03	-.75**		-.07**	.16**
4. Causal mapping (correlation)	-.03	.17	-.09		-.61**
5. Causal mapping (distance ratio)	.10	-.26*	.25*	-.63**	

For real partners, $n = 96$; for pseudo-partners, $n = 18,240$

* $p < .05$

** $p < .01$

We conducted two additional analyses in which we applied the distance ratio formula to the paired ratings data and applied the Pearson correlation to the causal mapping data. As Table 1 illustrates, the correlations were much higher in the analyses that used the same concept generation technique with different sharedness calculations (e.g., paired ratings with both the correlation and distance ratio formula) versus different rating techniques with the same sharedness calculation.

Discussion

In this study, we reasoned that if all the measures we used assessed task-based SMMs, which they were designed to do, they should be strongly correlated with each other. We examined this hypothesis in a sample consisting of participants who worked together on a task (real-partners analysis) and in a larger sample consisting of every possible pair of participants in that data set (pseudo-partners analysis). As was noted above, scores on these measures were not highly correlated.

In particular, scores on the concept mapping measure had low correlations with scores on both of the other measures (regardless of which sharedness calculation was used). To understand why this may have occurred, it is useful to examine features of the measures. The concept mapping measure, for example, may be a more “direct” measure than the others. Recall that the concept mapping measure requires participants to draw the sequential links that they saw among the eight concepts—in effect, producing their own mental model directly. For the other measures, participants merely provided paired ratings from which the researcher derived the mental models. Furthermore, the concept mapping measure may force structure to the concepts when such structures do not exist. For example, consider that participants may perform behaviors simultaneously; the concept mapping measure does not allow for such a structure to exist.

In addition, the measures required participants to make different types of judgments. Instructions for the concept

mapping measure asked participants to place the eight concepts in *sequential order*, from the first concept in the process of completing the flight mission to the last concept in the process of completing the flight mission. Instructions for the paired ratings measure asked for a *similarity rating* between each pair of concepts. Finally, the causal mapping measure required participants to rate each pair of concepts in terms of *causal influence*. Thus, although all of the measures may tap knowledge organization, each may tap knowledge organization in a different way. That is, measures differ in the extent to which they tap into similarity, hierarchy, or causal relationships. In support of this, Banks and Millward’s (2007) study suggested that SMM measures may not be assessing the same thing; that is, some measures may be targeting procedural knowledge organization, whereas others are assessing declarative knowledge organization. Cooke and colleagues (2000) defined declarative knowledge as “the facts, figures, rules, relations, and concepts in a task domain” (p. 153) and procedural knowledge as “the steps, procedures, sequences, and actions required for task performance” (p. 153). For example, a concept mapping measure may be more suited to procedural knowledge, whereas a paired ratings measure may be more suited to declarative knowledge (see Banks & Millward, 2007). An implication of our research, then, is that team members may have different levels of sharedness for different types of knowledge. If one were to train individuals to develop SMMs, the form of knowledge that is being trained must be kept in mind. Essentially, all mental model assessments are not created equal; researchers and practitioners should be cautious in their choice of measures.

We conducted two additional analyses in which we applied to the paired ratings data the distance ratio formula and applied to the causal mapping data the Pearson correlation. As shown in Table 1, results of these analyses revealed higher correlations when the same concept generation technique was used (e.g., paired ratings measure) than when the same sharedness calculation was used (e.g., Pearson correlation). This finding suggests that the nature of the relational judgment that participants are asked to make accounts for more variance in

Fig. 1 Scatterplots illustrating relations between sharedness scores on each of the concept mapping, paired ratings (correlation), and causal mapping (distance ratio) measures for real partners

the relations between these two rating techniques than does the sharedness calculations.

Other explanations for our results may be possible. One is that participants did not understand the task and, as such, did not develop mental models about it. We examined this possibility by looking at “crashes,” a key aspect of task performance. To do so, we examined the dyadic performance scores after participants had a chance to practice the game both alone and as a dyad. Due to technical complications, outcome data were not available for 8 dyads. Of the remaining 88 dyads with outcome data, only 12 dyads

(13.6 %) crashed their planes and, thus, received a score of 0 on the game. These data suggest that, despite the novelty of the task, most participants were able to understand the task at both factual and operational levels. Thus, it seems unlikely that our results were due to a lack of task understanding.

Another possibility is that the low correlations between measures are a result of floor or ceiling effects. If participants did not understand how to complete the measures and simply provided random ratings, sharedness between partners would be low. Little sharedness across all possible pairs would result in restriction of range. Similarly, if participants

Fig. 1 (continued)

had identical ratings and perfect sharedness, restriction of range would result as well. To explore these possibilities, we examined the means and standard deviations for each of the concept mapping, paired ratings, and causal mapping measures, and we inspected the scatterplots in Fig. 1, which illustrate the variability in our data. On the basis of this assessment, it seems unlikely that the low correlations among mental model measures were simply a result of floor or ceiling effects within our data set.

Limitations and future directions

It is important to point out that, in this study, as with much of the SMM research (Cooke et al., 2003; Mathieu et al., 2000), we used a task that relies heavily, albeit not exclusively, on spatial knowledge and skills. It is possible that different results might be obtained using team tasks that place greater emphasis on other types of knowledge and skills. A more comprehensive evaluation of the convergent validity issues that we raise here would include assessment of SMMs associated with a wide range of team tasks.

We administered the three mental model measures to all participants in the following order: concept mapping, paired ratings, and causal mapping. We realize that administering the measures in the same order to all participants may have introduced nonrandom error that served to inflate the correlations that we found. Given that the correlations found in our study may be inflated, the “true” correlations and “true” convergent validity may, in fact, be even lower than those found in this study.

Another limitation is the relatively short time given to participants to learn their task and interact within their dyads. However, many studies in the SMM literature use tasks and interaction periods that last for only a few hours (e.g., Mathieu et al., 2000; Stout et al., 1999). One question for future research is the extent to which convergent validity may change depending on the amount of time spent working together.

The present study, however, highlights what we see as an important and pressing issue: the convergent validity of SMM measures. Clearly, despite a growing body of research on SMMs, more research is needed that compares mental model measures. Future research might compare other mental model measures, such as the interview-based measures discussed earlier. Although the measures considered here were quantitative in nature, many of the interview techniques are more qualitative and would provide an interesting comparison with the results reported here.

Perhaps just as important as determining the convergent validity of mental model measures is determining divergent validity. If mental model measures are truly assessing mental models, their scores should not be highly correlated with scores on measures that are not assessing mental models. Future research should assess divergent validity as well.

In addition, future research might consider the specific mental model measures and mental model definitions used in previous mental model studies, to determine whether researchers’ measurement techniques match their theories. In this study, at least, the results suggest that researchers should define their SMM construct by more than just knowledge

organization. Instead, researchers should specify whether their measures tap, for example, sequence-, similarity-, or cause-based knowledge organization.

Conclusion

In summary, the results of this study and the properties of the measures themselves suggest that researchers should be cautious about comparing results among studies that use different task-based SMM measures. In particular, researchers should make clear the “type” of knowledge organization they are examining in their studies—for example, sequential organization or paired comparisons. Furthermore, and most critically, the results of this study call into some question the convergent validity of one, or more, SMM measures currently used by researchers.

Acknowledgments Support for this research was provided by the Social Sciences and Humanities Council of Canada (Grant 410-2007-0742 awarded to N.A.).

References

- Ambrosini, V., & Bowman, C. (2005). Reducing causal ambiguity to facilitate strategic learning. *Management Learning*, 36, 493–512.
- Badke-Schaub, P., Lauche, K., & Neumann, A. (Eds.). (2007). Team mental models in design [Special issue]. *CoDesign*, 3(1).
- Banks, A. P., & Millward, L. J. (2007). Differentiating knowledge in teams: The effect of shared declarative and procedural knowledge on team performance. *Group Dynamics: Theory, Research, and Practice*, 11, 95–106.
- Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan Jr. (Ed.), *Current issues in individual and group decision making* (pp. 221–246). Hillsdale, NJ: Erlbaum.
- Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior*, 18, 533–558.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. A. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, 7, 179–199.
- Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human Factors*, 42, 151–173.
- DeChurch, L. A., & Mesmer-Magnus, J. R. (2010a). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology*, 95, 32–53.
- DeChurch, L. A., & Mesmer-Magnus, J. R. (2010b). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, 14, 1–14.
- Dorsey, D. W., Campbell, G. E., Foster, L. L., & Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, 12, 31–57.
- Edwards, B. D., Day, E. A., Arthur, W., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, 91, 727–736.
- Ellis, A. P. J. (2006). System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal*, 49, 576–589.
- Gillespie, B. M., Chaboyer, W., Longbottom, P., & Wallis, M. (2010). The impact of organizational and individual factors on team communication in surgery: A qualitative study. *International Journal of Nursing Studies*, 47, 732–741.
- Haig, K. M., Sutton, S., & Whittington, J. (2006). SBAR: A shared mental model for improving communication between clinicians. *Joint Commission Journal on Quality and Patient Safety*, 32, 167–175.
- Heffner, T. S., Mathieu, J. E., & Cannon-Bowers, J. A. (1998, April). *The impact of mental models on team performance: Sharedness, quality, or both?* Poster presented at the annual meeting of the Society for Industrial and Organizational Psychology, Dallas, TX.
- Johnson, T. E., Sikorski, E. G., Mendenhall, A., Khalil, M., & Lee, Y. M. (2010). Selection of team interventions based on mental model sharedness levels measured by the Team Assessment and Diagnostic Instrument (TADI). In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Steel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 335–354). New York: Springer.
- Kivlighan, D. M., Jr., & Kivlighan, D. M., III. (2009). Training related changes in the ways that group trainees structure their knowledge of group counseling leader interventions. *Group Dynamics: Theory, Research, and Practice*, 13, 190–204.
- Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20, 403–437.
- Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social Networks*, 9, 171–186.
- Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10, 359–381.
- Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Techniques, methods, and analytic approaches. *Human Factors*, 42, 242–271.
- Langfield-Smith, K., & Wirth, A. (1992). Measuring differences between cognitive maps. *Journal of the Operational Research Society*, 43, 1135–1150.
- Lavigne, N. C., Salkind, S. J., & Yan, J. (2008). Exploring college students’ mental representations of inferential statistics. *The Journal of Mathematical Behavior*, 27, 11–32.
- 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, 135–144.
- Lim, B. C., & Klein, K. J. (2006). Team mental models and performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, 27, 403–418.
- Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, 87, 3–13.
- Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology*, 85, 971–986.
- 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, 37–56.
- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, 85, 273–283.
- Mathieu, J. E., Rapp, T. L., Maynard, M. T., & Mangos, P. M. (2010). Interactive effects of team and task shared mental models as

- related to air traffic controllers' collective efficacy and effectiveness. *Human Performance*, 23, 22–40.
- Minionis, D. P. (1995). Enhancing team performance in adverse conditions: The role of shared team mental models and team training on an interdependent task. *Dissertation Abstracts International*, 56(2), 1139A.
- Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, 36, 876–910.
- Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, 3, 123–165.
- Muchinsky, P. M. (2009). *Psychology applied to work*. Summerfield, NC: Hypergraphic Press.
- Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 327–345). Norwood, NJ: Ablex.
- Rasker, P. C., Post, W. M., & Schraagen, J. M. C. (2000). Effects of two types of intra-team feedback on developing a shared mental model in command and control teams. *Ergonomics*, 43, 1167–1189.
- Resick, C. J., Murase, T., Bedwell, W. L., Sanz, E., Jimenez, M., & DeChurch, L. A. (2010). Mental model metrics and team adaptability: A multi-facet, multi-method examination. *Group Dynamics: Theory, Research, and Practice*, 14, 332–349.
- Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 1296–1308.
- Salas, E., & Cannon-Bowers, J. A. (Eds.). (2001). Shared cognition in teams and organizations [Special issue]. *Journal of Organizational Behavior*, 22, 87–88.
- Schvaneveldt, R. W. (Ed.). (1990). *Pathfinder Associative Networks: Studies in Knowledge Organization*. Norwood, NJ: Ablex.
- Smith-Jentsch, K. A. (2009). Measuring team-related cognition: The devil is in the details. In E. Salas, G. Goodwin, & C. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pp. 491–508). New York: Routledge.
- 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.
- Spector, P. E. (2006). *Industrial and organizational psychology: Research and practice*. New York: Wiley.
- Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors*, 41, 61–71.
- Tegarden, D. P., Tegarden, L. F., & Sheetz, S. D. (2009). Cognitive factions in a top management team: Surfacing and analyzing cognitive diversity using causal maps. *Group Decision and Negotiation*, 18, 537–566.
- Ubisoft. (2003). *IL-2 Sturmovik: Forgotten Battles*. San Francisco, CA: Maddox.
- Waller, M. J., Gupta, N., & Giambatista, R. C. (2004). Effects of adaptive behaviors and shared mental models on control crew performance. *Management Science*, 50, 1534–1544.