COMPARATIVE STUDY OF MENTAL MODEL RESEARCH METHODS: RELATIONSHIPS AMONG ACSMM, SMD, MITOCAR & DEEP METHODOLOGIES

Tristan E. Johnson[1], Debra L. O'Connor[1], Pablo N. Pirnay-Dummer[2], Dirk Ifenthaler[2], J. Michael Spector[1], and Norbert Seel[1,2]
Learning Systems Institute, Florida State University, USA[1] and
Institute for Educational Sciences, Albert-Ludwigs-University of Freiburg, Germany[2]
Email: tjohnson@lsi.fsu.edu, debraoconnor@gmail.com, pablo.dummer@ezw.uni-freiburg.de, ifenthaler@ezw.uni-freiburg.de, mspector@lsi.fsu.edu, nmseel@gmx.de

Abstract. Measuring and assessing mental models of individuals and teams requires the capturing and analysis of key latent variables. This paper presents and compares four different research methods (ACSMM, SMD, MITOCAR and DEEP) that capture and create a conceptual representation of individual and team mental models. These methods use qualitative and quantitative techniques to investigate a single comparison of different groups or individuals' mental models with another group's mental model or to investigate the comparison of a group or individuals' mental model with themselves at a later time.

1 Introduction

Various fields of human science investigate many factors associated with cognition and mental models. Scientists are continually developing new techniques to capture key latent variables associated with cognition. One specific area of interest is with internal conceptual systems (mental models and schemata). These conceptual systems are theoretical constructs of science that are not observable. Therefore, individuals have to externalize their conceptual systems, and changes in these externalizations are interpreted as changes in the underlying conceptual systems; i.e. researchers can only learn of conceptual systems if individuals communicate their systems (Seel, 1991). A key research interest is to capture and/or create conceptual representations of these internal conceptual systems.

A common analytical requirement is to compare conceptual representations of one group or individual with the representations of another group or individual or even with oneself at a later time. These analytical comparisons are made to represent differences and change of cognitive function. As new techniques and methods are being developed, the question of a valid and reliable measurement of change is one of the central problems of conceptual systems research (Seel, 1999). As such, there are two comparative situations that merit considerations: 1) a single comparison of different groups or individuals with another group and 2) the comparison of a group or individual with themselves over time. Accordingly, the psychological and educational diagnosis of internal conceptual systems presupposes repeated measurements of these systems over the course of a given process (Ifenthaler & Seel, 2005), both for individual and team performance.

Given these research interests, several researchers are conducting a series of studies evaluating four different research methods (ACSMM, SMD, MITOCAR and DEEP) that are specifically aimed at comparing internal conceptual representations of both individuals and teams. The purpose of the study is to determine the strengths for each method, to understand how these tools are unique, and to determine if the specific techniques could be collectively used to ultimately further mental model research and theory development. Each of these research methods employ concept maps as part of the conceptual system elicitation and/or part of the analysis results. Due to the complexity of the studies, this paper will not present the preliminary findings, but will describe each method as well as the series of comparative studies.

2 Methodology Descriptions

While the initial rationale for each method was different, this section seeks to describe each method. The methods employ various techniques (quantitative, qualitative) but each use concept maps as part of their methodological tasks.

2.1 *Analysis Constructed Shared Mental Model (ACSMM) Methodology*

The ACSMM methodology (O'Connor & Johnson, 2004) was developed as a method to determined sharedness among team members. The basic technique involved eliciting concept maps from individual teammates and then using the ACSMM phases to analyze the individual concept maps thereby creating a single concept map that represents the shared components of the team.

Through concept mapping, similarity of mental models can be measured in terms of the proportion of nodes and links shared between one concept map (individual mental model) and another (Rowe, 1995). Utilizing qualitative techniques with an aggregate method of creating an analysis constructed shared mental model (ACSMM), we can capture a more descriptive understanding than by using quantitative techniques.

The ACSMM technique translates individual mental models into a team sharedness map without losing the original perspective of the individual, thereby representing a more accurate representation of team sharedness. The methodology includes several phases: elicitation design and preparation, elicitation of individual team member mental models, coding of individual data, analysis of data to determine what is shared among team members, and construction of the team conceptual representation.

Phase I: Elicitation Design—In order for individuals to create a fully constrained or semi-constrained concept map, a topic analysis is preformed to generate the list of related terms. The process of analysis focuses on determining the various components of a concept and logical relationships, if any. Many topics are ill-structured and allow for multiple logical arrangements. Once the components of the topic have been generated, these terms are then used to help elicit individual mental models.

Phase 2: Individually Constructed Mental Model (ICMM) Elicitation—Prior to the data collection process, a guided practice is provided to demonstrate how to create a concept map. Participants also create a simple concept map by themselves or with others. Concepts maps are reviewed and feedback is provided. If subsequent elicitations are to take place, no guided practice is required. Individuals are then ready to create their individually constructed mental model (ICMM).

Phase 3: ICMM Coding—In order to compare and measure a degree of sharedness in ICMMs, factors such as concepts, links, and cluster combinations are used in analyzing individually constructed concept maps (Doyle, 1998; Jonassen, 1997; Novak & Gowin, 1984). Coding also considers implicit relationships between concepts and structural factors such as directional links and sequence indicators. If the topic of study focuses on a process task, it would be appropriate to use causal measures (directional links, sequence of concepts, and clusters) rather than hierarchical measures and cross-links for coding as suggested by Novak and Gowin (1984).

The ACSMM method accounts for map relatedness at the concept, link, and cluster levels. Because maps appear so unique, the coding strives to reduce the logical, spatial, and structural information and code them so that comparisons among maps can be made. The coding process involves documenting the explicit information on the maps as well as making assessments regarding implicit information. This assessment allows for explication of implicit relationships by considering the spatial, structural and logical information in the map. The coding process involves studying concepts, links, and clusters within each ICMM. The process of coding is much like the process of interpretation. Each map is analyzed and then the researcher codes their interpretation in a spreadsheet or other appropriate tool.

Phase 4: Shared Analysis—After coding the ICMMs, the next step involves an analysis of the ICMM dataset to determine what items were shared by team members. The data tables resulting from ICMM coding are compared for similarity across team members. We suggest starting with a sharedness criterion of 50%. Then determine the shared items for that sharedness level. Depending on the sharedness level sensitivity, the level can be adjusted up or down to increase or decrease the sensitivity. The percentage or number of team members sharing each item is recorded, and all shared items are carried forward for use in constructing the representation of the team's shared mental model, the ACSMM.

Phase 5: ACSMM Construction—The analysis constructed shared mental model (ACSMM) is constructed from the shared ICMM dataset generated in Phase 4. This construction process includes the following steps: Step 1: List Shared Concepts, Step 2: Configure Shared Clusters, Step 3: Configure Shared Links, and Step 4: Configure Non-linked Concepts. The ACSMM methodology is repeated for each subsequent set of concept maps. Once the sharedness between all ICMMs has been identified, the analysis constructed shared mental models can be compared looking at change over time for a specific team or the ACSMM s can be compared among teams to show variation of shared mental model among teams.

2.2 Surface, Matching, and Deep Structure (SMD) Methodology

The Surface, Matching, and Deep Structure (SMD) methodology was developed as a means of calculating the mental model development of students using three different instructional treatments. The basic technique involved eliciting concept maps from individuals at multiple intervals over time as they use a specific instructional intervention. SMD is then used to determine if the concept maps for the individuals change over time and then combine the concept maps of students in each treatment group and compare the findings among treatment groups to determine the effect.

The question of a precise assessment of mental models led to the development of a new methodology named SMD-Technology (Ifenthaler, 2006; Seel, Ifenthaler, & Pirnay-Dummer, 2006). As a basis for the assessment of models, SMD-Technology uses graphical drawings (concept maps) or natural language statements (that is then converting into a concept map) by the subjects. Both, the graphical drawings or natural language statements are transferred into a dataset for further analysis. The SMD-Technology is composed of three levels - Surface, Matching, and Deep Structure.

Phase 1: Surface Structure Analysis—The first level of SMD-Technology constitutes the Surface Structure, on which a rapid and economical assessment of the number of propositions (2 linked nodes) is made possible. The Surface Structure θ is defined as the sum of all propositions P in an individual model.

$$\theta = \sum_{i=0}^{n} P_i \tag{1}$$

Phase 2: Structural Properties Analysis—The assessment of the structural properties of the externalized models is realized on the Matching Structure. The Matching Structure μ is defined as the quantity of links L of the shortest path between the most distant nodes K.

$$\mu = \max_{i,j}\{d(i,j)\} \tag{2}$$

Phase 3: Deep Structure Analysis—The third SMD-Technology level is defined as the Deep Structure. This is the level on which the models are assessed in terms of their semantic structure. The Deep Structure δ is calculated as the similarity (Tversky, 1977) between a shared group model M_{gr} or a domain-specific expert model M_{ex} and the individually assessed model M_i.

$$\delta = \frac{f(A \cap B)}{f(A \cap B) + \alpha \cdot f(A - B) + \beta \cdot f(B - A)} \tag{3}$$

2.3 Model Inspection Trace of Concepts and Relations (MITOCAR) Methodology

Model Inspection Trace of Concepts and Relations (MITOCAR) is a software tool that is based on mental model theory (Seel, 1991) and uses natural language expressions as input data for model re-representation instead of using graphical drawings by the subjects. This is made possible by parsing and corpus linguistics technologies that are similar to those used to implement automated learning of concept hierarchies from text corpora to construct ontology (Maedche et.al. 2002).

One of the goals of MITOCAR is to dig deeper into the semantics of models, especially shared models within groups of experts. This is done during a phase of assessment and an inferential phase. During the assessment phase of MITOCAR, subjects usually go through two different rounds. In the first round they only provide a number of natural phrases (usually sentences) about their specific subject matter. Before the second round, the parser extracts the most frequent concepts from the text corpus of the group and connects them to pairs of concepts.

In the second round the subjects rate how close the concepts are and how sure they are about their assessment. The participants also cluster their concepts from a random list into a list of groups – a method that is sometimes used in knowledge tracking (Janetzko, 1996). Additionally, subjects rate the plausibility of their fellow group members' source phrases.

From this data MITOCAR calculates a proximity vector that represents the whole model that is used to build the model representation (concept map). These data allow one to derive models and even graphical models (Fruchterman & Reingold, 1991; Ganser & North, 1999) that can then be compared in several new ways. This is done with the inferential modules of MITOCAR.

While the semantic comparison of MITOCAR uses traditional measures of similarity (Tversky, 1977) the technology of structural comparison is unique to MITOCAR and can compare models from different subject domains (Pirnay-Dummer, 2006). Up to now we have succeeded in re-representing the models graphically (concepts and structures) on non-directed graphs and comparing them by using conceptual, structural and combined similarity measures. Those measures are controlled using different statistical tests and controls for homogeneity and reliability of the group-consensus models, multidimensional scaling (MDS) of the proximity vectors to test for the representation in 2D and for the comparison and tracking of model complexity.

The final outputs of MITOCAR are the graphical representations to assess a group consensus model from any subject domain and the comparison between different groups of experts. For both outputs there are automated reports which are computing, presenting and interpreting all the above mentioned measures. It has also been investigated how the results of MITOCAR can be used in needs assessment (Pirnay-Dummer, 2006). In addition, the methods of MITOCAR have been applied successfully to tracking user behavior in e-learning environments (Dummer & Ifenthaler, 2005).

Due to the modular design of MITOCAR the assessment tools (re-representation of models by means of natural language, parsing and graph theory) can be separated from the inferential tools (comparing structures and semantics and both). This opens the MITOCAR technology to use on all kinds of model-related data.

2.4 Dynamic Evaluation of Enhanced Problem-solving (DEEP) Methodology

The Dynamic Evaluation of Enhanced Problem-solving (DEEP) methodology is based on a view of learning as becoming more expert-like (Ericsson & Smith, 1991) and more skilled in higher-order causal reasoning and problem solving (Grotzer & Perkins, 2000). The learning-as-becoming-like-an-expert perspective treats learning as a continuing process of growth rather than a single end-point measurable by a simple test. A fundamental assumption is that with regard to complex task performance it is possible to predict performance and assess relative level of expertise by examining a person's conceptualization of the problem space (a problem conceptualization that suggests likely solution alternatives) for specific complex problems.

Learning assessment in DEEP involves: 1) identifying characteristic problems in a particular complex task domain; 2) eliciting both expert and novice patterns of responding to these problems; 3) representing these responses (problem-solution conceptualizations) in a standard form; 4) measuring similarities and differences among experts and novices; and, 5) assessing changes in problem-space conceptualizations over time and with experience.

Learners, either individually or in small groups, are presented with a short problem scenario. They are asked to identify the most relevant factors and issues to consider in developing a solution and then to illustrate the specific nature of relationships among these factors. These annotated causal representations are compared to prior representations and to those of experts to determine progress of learning. Three levels of analysis were applied to these representations. Simple counts of nodes, links, and words per node or link constituted a surface level analysis. Determining the extent of similar nodes and links and how they were connected constituted the structural analysis. Understanding what is said about a particular node or link constituted the semantic analysis. Because the tool allowed nodes and links to be named according to respondent preferences, it was not possible to easily separate the structural and semantic analysis. One purpose of the study was to determine the extent to which a semantic analysis would be required to determine relative level of expertise.

A unique aspect of the DEEP methodology is that it is intended for complex problems involving causal relationships that are interrelated and that may change over time. Moreover, a variety of graphical representations (e.g., semantic networks, flowcharts, causal diagrams, etc.) can be accommodated in this methodology. The graphical representations are converted into standard causal representations (i.e., annotated causal influence diagrams). The reason for using causal representations as the basis for analysis is that such representations reflect internal relationships among factors and components (i.e., problem dynamics) and causal representations can be derived from many other graphical representations when the appropriate documentation is provided (e.g., the descriptions of individual factors). The DEEP methodology supports assessments of individual learning in problem-centered instructional modules, which can be used in evaluating problem-centered instructional programs (Baker, 1999; Herl et al., 1999). The data provides information about how well individual learners are doing in specific problem-centered modules. This enables teachers to adjust instructional supports appropriately and it enables learners to adjust their learning strategies. Additionally, instructional designers are provided with information on which to base specific modifications to the structure and sequence of various learning activities.

Variations of this methodology have been effectively demonstrated in simpler domains (Herl et al., 1999; Novak, 1998; Schvaneveldt, 1990). Those who have employed an analogous method for simpler learning tasks have relied on: simple quantitative measures for measures of similarity to expert responses (e.g., presence/absence of salient features and their location in a concept map); and, qualitative analysis of responses, which are notoriously time-intensive and costly and, consequently, hardly ever used when a laboratory or demonstration effort of a learning environment or instructional system scales up to full-scale implementation and deployment.

2.5 Methodology Comparison

The general characteristics of each of these methods have similarities and differences (Table 1). In each case, the methodology collects data that then is transformed before specific analysis can take place. Analytical methods for each method convert the data to emphasize a decomposition or composition of the initial data set. Comparisons describe the functionality of the methods to simultaneously compare/analyze unlimited or paired groups/individual data.

3 Studies: Method Comparisons

In order to meet the goals of our research (determine the method strengths, unique characteristics, and collective viability), we have setup a series of comparative studies to answer these questions. A series of pair-wise comparative studies are followed in order to detect analytical differences among the mythologies. There are two points of method comparisons: data conversion techniques and data analysis techniques.

Each of these methodologies have been used in specific settings, however the work represented in these comparisons is the first project utilizing a single set of data to compare the results of these four methodologies. There are six studies involving two parts each. The studies compare the following pairs of methods: 1) ACSMM & MITOCAR, 2) SMD & MITOCAR, 3) ACSMM & SMD, 4) DEEP & SMD, 5) DEEP & MITOCAR, and 6) DEEP & ACSMM.

Method	Data Collection	Analysis	Data Conversion	Comparison Fxn
ACSMM	Concept Map	Qualitative with Descriptive Statistics—analysis is done mostly by hand	Structural Decomposition into 3 Key Categories (manual), Structural Re-composition into 1 Representation	Unlimited comparisons, show details relating to concepts
SMD	Concept Map or Natural Language	Quantitative—analysis is calculated using tools	Structural Decomposition into 3 Key Categories (manual)	Unlimited comparisons
MITOCAR	Natural Language	Quantitative—analysis included multiple calculations using tools	Structural Composition into 1 Category (automatic)	Paired comparisons for semantic and structural model distance measures
DEEP	Annotated Influence Diagrams	Quantitative/Qualitative—analysis is done mostly by hand	Structural Decomposition into 3 Key Categories (automatic)	Unlimited comparisons, show details relating to concepts

Table 1: Methodologies' features and techniques comparisons.

We first look at the data analysis techniques. Since each method has different techniques for data conversion prior to data analysis, the initial comparison among the methods will focus on the specific analytic techniques employed. As such, the comparisons begin with the same dataset in order to control for the analysis phase of each method. This involves taking a set of decomposed data from concept maps or natural language and the analysis techniques used in each method will be employed and the output will be compared.

The second part compares the data conversion techniques. This involves collecting data according to the method specifications and then going through the process of data conversion. The converted data are compared for similarities and differences.

For example, the application of the SMD-Technology in different subject domains and the comparison with other quantitative methodologies, e.g. MITOCAR (Pirnay-Dummer, 2006), or qualitative methodologies, e.g. ACSMM (O'Connor & Johnson, 2004) and DEEP (Spector & Koszalka, 2004), could cross-validate the SMD-Technology (and vice versa) and give a more detailed understanding of the changes of mental models within individuals and teams.

4 Summary

From the study of these methodologies, we hope to better qualitatively and quantitatively represent individual and team mental models thereby facilitating greater understanding of the notion of individual and team processes and the development of mental models. We hope to better understand mental model development by the comparison of individuals to experts.

Further, by comparing the teams at various points during team processes, we should be able to determine how team metal models change over time. With each methodology we test the progression/development of conceptual representations with each other from the initial state to the post state and the similarity between the individually constructed conceptual representations. Not only will this information benefit further study in individual and team dynamics, but also if we can identify how team mental models change over time and find indicators of why they change, we should be able to develop methods for improving overall individual and team performance.

Conducting experiments for cross validation gives us a chance to add quantitative control for qualitative assessment methods looking at individuals and aggregation to be used for teams. Likewise this type of experimentation gives us added qualitative control for quantitative assessment methods.

References

Baker, E. (1999, July). Technology: How do we know it works? Paper presented at the Secretary's Conference on Educational Technology, July 12-13, 1999, Spector – Page 18 of 23 Washington, D.C. (Available at http://www.ed.gov/Technology/TechConf/1999/whitepapers/paper5.html).

Doyle, J. K., Radzicki, M. J., & Trees, W. S. (1998). Measuring Change in Mental Models of Dynamic Systems: An Exploratory Study.Unpublished manuscript.

Dummer, P., Ifenthaler, D. (2005). Planning and assessing navigation in model-centered learning environments. Why learners often do not follow the path laid out for them. In G. Chiazzese , M. Allegra, A. Chifari & S. Ottaviano (Eds.), Methods and technologies for learning, pp. 327-334. Southampton, Boston: WIT Press.

Ericsson, K. A., & Smith, J. (Eds.) (1991). Toward a general theory of expertise: Prospects and limits. New York: Cambridge University Press.

Fruchterman, T.M.J., & Reingold, E.M. (1991). Graph drawing by force-directed placement. Software - Practice and Experience, 21 (11), 1129-1164.

Ganser, E.R, & North, S.C. (1999). An open graph visualization system and its applications to software engineering. Software - Practice and Experience, 00 (S1), pp. 1-5.

Grotzer. T. A., & Perkins, D. N. (2000, April). A taxonomy of causal models: The conceptual leaps between models and students' reflections on them. Paper presented at the National Association of Research in Science Teaching (NARST), New Orleans, LA, 28 April - 1 May 2000.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. L. W. K., Bianchi, C., Wang, S., Mayer, R., Lee, C. Y., Choi, A., Suen, T., & Tu, A.(1999). Final report for validation of problem solving measures. CSE Technical Report 501. Los Angeles: CRESST.

Ifenthaler, D. (2006, in press). Diagnose lernabhängiger Veränderung mentaler Modelle. Entwicklung der SMD-Technologie als methodologisches Verfahren zur relationalen, strukturellen und semantischen Analyse individueller Modellkonstruktionen. Freiburg: Universitäts-Dissertation.

Ifenthaler, D., Seel, N.M. (2005). The measurement of change: learning-dependent progression of mental models. Technology, Instruction, Cognition and Learning, 2 (4), 317-336.

Janetzko, D. (1996) Knowledge Tracking. A Method to Analyze Cognitive Structures. Freiburg: IIG-Berichte 2.

Jonassen, D. H., Reeves, T. C., Hong, N., Harvey, D., & Peters, K. (1997). Concept Mapping as Cognitive Learning and Assessment Tools. Journal of Interactive Learning Research, 8(3/4), 289-308.

Maedche, A., Pekar, V., Staab, S. (2002) Ontology learning part one - on discovering taxonomic relations from the web. In Proceedings of the Web Intelligence Conference, pp. 301-322. Springer

Novak, J. D. (1998). Learning, creating and using knowledge: Concept maps as facilitative tools in schools and corporations. Mahwah, NJ: Erlbaum.

Novak, J. D., & Gowin, D. B. (1984). Learning How to Learn. Cambridge Cambridgeshire; New York: Cambridge University Press.

O'Connor, D.L., Johnson, T.E. (2004). Measuring team cognition: Concept mapping elicitation as a means of constructing team shared mental models in an applied setting. First International Conference on Concept Mapping, September 14-17, 2004, Pamplona, Spain.

Pirnay-Dummer, P. (2006, in press). Expertise und Modellbildung: MITOCAR. Freiburg: Universitäts-Dissertation.

Rowe, A. L., & Cooke, N. J. (1995). Measuring Mental Models: Choosing the Right Tools for the Job. Human Resource Development Quarterly, 6(3), 243-255.

Schvaneveldt, R. W. (Ed.) (1990). Pathfinder associative networks: Studies in knowledge organization. Norwood, NJ: Ablex.

Seel N.M., Ifenthaler, D., Pirnay-Dummer, P. (2006, in press): Mental Models and Problem Solving: Technological Solutions for Measurement and Assessment. In: Strobel, J. Jonassen, D. Learning by Modelling Systems, Mawah, NJ: Erlbaum

Seel, N.M. (1991). Weltwissen und mentale Modelle. Göttingen: Hogrefe.

Seel, N.M. (1999). Educational diagnosis of mental models: Assessment problems and technology-based solutions. Journal of Structural Learning and Intelligent Systems, 14 (2), 153-185.

Spector, J. M., & Koszalka, T. A. (2004). The DEEP methodology for assessing learning in complex domains (Final report to the National Science Foundation Evaluative Research and Evaluation Capacity Building). Syracuse, NY: Syracuse University.

Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2005). Causal maps, mental models and assessing acquisition of expertise. Technology, Instruction, Cognition and Learning, 3.

Tversky, A. (1977). Features of similarity, Psychological Review, 84, 327-352. From: Spector, J. M., Dennen, V. P., & Koszalka, T. A. (2005). Causal maps, mental models and assessing acquisition of expertise. Technology, Instruction, Cognition and Learning, 3.