

Model-Based Methods for Assessment, Learning, and Instruction: Innovative Educational Technology at Florida State University

Valerie J. Shute, Allan C. Jeong, J. Michael Spector, Norbert M. Seel, and Tristan E. Johnson

Abstract In this chapter, we describe our research and development efforts relating to eliciting, representing, and analyzing how individuals and small groups conceptualize complex problems. The methods described herein have all been developed and are in various states of being validated. In addition, the methods we describe have been automated and most have been integrated in an online model-based set of tools called HIMATT (Highly Interactive Model-based Assessment Tools and Technologies; available for research purposes at <http://himatt.ezw.uni-freiburg.de/cgi-bin/hrun/himatt.pl> and soon to be available on a server at Florida State University). HIMATT continues to expand in terms of the tools and technologies included. Our methods and tools represent an approach to learning and instruction that is now embedded in many of the graduate courses at Florida State University and also at the University of Freiburg. We call our approach model-based because it integrates representations of mental models and internal cognitive processes with tools that are used to (a) assess progress of learning, and (b) provide the basis for informative and reflective feedback during instruction.

Keywords Belief networks · Causal diagrams · Cognitive modeling · Concept mapping · Mental models · Model-based assessment · Technology-based assessment

Introduction

Knowledge is no longer an immobile solid; it has been liquefied. It is actively moving in all the currents of society itself (Dewey, 1915, p. 25).

This quote by John Dewey nearly 100 years ago is particularly relevant now. That is, in our increasingly technological society, understanding the ebb and flow of

V.J. Shute (✉)
Instructional Systems Program, Educational Psychology and Learning Systems Department,
College of Education, Florida State University, Tallahassee, FL 32306, USA
e-mail: vshute@fsu.edu

mental models, and figuring out how to help people develop and hone good mental models—alone and in collaboration with others—are important goals with potentially large educational and economic benefits (e.g., Seel, 1999a; Shute & Zapata-Rivera, 2008; Spector, Dennen, & Koszalka, 2006).

Mental models have been researched extensively over the past several decades, and have been implicated in many phenomena that are fundamental parts of human cognition, such as the ability to reason—inductively and deductively—about complex physical and social systems, to generate predictions about the world, and to form causal explanations for what happens around us (e.g., Gentner & Stevens, 1983). As part of the Instructional Systems program at FSU, we have been building on the theoretical and empirical foundations of mental model research. Currently, we're using a model-based approach to design and develop innovative educational technologies to (a) represent mental models (i.e., externalized constructions of internalized structures), (b) analyze their changes over time, and (c) create instructional interventions to support learning. We have also been developing tools to aggregate mental model representations, compare those representations, and identify the reasons for change. We call our approach model-based because it integrates representations of mental models and internal cognitive processes with tools that are used to assess progress of learning, and provide the basis for informative and reflective feedback during instruction.

This chapter focuses on the role of internal representations (i.e., mental models) in interpreting experience and making sense of things. While internal representations are not available for direct and immediate observation, we accept the general notion that the quality of internal representations is closely associated with the quality of learning. So, to help instructional designers and educational technologists improve support for learning, we have devised a theoretical foundation and a collection of tools to facilitate assessment and to provide personalized, reflective, and meaningful feedback to learners, particularly in relation to complex and challenging problem-solving domains.

We first review the foundations of our model-based approach to assessment, learning, and instruction. Then we discuss a variety of tools and technologies that we have been developing and validating in a number of different problem-solving domains. We expect these tools and technologies to evolve and perhaps be replaced with other tools and technologies. We also expect that the underlying foundations will evolve as scientists learn more about specific human learning mechanisms. However, in the near-term we believe that a model-based approach to learning and instruction supported by the kinds of tools described here are important for progress in educational technology research.

Foundations of Our Model-Based Approach

Our model-based research and development rests on two foundations: (1) mental models research and systems thinking (*internal* constructs and processes), and (2) concept maps and belief networks (*external* representations and entities). We aim to assess the quality of the former via aspects of the latter.

Internal Constructs: Mental Models and Systems Thinking

As philosophers have long argued, we create internal representations of things that we experience. The most direct statement of this capability can be found in Wittgenstein's *Tractatus Logico-Philosophicus*: "*We make to ourselves pictures of facts*" (Wittgenstein, 1922; for an online version, see <http://www.kfs.org/~jonathan/witt/tlph.html>). Psychologists have expanded and elaborated this notion of internal representations in the last several decades to include the key constructs of mental models and schema. Because these internal constructions are vital to how people come to make sense of and learn about the world, we place particular emphasis on such internal representations as the basis for developing proper support for learning.

What is the nature of these internal constructs? We each hold many different beliefs about the world based on our unique experiences, and we can conceive of these beliefs as structures or networks of concepts (nodes) and their relationships (links). Some beliefs may be more accurate than others, depending on the existence and quality of the underlying evidence. Some beliefs may be more or less firmly held, depending on the strength of the links. As educators, we would like to be able to make valid inferences about what a person knows and believes. Beliefs are not fixed and unchanging.¹ Instead, belief structures or mental models: (a) are incomplete and constantly evolving; (b) may contain errors, misconceptions, and contradictions; (c) may provide simplified explanations of complex phenomena; and (d) often contain implicit measures of uncertainty about their validity that allow them to be used even if incorrect (e.g., Ifenthaler & Seel, 2005; Seel, 2003). So knowledge and beliefs can change, but seldom randomly—there are typically triggering events that provide the impetus for change. We will explore this in more detail later when we describe our tools that model evolving belief networks, and attempt to identify the basis for change.

Mental models also play a key role in qualitative reasoning. For example, Greeno (1989) argued that model-based reasoning in specific situations (e.g., physics, economics, and so on) occurs when an individual interacts with the objects involved in a situation in order to manipulate them mentally so that the cognitive operations simulate (in the sense of thought experiments) specific transformations of these objects which may occur in real-life situations. In line with symbolic models of cognition, it is widely recognized that the construction of mental models necessarily presupposes the use and manipulation of signs (used as index, icon, or symbol) to the extent that mental models are used to organize the symbols of experience and thinking to achieve a systematic representation of this thinking as a means of understanding and explanation (Seel & Winn, 1997; Seel, 1999a). Accordingly, in cognitive and educational psychology, mental models are considered *qualitative mental representations which are developed by individuals (or groups) on the basis of their available world knowledge (or beliefs) aiming at solving problems or acquiring competence in a specific domain.*

¹To illustrate, your belief that *Pluto is a planet* likely changed in 2006 when the International Astronomical Union decided to re-classify Pluto as a "dwarf planet."

In short, mental models are cognitive artifacts; that is to say they are inventions of the mind that represent, organize, and restructure a person's knowledge and beliefs in such a way that even complex phenomena of the (observable or imagined) world become plausible. For our purposes, complex phenomena include social, technological, and natural systems, whereby a system is understood as a designed entity (designed by humans or by nature) that maintains its existence and functions as a whole through the dynamic interaction of its parts. A system's interdependent parts form a unified whole, driven by a purpose; and the various parts generally attempt to maintain stability or equilibrium through feedback (examples of such systems include human respiration, energy consumption in a hybrid vehicle, and the caucus system to determine U.S. presidential nominations). The ability to understand and reason about such complex systems is often called *systems thinking* and has been identified as an essential skill for the 21st century (Federation of Scientists, 2006). The International Board of Standards for Training, Performance and Instruction also regards systems thinking as a fundamental skill (see <http://www.ibstpi.org>).

External Entities: Concept Maps, Causal Models, and Belief Networks

As we mentioned earlier, our high-level goal is to infer the quality of presumed internal constructs and processes (mental models and systems thinking) via valid techniques that seek to externalize internal, invisible structures. This task is made simpler because humans have developed an amazing ability to talk about (or otherwise represent) their private, internal representations (thoughts, feelings, beliefs). Discourse is a vital component of most learning experiences, and Wittgenstein (1953) recognized the criticality of discourse in his later work, referring to this ability as engaging in what he termed *language games*. A language game is specific to a group of people who share a common purpose or enterprise. A language game is context specific as well as specific to a community of speakers. Key aspects of a language game include a common vocabulary, a set of accepted conventions and rules, and sets of expected statements and responses. This notion is relevant to our focus on assessing learning in complex domains. That is, what people say and how they relate various aspects of a problem situation are indicative of their understanding. Examining these external representations, then, provides evidence of the nature and quality of the internal representations that are the basis for action. These external representations come in (and can be shaped into) various forms including concept maps, causal models, and belief networks.

A *concept map* is a diagram showing the relationships among concepts. Concepts are connected with labeled arrows, often in a hierarchical structure. The relationship between concepts is specified via linking phrases, such as, "results in," "is required by," or "is part of." Concept mapping is the term used for visualizing the relationships among different concepts. Concept maps are frequently used to examine and assess learners' understanding of complex domains and their progress

towards increased understanding (e.g., Spector & Koszalka, 2004). However, many of the current studies on concept maps focus on well-defined problems (Freeman & Urbaczewski, 2001; Ruiz-Primo & Shavelson, 1996) and are restricted to a closed format where concepts are provided by the evaluator (Zelev, 2004). This closed format, while making it easy to score, provides little insight into the actual process of learning, or more specifically, the cognitive processes underlying the changes learners make to their concept maps. To examine the underlying processes of concept mapping, researchers can provide learners with the opportunity to create and annotate nodes and links in their concept maps (Alpert, 2003), yielding richer and more accurate maps (Spector et al., 2006). These annotated maps enable researchers to access, study, and determine some of the cognitive processes that underlie, trigger, and explain changes (both good and bad) in learners' mental models.

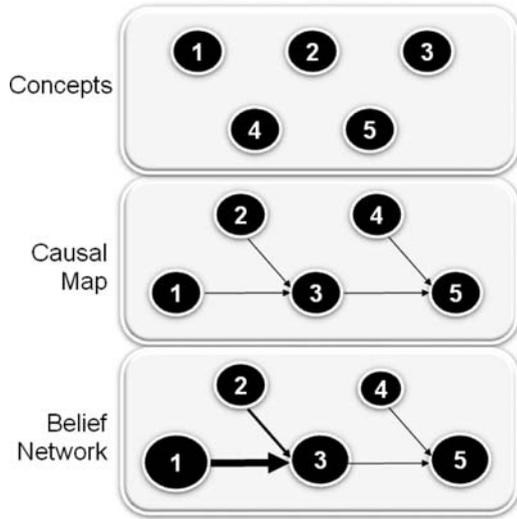
A *causal model* is like a concept map, only instead of allowing any type of link between nodes or concepts, it uses cause and effect logic to describe the behavior of a system. In traditional causal modeling, a network of variables is developed and the causal relationships between variables are explicitly delineated. It is a model in which the variables of interest (the dependent variable or variables) are related to various explanatory variables (or causal variables) based on a specified theory.

A *belief network* is a probabilistic graphical model that represents a set of variables and their probabilistic independencies. This goes beyond causal models (e.g., "A causes B") in that belief networks allow for the specification of degree or level of relationships (e.g., "If A occurs, that will strongly influence B"). Belief networks are in line with our goal of wanting to represent individuals' understanding of complex phenomena (e.g., systems thinking), and encompass a wide range of different but related techniques which deal with reasoning under uncertainty. Both quantitative (mainly using Bayesian probabilistic methods) and qualitative techniques can be used to interpret belief networks. Our approach involves representing a learner's (or group of learners') current set of beliefs about a topic by overlaying Bayesian networks (Pearl, 1988) on top of students' causal maps. Again, this allows us to model and to question the degree to which relationships among concepts/nodes hold as well as the *strength* of the relationships. In addition, prior probabilities can be used to represent preconceived beliefs. A probabilistic network provides us with a richer set of modeling tools that we can use to represent the degree to which people ascribe to a particular belief structure (for more, see Shute & Zapata-Rivera, 2008).

Figure 1 illustrates a simplified example of the progression from concepts, to causal maps, to belief nets when Bayesian networks are overlaid to specify structure, node size, and links (i.e., type, directionality, and strength of association). Furthermore, evidence can be attached to each node-relationship which either supports or counters a given claim.

The *size* of the node in the belief network indicates a given node's marginal probability (e.g., $p(\text{node } 1 = \text{True}) = 0.55$ —a medium node with a slightly better-than-average probability of being true). *Links* illustrate the perceived relationships among the nodes in terms of *type*, *direction*, and *strength*. *Type* refers to the probabilistic or deterministic representation—defining the nature of the relationship (in this case, causes). The *strength* of the relationship is shown by the thickness of the

Fig. 1 Progression from concepts to causal map to belief network (from Shute & Zapata-Rivera, 2008)



link, and the *direction* indicates that the relationship has an origin and a destination. The belief structure in Fig. 1 models the beliefs of a person (or group of people) that, for example: (a) nodes 1 and 2 exist, (b) the current probability of node 1 is greater than node 2, and (c) there is a positive and strong relationship between nodes 1 and node 3 (represented by a thick line).

When comparing two belief nets (e.g., the same student at different points in time; a student with an expert), they may contain the same concepts, but the size of the respective nodes, the directionality of relations, and/or the strength of the links may be very different. Because we have chosen to use Bayesian networks to

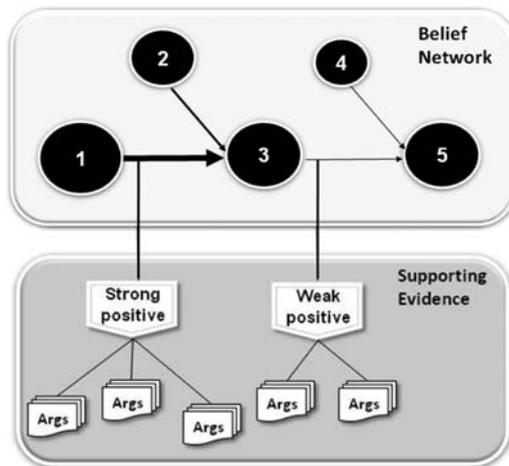


Fig. 2 Supporting evidence underlying an example belief network

represent belief structures, this enables us to examine not only (a) the structure of the map, but also (b) the content (nodes and links), as well as (c) the underlying evidence that exists per structure (and per node). That is, as part of creating a current belief structure, the student arranges concepts and establishes links, and he or she includes specific evidence (sources) per claim (i.e., arguments and relevant documentation in support of, or in opposition to a given claim). Figure 2 shows a generic belief network with its supporting evidence attached.

Tools and Technologies

The complexity and quantity of data that can be produced in relation to concept maps, causal models, and belief networks has motivated our design and development of new software tools and methods. These tools are designed to produce numerical indices (e.g., structural similarity between a pair of maps) as well as visual representations (often automatically generated) that can simultaneously reveal: (a) global patterns emerging in the maps and the cognitive processes, events, and/or conditions that trigger changes in the maps; (b) the extent to which the changing patterns are progressing toward a target model; and (c) detailed and precise information on what and where changes are occurring within the maps.

To date, we have developed six tools and technologies, detailed in this section, for purposes of assessing mental models and using that information as the basis to improve learning. The names of the six tools are: DAT, jMap, DEEP, ACSMM, SMD, and MITOCAR. The last four have been integrated in a Web-based assessment tool kit called HIMATT (Highly Interactive Model-based Assessment Tools and Technologies), while DAT and jMap are in the process of being integrated. These tools are currently available at <http://himatt.ezw.uni-freiburg.de/cgi-bin/hrun/himatt.pl>) and soon will be available on a server at Florida State University. The six tools are summarized below.

DAT (Discussion Analysis Tool)

As described earlier, belief networks represent and analyze links and nodes in causal maps. Similarly, sequential analysis (Bakeman & Gottman, 1997) has been used to model and analyze *sequential links* between behavioral events to determine how likely one given event is followed by another given event. Jeong (2004, 2005) developed DAT to compute the transitional probabilities between dialog moves observed in online debates. For example, DAT produces a transitional probability matrix to report the percentage of replies to stated arguments (ARG) that are challenges (BUT) vs. explanations (EXPL) vs. supporting evidence (EVID); and the percentage of replies to challenges that are counter-challenges vs. explanations vs. supporting evidence (see Fig. 3).

The matrix shown in Fig. 3 represents actual data from an online debate. The circled number indicates that 48% of all replies to opposing arguments (–ARG) were

	+ARG	+BUT	+EXPL	+EVID	-ARG	-BUT	-EXPL	-EVID	Replies	No Replies	Givens	Reply Rate	% replies	% givens
+ARG	.02	.03	.25	.17	.00	.49	.00	.04	213	21	127	.83	.25	.10
+BUT	.00	.10	.05	.10	.00	.66	.06	.03	135	162	289	.44	.16	.23
+EXPL	.00	.02	.08	.15	.00	.67	.06	.02	52	64	112	.43	.06	.09
+EVID	.00	.00	.10	.13	.00	.71	.00	.06	31	50	84	.40	.04	.07
-ARG	.00	.48	.03	.02	.00	.02	.26	.19	174	21	124	.83	.20	.10
-BUT	.00	.61	.11	.02	.00	.08	.08	.09	157	185	328	.44	.18	.26
-EXPL	.00	.56	.13	.00	.00	.04	.17	.10	52	56	102	.45	.06	.08
-EVID	.00	.62	.05	.03	.00	.00	.15	.15	39	49	81	.40	.05	.06

Fig. 3 Transitional probability matrix produced by DAT

challenges (+BUT), for this group of students. DAT also produces a corresponding z-score matrix to identify and automatically highlight transitional probabilities that are significantly higher/lower than expected probabilities to determine which behavioral sequences can be considered a “pattern” in a group’s behaviors.

To visually and more efficiently convey the complex data revealed in the transitional probability matrix, DAT converts the observed probabilities into transitional state diagrams (see Fig. 4). Potential differences in behavior patterns between experimental groups—such as groups with students that are high vs. low in intellectual openness (Jeong, 2007)—can be easily seen by juxtaposing state diagrams and observing the differences in the thickness of the links between events (signifying the strength of the transitional probabilities between given events).

Once specific patterns and differences are identified between particular events, DAT automates the process of tabulating raw scores that reveal, for example, how

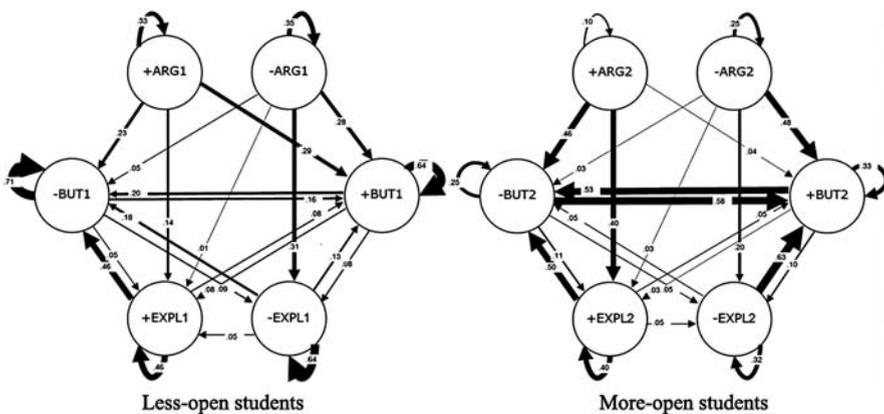


Fig. 4 Transitional state diagrams of response patterns produced by less- vs. more-intellectually open students

many challenges are elicited by *each* argument, or how many explanations are elicited by *each* challenge. These raw scores can then be used to test for differences in the mean number of challenges elicited per argument and the mean number of explanations elicited per challenge between two or more experimental groups using two-way analysis of variance.

jMap

Another tool we have recently developed is an Excel-based software application called *jMap* (Jeong, 2008; Shute, Jeong, & Zapata-Rivera, 2008), designed to accomplish four specific goals: (1) elicit, record, and automatically code mental models; (2) visually and quantitatively assess changes in mental models over time; (3) determine the degree to which the changes converge towards an expert’s or the aggregated group model; and (4) measure how specific social and/or cognitive events and processes (e.g., degree to which evidence is presented, degree to which the merits of presented evidence is thoroughly cross-examined) trigger changes in mental models.

Using *jMap*, students (and experts, as warranted) individually create their causal maps using Excel’s *autoshape* tools. Causal link strength is designated by varying the densities of the links. The strength of evidentiary support for a link (not shown in Fig. 5) is designated by dashed lines where longer dashes convey stronger evidence. *jMap* automatically codes each map into a transitional frequency matrix by inserting two values into each matrix cell—*causal strength* of the links between nodes

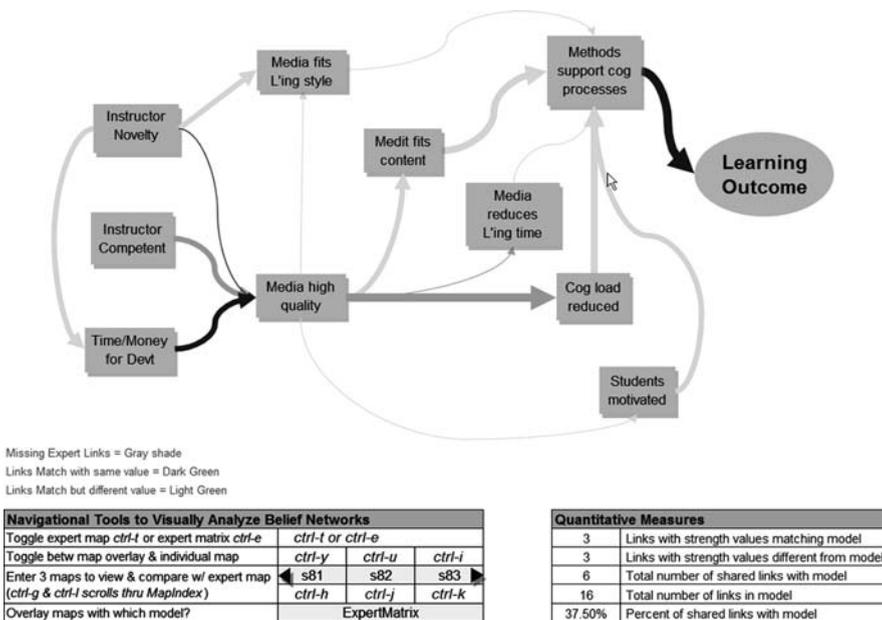
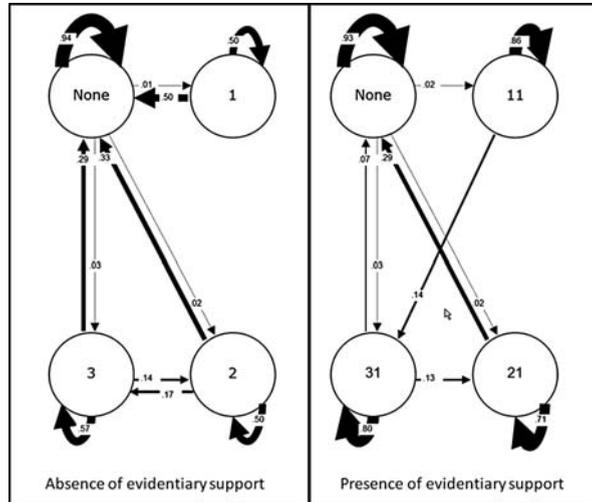


Fig. 5 Student’s causal map superimposed over an expert’s map

Fig. 6 Transitional state diagrams revealing how absence vs. presence of evidentiary support affects how causal link strengths change over time



(1 = weak, 2 = moderate, 3 = strong) and *strength of evidentiary support* underlying the links (0 = none, 1 = weak, 2 = moderate, 3 = strong). Figure 5 shows a student’s map overlaid on an expert’s map.

Once maps are tabulated, *jMap reproduces and presents each student’s map using a standardized map template* (e.g., based on an expert’s map). Using this approach, the maps of two or more learners and/or experts can be superimposed over one another. Visual comparisons can be performed between: (a) student A’s map produced at time 1 vs. time 2; (b) student A’s map vs. an expert’s map; and (c) a group map (produced by aggregating all maps across all students) vs. an expert’s map. Users (e.g., teachers, researchers, students, etc.) can rapidly toggle between maps produced over different times to animate and visually assess how maps change over time and the extent to which the changes are converging toward an expert or collective group map. Additional *jMap* tools enable users to compile raw scores to: (a) compare quantitative measures (e.g., test the rate of change in the number of matching links); and (b) sequentially analyze and identify patterns in the way causal link strengths change over time using both *jMap* and *DAT* software combined. Figure 6 shows state diagrams for two groups of students—those who did not include evidentiary support in their causal maps (left) and those who did (right). The presence of evidence appears to stabilize students’ causal maps.

ACSMM

Our next tool is called Analysis Constructed Shared Mental Model (ACSMM). This methodology was developed primarily as a way to assess *team processes* and predict team performance by determining the degree of overlap or “sharedness” of mental

models among team members (O'Connor & Johnson, 2004; Johnson & O'Connor, 2008). The ACSMM methodology is based on the understanding that: (1) teams with similar ways of thinking are likely to work more effectively together than teams with different ways of thinking (Cannon-Bowers & Salas, 1998; Guzzo & Salas, 1995; Hackman, 1990), and (2) the degree to which a team shares similar conceptualizations is seen as a key indicator of overall team performance (Salas & Cannon-Bowers, 2000). That is, as teammates interact with one another, they begin to share knowledge. This knowledge sharing enables them to create cues in a similar manner thus helping them to make compatible decisions and to take proper actions (Klimoski & Mohammed, 1994; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000). Shared knowledge can help team members understand what is occurring with regard to the task at hand, develop accurate expectations about future member actions and task states, and communicate meanings efficiently.

A common method for assessing team knowledge has been via concept maps (e.g., Herl et al., 1999; Ifenthaler, 2006; O'Connor & Johnson, 2004; O'Neil, Wang, Chung, & Herl, 2000). Through concept mapping, similarity of mental models can be measured in terms of the proportion of nodes and links shared between one concept map and another (Rowe & Cooke, 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 only quantitative techniques. Specifically, ACSMM can retain not only the logical structure, but also a general semantic meaning of the shared mental models.

How does it work? ACSMM involves a methodology where individually-constructed mental models (ICMMs) are elicited, and then a technique is used such that the sharedness is determined not by the individuals who provided their mental models, but by an analyst or analytical procedure. That is, ACSMM provides a set of heuristics to code the individual maps and then transform the ICMMs into a team map (i.e., the ACSMM) without losing the original perspective of the individual (see Fig. 7).

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

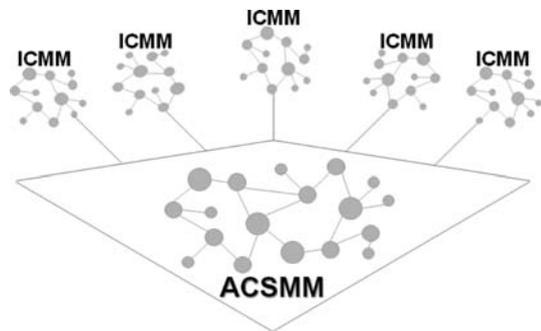


Fig. 7 Relationship between ICMMs (Individual Constructed Mental Model) and ACSMM (Analysis Constructed Shared Mental Model)

of the team conceptual representation (i.e., the team map). One of the key features of ACSMM is that this method accounts for map relatedness at the concept, link, and cluster levels. Because individual maps are so unique, the coding strives to reduce the spatial, structural, and logical information thereby permitting comparisons among maps. The coding process involves documenting the *explicit* information on the maps as well as making assessments regarding *implicit* information, which allows for explication of implicit relationships by considering the spatial, structural, and logical information in the map. The process of coding each ICMM is much like the process of interpretation. That is, each map is analyzed and then the researcher codes her interpretation in a spreadsheet (or other appropriate tool). At least one of two congruency guidelines must be satisfied before coding implicit clusters or links: (1) logical and spatial congruency, or (2) logical and structural congruency.

This technique was initially carried out by hand, but there are parts of the methodology that are automated and can be carried out in HIMATT. ACSMM is designed to quickly and easily capture mental models and that is the extent at intervening in the teams' activities. An alternative approach (not addressed by the ACSMM methodology) involves the team members themselves co-constructing a team mental model.

DEEP

The Dynamic Evaluation of Enhanced Problem-solving (DEEP) (Spector & Koszalka, 2004) methodology is based on the notion that learning in a complex domain implies becoming more like an expert (Ericsson & Smith, 1991) and more skilled in higher-order causal reasoning and problem solving (Grotzer & Perkins, 2000). A fundamental assumption is that it is possible to predict performance and assess progress of learning by examining a person's conceptualization of the problem space that person associates with a representative problem. Representations can then be compared with other representations using the analytic methods of MITO-CAR (Model Inspection Trace of Concepts and Relations Methodology) and SMD (Surface, Matching and Deep Structure Methodology), described later in this section. Moreover, these representations can be created by small groups, as well as individuals, and then analyzed using the ACSMM (Analysis Constructed Shared Mental Model) methodology or jMAP procedure, discussed earlier.

In DEEP, learners are presented with a short problem scenario and then asked to identify the most relevant factors influencing the problem situation. Next, learners are asked to describe each factor and indicate how it is related to other factors, again describing the nature of each identified relationship. These representations amount to annotated causal maps used in system dynamics to elicit expert models of complex, dynamic systems (i.e., intended to reflect systems thinking); although DEEP also allows for non-causal links (e.g., correlations, steps in a procedure, examples, and formulas). A sample DEEP representation is shown in Fig. 8 .

Two reflection questions are asked to complete the problem conceptualization: (1) What else would you need to know in order to actually resolve this problem situation? and, (2) What assumptions have you made in responding to this problem

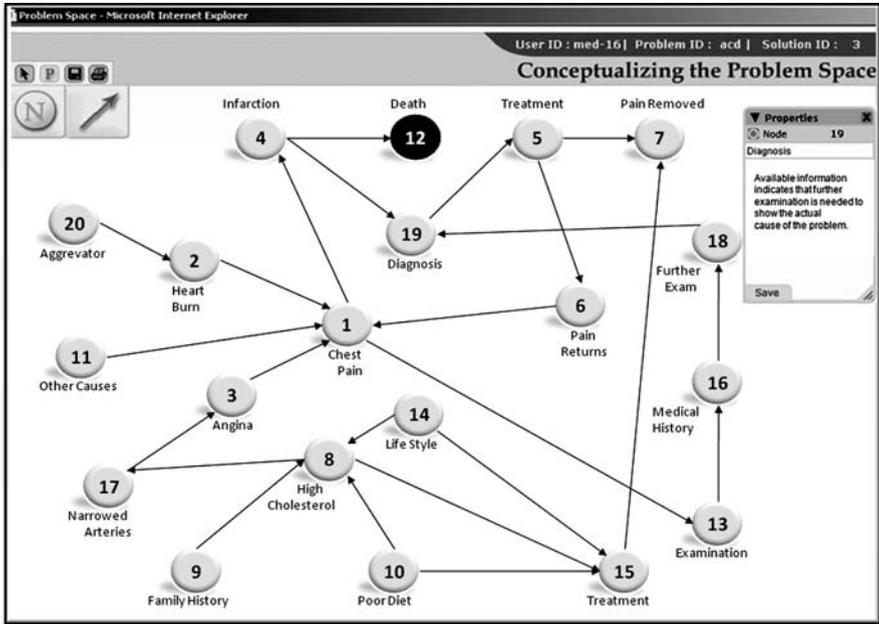


Fig. 8 A sample DEEP problem representation

situation? One strength of this methodology is that it is relatively simple to use and understand, minimizing the influence of the elicitation method on the representation.

The annotated causal representations in DEEP can be compared with prior representations and with those of experts, using some of the other tools described in this section (e.g., ACSMM, jMap, SMD, etc.). Three general levels of analysis can be applied to these representations: surface, structural, and semantic. 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 networks for analysis (e.g., causal maps or belief networks). 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).

SMD

The SMD (Surface, Matching, and Deep Structure) methodology (Ifenthaler, 2006, 2007) takes graphical representations in the form of causal diagrams (e.g., DEEP,

jMap) or association networks (e.g., MITOCAR) as inputs and provides *similarity metrics* for purposes of analysis of mental model development and progress of learning. The similarity metrics are derived from graph theory, and allow for comparisons among surface, matching, and deep structures.

Surface structure analysis is based on the sum of all propositions (node-link-node) in a particular representation. *Matching* structure is based on an analysis of the shortest path between the most distant nodes of the representation (Harary, 1974; Ifenthaler, Masduki, & Seel, 2008). *Deep* structure is based on an analysis of the semantic similarity of propositions (Tversky, 1977) between a domain-specific expert representation and a particular representation. The automated, on-the-fly analysis of SMD enables instructors to give learners immediate feedback during the learning process or while solving complex problems. The same metrics also provide researchers with powerful tools to analyze causal representations and association networks created using DEEP and MITOCAR, described next.

MITOCAR

The Model Inspection Trace of Concepts and Relations (MITOCAR) methodology (Pirnay-Dummer, 2006, 2007) is the final tool in our current HIMATT collection. And like the others, it is based on mental model theory (Seel, 1991). One of the unique features of MITOCAR is its ability to dig deeper into the semantics of various representations. Towards that end, MITOCAR operates in two phases—an assessment phase and an inferential phase.

During the assessment phase of MITOCAR, students usually respond in two rounds. In the first round they only provide a number of natural language phrases (usually sentences, and the program currently accepts English and German language as input) about a specific subject matter or problem area. The program's parser then extracts the most frequent concepts from the text corpus and creates an internal network of pairs of concepts from which a proximity vector is constructed. These data allow one to derive graphical models from text and compare them in several ways (Fruchterman & Reingold, 1991; Ganser & North, 1999; Maedche, Pekar, & Staab, 2002).

Like SMD, MITOCAR provides a variety of analysis measures based on graph theory and Tversky-Similarity (Tversky, 1977). For example, concept matching (surface level) compares the use of terms between different models, and structural matching introduces an algorithm that compares concepts maps in relation to (a) structure only (e.g., providing a testing ground for hypotheses about the structure of expertise), and (b) several density measures (Pirnay-Dummer, 2006).

In the second round of assessment, the students are asked to rate how close the concepts, output by MITOCAR, are to their current conceptualization (i.e., confidence in the validity of the MITOCAR 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). Finally, they are asked to rate the plausibility of their fellow group members' source phrases.

Table 1 Summary of six model-based assessment tools to support learning

Method	Data collection	Analysis	Data conversion	Comparison(s)
DAT	Annotated natural language from discussion or debate	Quantitative analysis—analysis is calculated using tools	Structural decomposition of event sequences and mean response scores (e.g., mean number of times X elicits Y)	Produces diagrams for visual comparison; produces raw scores for performing statistical comparisons
jMAP	Concept maps, causal maps, or belief networks	Quantitative analysis—analysis is calculated using tools	Structure decomposition into link strengths between causal factors and evidentiary strength	Superimposes maps of individual ($n=1$) and group of learners ($n = 2+$) over a specified target map
ACSM	Concept map	Qualitative with descriptive statistics. Analysis is done mostly by hand.	Structural decomposition into 3 categories (manual), structural re-composition into 1 representation (team map)	Unlimited comparisons, showing details relative to concepts.
DEEP	Annotated causal maps	Quantitative/qualitative—analysis is done mostly by hand	Structure decomposition into 3 categories (automatic)	Unlimited comparisons, showing details relative to concepts
SMD	Concept map or natural language	Quantitative—analysis is calculated using tools.	Structural decomposition into 3 categories (manual and semi-automatic)	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

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). The outputs of MITOCAR are graphical representations created from individual and group statements about a problem domain or situation. MITOCAR provides similarity measures, or a researcher can import MITOCAR outputs (graphs) into SMD for analysis.

Table 1 summarizes each of the six tools described in this chapter in relation to its (a) data collection requirements, (b) primary form of analysis, (c) data conversion procedure, and (d) permissible comparisons.

Conclusions

At any given time, students hold various beliefs about concepts, procedures, and other phenomena, which are all unobservable. Educators need valid, reliable, and efficient ways to externalize students' internal beliefs in order to accurately assess understanding and provide timely and meaningful assistance. Our chapter has presented a set of tools and technologies we are developing to support this assessment of individual and group mental models in different instructional contexts (e.g., problem-centered modules, discussion forums, informal settings).

In general, our tools aim to produce external representations (i.e., concept maps, causal models, and belief networks) that provide insight into internal constructs and processes (e.g., mental models and systems thinking). These external representations can provide useful information on how well students are conceptualizing some content area; and then teachers or automated instructional systems can adjust instructional supports appropriately. In addition to helping instructors and researchers, our tools can also help students to adjust their learning strategies and enhance their metacognitive skills if they are permitted to view, compare, and otherwise interact with their maps. Open or visible student models, as they're called, have been used to support knowledge awareness, student reflection, group formation, student modeling accuracy, and student learning (Bull & Pain, 1995; Kay, 1998, Hartley & Mitrovic, 2002; Zapata-Rivera & Greer, 2004). Finally, the tools can provide instructional designers with valuable information on which to base specific modifications to the structure and sequence of various learning activities.

As society becomes more complex, and new educational tools and technologies are being developed to keep pace with these changes, there is a growing need for assessment tools that can capture and measure mental models. Research in this area, however, must be based on sound theoretical foundations, and employ validated, scalable, and easy-to-use assessment tools. Moreover, these tools need to allow for measurement of change—one of the central problems of mental model research (Seel, 1999b). Towards that end, we have been designing and developing tools to allow for an assortment of comparisons between maps/models, of individuals and

groups, and at various points in time—to show not only where students began, but also their learning trajectories, similar to the benefits of motion pictures over still photographs.

References

- Alpert, S. R. (2003). Abstraction in concept map and coupled outline knowledge representation. *Journal of Interactive Learning Research, 14*(1), 31–49.
- Bakeman, R., & Gottman, J. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge: University Press.
- Bull, S., & Pain, H. (1995). Did I say what I think I said, and do you agree with me?: Inspecting and questioning the student model. *Proceedings of the Artificial Intelligence in Education (AAIE)*, Charlottesville, VA, pp. 501–508.
- Cannon-Bowers, J. A., & Salas, E. (Eds.). (1998). *Making decisions under stress: Implications for individual and team training*. Washington, DC: American Psychological Association.
- Dewey, J. (1915). *The school and society*. (2nd ed.). Chicago: University of Chicago Press.
- Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.
- Freeman, L. A., & Urbaczewski, A. (2001). Using concept to assess students' understanding of IS. *Journal of Information Systems Education, 12*, 3–8.
- 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), 1–5.
- Gentner, D., & Stevens, A. (1983). *Mental Models*. Hillsdale, NJ: Erlbaum.
- Greeno, J. G. (1989). A perspective on thinking. *American Psychologist, 44*(2), 134–141.
- Grotzer, T. A., & Perkins, D. N. (2000). *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.
- Guzzo, R. A., & Salas, E. (1995). *Team effectiveness and decision-making in organizations*. San Francisco, CA: Jossey-Bass, Inc.
- Hackman, R. A. (1990). *Groups that work (and those that don't): Creating conditions for effective team work*. San Francisco, CA: Jossey-Bass.
- Harary, F. (1974). *Graphentheorie*. München: Oldenbourg.
- Hartley, D., & Mitrovic, A. (2002). Supporting learning by opening the student model. *Proceedings of ITS 2002*, pp. 453–462.
- Herl, H. E., O'Neil, H. F., Jr., Chung, G. L. W. K., Bianchi, C., Wang, S., Mayer, R., et al. (1999). *Final report for validation of problem solving measures* (CSE Technical Report 501). Los Angeles: CRESST.
- Ifenthaler, D. (2006). *Diagnosis of the learning-dependent progression of mental models. Development of the SMD-Technology as a methodology for assessing individual models on relational, structural and semantic levels*. Freiburg: Universitäts-Dissertation.
- Ifenthaler, D. (2007). *Relational, structural, and semantic analysis of graphical representations and concept maps*. Paper presented at the Annual Convention of the AECT, Anaheim, CA.
- Ifenthaler, D., Masduki, I., & Seel, N. M. (2008). *Tracking the development of cognitive structures over time*. Paper presented at the AREA 2008, New York.
- 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.
- Jeong, A. (2005). A guide to analyzing message-response sequences and group interaction patterns in computer-mediated communication. *Distance Education, 26*(3), 367–383.

- Jeong, A. (2007). The effects of intellectual openness and gender on critical thinking processes in computer-supported collaborative argumentation. *Journal of Distance Education*, 22(1), 1–18.
- Jeong, A. (2008). *jMap*. Retrieved May 5, 2008, from <http://garnet.fsu.edu/~ajeong>
- Jeong, J. C. (2004). *Discussion Analysis Tool (DAT)*. Retrieved March 4, 2008, from <http://garnet.fsu.edu/~ajeong/DAT>
- Johnson, T. J., & O'Connor, D. L. (2008). Measuring team shared understanding using the analysis-constructed shared mental model methodology. *Performance Improvement Quarterly*, 21, 113–134.
- Kay, J. (1998). *A scrutable user modelling shell for user-adapted interaction*. Ph.D. Thesis, Basser Department of Computer Science, University of Sydney, Sydney, Australia.
- Klimoski, R., & Mohammed, S. (1994). Team mental model – Construct or metaphor. *Journal of Management*, 20(2), 403–437.
- Maedche, A., Pekar, V., & Staab, S. (2002). Ontology learning part one – On discovering taxonomic relations from the web. *Proceedings of the Web Intelligence Conference*, pp. 301–322, Springer.
- 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(2), 273–283.
- 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.
- O'Neil, H. F., Wang, S., Chung, G., & Herl, H. E. (2000). Assessment of teamwork skills using computer-based teamwork simulations. In H. F. O'Neil & D. H. Andrews (Eds.), *Aircrew training and assessment* (pp. 244–276). Mahwah, New Jersey: Lawrence Erlbaum.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufman Publishers.
- Pirnay-Dummer, P. (2006). *Expertise und Modellbildung: MITOCAR*. Freiburg: Universitäts-Dissertation.
- Pirnay-Dummer, P. (2007). *Model inspection trace of concepts and relations. A heuristic approach to language-oriented model assessment*. Paper presented at the AERA 2007, Division C, TICL SIG, April 2007, Chicago.
- 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.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33, 569–600.
- Salas, E., & Cannon-Bowers, J. A. (2000). The anatomy of team training. In S. Tobias & J. D. Fletcher (Eds.), *Training & retraining: A handbook for business, industry, government, and the military* (pp. 312–335). New York: Macmillan Reference.
- Seel, N. M. (1991). *Weltwissen und mentale Modelle [World knowledge and mental models]*. Göttingen, Germany: Hogref.
- Seel, N. M. (1999a). Semiotics and structural learning theory. *Journal of Structural Learning and Intelligent Systems*, 14(1), 11–28.
- Seel, N. M. (1999b). Educational diagnosis of mental models. Assessment problems and technology-based solutions. *Journal of Structural Learning and Intelligent Systems*, 14(2), 153–185.
- Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning*, 1(1), 59–85.
- Seel, N. M., & Winn, W. D. (1997). Research on media and learning: Distributed cognition and semiotics. In R. D. Tennyson, F. Schott, S. Dijkstra, & N. M. Seel (Eds.), *Instructional design international perspectives: Vol. 1. Theories and models of instructional design* (pp. 293–326). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Shute, V. J., Jeong, A., & Zapata-Rivera, D. (2008). *Assessing mental models and discourse patterns with evidence-based flexible belief networks*. Paper presented at the American Educational

- Research Association conference for the Technology, Instructional, Cognition and Learning (TICL) Symposia, New York, NY.
- Shute, V. J., & Zapata-Rivera, D. (2008). Using an evidence-based approach to assess mental models. In D. Ifenthaler, P. Pirnay-Dummer, & J. M. Spector (Eds.), *Understanding models for learning and instruction: Essays in honor of Norbert M. Seel*. New York: Springer.
- Spector, J. M., Dennen, V. P., & Koszalka, T. (2006). Causal maps, mental models and assessing acquisition of expertise. *Technology, Instruction, Cognition and Learning*, 3, 167–183.
- Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains (Final report the National Science Foundation Evaluative Research and Evaluation Capacity Building)*. Syracuse, NY: Syracuse University.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352.
- Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (C. K. Ogden, Trans.). London: Routledge and Kegan Paul.
- Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). London: Blackwell.
- Zapata-Rivera, D., & Greer, J. E. (2004). Interacting with inspectable Bayesian models. *International Journal of Artificial Intelligence in Education*, 14, 127–163.
- Zelev, E. C. (2004). Improving the usefulness of concept maps as a research tool for science education. *International Journal of Science Education*, 26(9), 1043–1064.