DEVELOPMENT ARTICLE

Evaluating mental models in mathematics: a comparison of methods

Aytac Gogus

Abstract Cognitive scientists investigate mental models (how humans organize and structure knowledge in their minds) so as to understand human understanding of and interactions with the world. Cognitive and mental model research is concerned with internal conceptual systems that are not easily or directly observable. The goal of this research was to investigate the use of Evaluation of Mental Models (EMM) to assess the mental models of individuals and groups in solving complex problems and to compare novices and experts models as bases for providing feedback to learners. This study tested a qualified web-based assessment tool kit, Highly Interactive Model-based Assessment Tools and Technologies (HIMATT), in an as yet untested domain—mathematics. In this study, university students and their mathematics instructors used two tools in HIMATT, Dynamic Evaluation of Enhanced Problem Solving (DEEP) and Text-Model Inspection Trace of Concepts and Relations (T-MITOCAR). The research questions include: Do novice participants exhibit common patterns of thoughts when they conceptualize complex mathematical problems? Do novices conceptualize complex mathematical problems differently from experts? What differences in DEEP and T-MITOCAR patterns and responses exist according to the measures of HIMATT? Findings suggest that EMM and HIMATT could effectively support formative assessment in a complex mathematical domain. Finally, this study confirms a common assumption of cognitive scientists that the tool being used could affect the tool user's understanding of the problem being solved. In this case, while DEEP and T-MITOCAR led to somewhat different expert models, both tools prove useful in support of formative assessment.

Keywords Mental models · Formative assessment · Complex learning · HIMATT · DEEP · T-MITOCAR

A. Gogus (✉)
Sabancı University, CIAD, Orta Mahalle, Tuzla, Istanbul 34956, Turkey
e-mail: agogus@sabanciuniv.edu

Introduction

Mental models are internal constructs that comprise an individual's conceptual systems and enable him or her to make sense of experience. Cognitive and mental model research is concerned with internal conceptual systems that are not easily or directly observable; researchers can only learn about internal conceptual systems by interpreting that which is more easily and directly observed—namely, individuals' communications about and representations of their own knowledge. External representations of mental models in the form of text (written or spoken), diagrams, physical models and such enable researchers to understand human perceptions of and interactions with the world so as to explain how humans organize, perceive and structure knowledge in their mind (e.g. Alexander 1963; Gentner and Stevens 1983; Gogus 2012a; Johnson-Laird 1983; Seel 1999). This study investigates the use of Evaluation of Mental Models (EMM) methodology that includes a specific approach for collecting, analyzing and interpreting data and an aggregation technique in terms of using three different tools of Highly Interactive Model-based Assessment Tools and Technologies (HIMATT): (1) Dynamic Evaluation of Enhanced Problem-solving (DEEP) (Spector and Koszalka 2004) using concept mapping (2) Text- Model Inspection Trace of Concepts and Relations (T-MITOCAR) (Pirnay-Dummer 2007) employing text form, and (3) surface structure, matching relations, deep semantics (SMD) (Ifenthaler 2006; Ifenthaler et al. 2007) to capture external representations of mental models of individuals who are solving complex mathematical problems and compares these models with each other.

Previous studies have conducted research to create reliable and valid methods to investigate conceptual representations of individuals and groups (e.g. Ifenthaler 2010; Johnson et al. 2006 as cited in Gogus 2009; Gogus and Gogus 2009; Gogus 2012a). As mental models are internal entities that have dynamic structures and subjective explanations, it is very difficult to determine the details of mental models and identify how they affect learning (Alexander 1963; Ariely 2009; Gentner and Stevens 1983; Gogus 2009, 2012a; Gogus and Gogus 2009; Ifenthaler 2010; Johnson-Laird 1983; Lakoff and Johnson 1981; Payne 1991; Polk and Seifert 2002; Seel 2003; Taleb 2008; Young 2008). Mental models represent constructs developed in order to understand human learning and problem solving processes. Seel (1991) emphasizes that constructing a mental model involves both assimilation (adding to existing mental model structures) and accommodation (reorganizing mental model structures and/or creating new mental models). Gentner and Stevens (1983) propose that there are three key dimensions of mental model research: (1) the domain studied (2) the theoretical approach involved, and (3) the methodology. There are relationships among these three dimensions, although the first is somewhat independent of the latter two which are more closely interrelated. Spector and Koszalka (2004) proceed on the assumption that the theoretical approach and methodology could be developed independently of a problem-solving domain and then applied in multiple domains. Their research applied a particular theoretical approach with regard to expertise development and an associated methodology for representing and assessing mental models in three domains (biology, engineering design and medical diagnosis) with the specific purpose of reliably distinguishing less experienced problem conceptualizations from those generated by experts. Their approach was successful in all cases that they examined in those three domains (15 experts and 60 novices).

Since mental models are internal representations of human conception on the problem, the structure of the problem and its presentation serve as important bases for the development of mental models on the problem. Therefore, this study chooses complex mathematical problems. To further elaborate the close connection between the theoretical

approach and methodology, this study considers the case of complex learning. One theory of learning in complex domains suggests that expertise develops over time and through sustained and systematic engagement with increasingly challenging problems (Ericsson and Smith 1991; Milrad et al. 2003). This view of expertise posits the progressive development of mental models. A related methodology, then, is to capture or represent in some way expert models and use them as reference points to determine how less experienced individuals think about a complex problem (Ifenthaler 2010; Spector and Koszalka 2004).

This research approach also has implications for the design of instruction and teaching. Developing assessment tools on the basis of a theoretical approach and an associated methodology is beneficial in assessing complex problem-solving skills and improving the quality of the most suitable instructional methods, course materials and learning activities (Andrews and Halford 2002; Gogus et al. 2009; Gogus 2012a; Dabbagh et al. 2000; Ericsson and Smith 1991; Herl et al. 1999; Ifenthaler 2010; Ifenthaler et al. 2009; Jacobson 2000; Jonassen 1997; Pirnay-Dummer et al. 2010). The theoretical approach and methodology developed by Spector and Koszalka (2004) involve an annotated concept map or influence diagram to elicit and represent a problem solver's mental model.

Concept mapping has been identified as a promising strategy to support assessment of learning in complex problem solving domains (Fisher 2000; Herl et al. 1999; Gogus and Gogus 2009; Gogus et al. 2009; Liu and Hichey 1996; McClure et al. 1999; Polk and Seifert 2002; Ruiz-Primo and Shavelson 1997; Taricani and Clariana 2006). Based on prior research and the success of recent efforts (Ifenthaler 2010; Ifenthaler and Seel 2005; Pirnay-Dummer 2007; Spector and Koszalka 2004), concept mapping tools are embedded in an integrated set of tools called HIMATT (Pirnay-Dummer et al. 2008, 2010).

The study reported here adopts the theoretical development of expertise characterized above and uses an associated methodology, EMM along with HIMATT to evaluate the mental models of individuals as they solve complex mathematical problems. EMM involves analyzing and comparing individual mental models and adds the construct of shared understanding and a method for capturing the common models of groups. Shared understanding means that a shared mental model can be created from common representations and trajectories of individual meaning within a group. This study assesses mental models using extensions to the previous concept mapping methods while participants are engaged in solving complex mathematical problems. The concept maps are drawn by analyzing key concepts and their relationships as the previous research on HIMATT suggests. This research extends prior studies into a new subject domain and focuses on higher-order thinking and problem solving skills for complex mathematical problems. These problems typically involve many interrelated factors without an obvious or simple algorithmic solution (Gogus and Gogus 2009; Gogus et al. 2009; Ifenthaler 2010; Spector et al. 2005; Spector and Koszalka 2004).

This research considers the process of learning as gaining expertise in problem solving. Therefore, experts' mental models are compared with those of students. The research questions include: do expert participants and novice participants exhibit common patterns of thoughts when they conceptualize complex mathematical problems? Do novices conceptualize complex mathematical problems differently than do experts? What differences in DEEP and T-MITOCAR patterns and responses exist according to the measures of HIMATT? Patterns mean a model that includes common concepts and links that are used in problem conceptualization. This study has aimed primarily at determining (a) what participants conceptualize when solving complex mathematical problems, and (b) whether there exists recognizable patterns in expert and student conceptualizations that can

facilitate the development of expertise (Gogus 2009, 2012a). In addition, differences between two different forms of externalizing mental models were examined to explore the notion of the impact of a tool on the conceptualization process.

Research methodology

Data collection tools

HIMATT, a set of Web-based assessment tools to measure and assess mental models of individuals and groups (Pirnay-Dummer et al. 2008, 2010) was used for the analysis of problem conceptualizations. HIMATT consists of three integrated tools: (1) DEEP (Spector and Koszalka 2004), (2) T-MITOCAR (Pirnay-Dummer 2007), and (3) SMD (Ifenthaler 2006). These three integrated tools of HIMATT have been used in more than a dozen studies but have not been previously tested in the area of mathematics and HIMATT was used in this study for the first time in the mathematics domain in 2008 (Gogus 2009, 2012a).

DEEP was developed in a project funded by the U.S. National Science Foundation entitled "Enhanced Evaluation of Learning in Complex Domains" which demonstrated the validity of the method (Spector and Koszalka 2004). In DEEP, annotated causal concept maps are used to elicit a conceptualization of how an individual (or small group of persons) is thinking about a problem situation. MITOCAR was found to visualize knowledge from groups of experts by using natural language expressions and several subsequent tests which the experts worked on individually (Pirnay-Dummer 2007). The whole process of knowledge elicitation methodology of MITOCAR was homogeneous, highly reliable and valid (Pirnay-Dummer 2007). Based on MITOCAR, Text-MITOCAR (T-MITOCAR) was designed to visualize knowledge structures on the basis of written text (Pirnay-Dummer 2007). T-MITOCAR relies on the structures and meanings of the words and uses the associative features of text to represent knowledge from text sources in a heuristic process (Pirnay-Dummer et al. 2008, 2010). The SMD Technology was developed to meet the demand for valid and reliable analytical instruments for educational research (Ifenthaler 2006). The reliability and validity of the computer-based and automated SMD Technology was tested in three experimental studies with 106 participants by Ifenthaler (2006). The SMD Technology measures relational, structural and semantic levels of graphical representations and concept maps (Ifenthaler 2006). The development of SMD and T-MITO-CAR were mainly funded by ParaDocks Omnimedia (R), and also in minor parts by the University of Freiburg. HIMATT had funding from multiple sources, including a Social Science Program Enhancement Grant from Florida State University managed by the International Center for Learning, Education and Performance.

T-MITOCAR and DEEP are compatible with a theoretical perspective but take different methodological approaches. Whereas DEEP supports the creation of an annotated concept map or influence diagram (Spector and Koszalka 2004), T-MITOCAR generates an association network based on text input. In that sense, T-MITOCAR is more closely aligned with the talk-aloud protocol method that was mentioned earlier. There are minor differences between the concept maps created using DEEP and the association networks created by T-MITOCAR. More specifically, the links in DEEP are directional while they are not directional in T-MITOCAR. Nonetheless, the same analytical methods (based on the SMD component of HIMATT) are used for analysis and for comparing two concept maps or two association networks. What have been left unanalyzed are the differences in

how T-MITOCAR and DEEP representations might be useful for predicting relative levels of expertise and as a basis for formative feedback to learners. Additionally, the impact of the tool being used to externalize mental models on those externalizations has not been investigated previously.

HIMATT have been used successfully in many research areas, including: assessments of performance and understanding of complex problems for teacher technology integration (McKeown 2008); acquiring instructional design expertise (Kim 2008); and diagnosis of stage-sequential learning (Kim 2012).

Research design

This study uses HIMATT in the mathematics domain for the first time. In addition, this study explores differences in DEEP and T-MITOCAR representations. Figure 1 shows the general approaches to using these two tools.

This mental model research design (see Fig. 1) uses both DEEP and T-MITOCAR to investigate how an internal representation can be re-represented with these two different tools. DEEP uses annotated concept maps while T-MITOCAR is a natural language capture of the thinking and mental models of the problem solver as she/he responds to a particular problem scenario. Since T-MITOCAR allows collecting the direct response to the problem as a text document, it is assumed to closely reflect the mental model. Using DEEP is good for gaining a sense of the entire problem space and for visualization whereas T-MITOCAR lends itself to natural language elicitation and is appropriate for general use and for those who might not find the visual concept map approach strange. This research intends to benefit from the advantages of both DEEP and T-MITOCAR and also explores differences in two different representations.
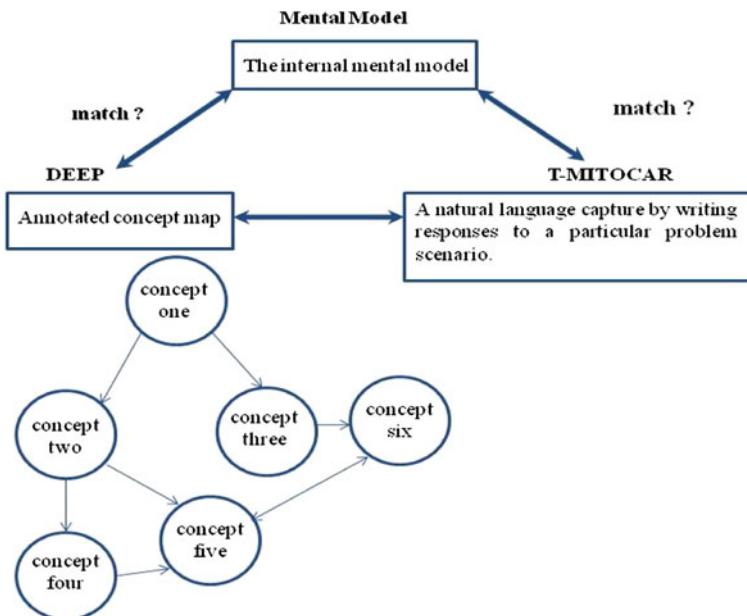
Fig. 1 The research design

EMM methodology: a research procedure and strategy

In EMM methodology, experts and novices develop annotated concept maps as conceptual representations of their understanding of a complex problem. This method constitutes a specific form of the talk-aloud protocol developed by Ericson and Simon (1993). It might be more accurately described as a think and externalize protocol that lends itself to a visual representation as well as a representation that can support formative feedback to learners. Within EMM, HIMATT tools are used to analyze the underlying cognitive structures in the concept maps and to get predictable patterns of structural understanding that can both inform and facilitate problem-solving processes (Markham et al. 1994). EMM methodology consists of the following stages:

1. Decide problem scenarios.
2. Organize an experimental group and ask them to describe what they think about the solution of the problem they have individually presented in graphic and text formats.
3. Create a reference model of experts.
4. Use the reference model to look for a common model between experts and compare this model with those of the novices and observe how to give feedback while comparing two different models with each other.

In this study, university students and their mathematics instructors use two tools (DEEP and T-MITOCAR) in HIMATT during data collection. This study also uses SMD, the feedback function of HIMATT, during data analysis. By taking into account that effective model-based feedback is composed of externalized representations (re-representations) of mental models such as a causal model, a concept map, written text and aims at the improvement of expertise and expert performance, the data will be collected with a concept map tool (DEEP) and a written text tool (T-MITOCAR) and will be analyzed by using an automatic model-based feedback tool (SMD). HIMATT has a model-based feedback function to generate a reference model (e.g. an expert's model), a participant model (e.g. a learner's model), and a cutaway model which is generated by the comparison of a participant's model with a reference model to show similarities and differences between two models (Ifenthaler 2010).
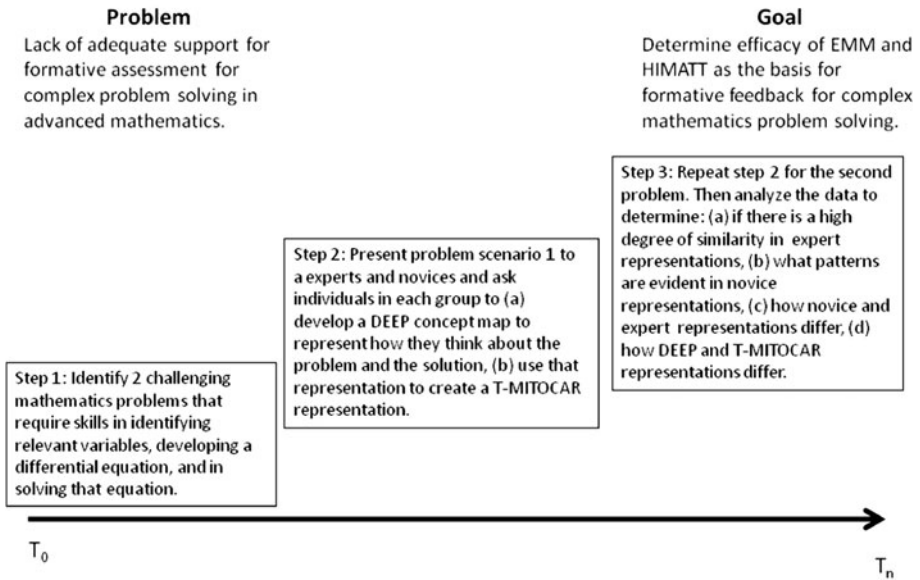
Research questions

The purpose of this research study is to assess the mental models of individuals and groups in solving complex and challenging mathematics problems and to compare novices and experts models as a basis for providing feedback to learners. The primary research questions include:

1. Do novice participants exhibit common patterns of thoughts when conceptualizing complex mathematical problems?
2. Do novices conceptualize complex mathematical problems differently than do experts?
3. What differences in DEEP and T-MITOCAR patterns and responses exist according to the measures of HIMATT?

Figure 2 provides a general overview of the research approach.

Firstly, this study aims to test HIMATT in an as yet untested domain—mathematics—to see if and to what extent annotated concept maps can distinguish expert responses from those of less experienced persons.

Research Overview

Problem
Lack of adequate support for formative assessment for complex problem solving in advanced mathematics.

Goal
Determine efficacy of EMM and HIMATT as the basis for formative feedback for complex mathematics problem solving.

Step 3: Repeat step 2 for the second problem. Then analyze the data to determine: (a) if there is a high degree of similarity in expert representations, (b) what patterns are evident in novice representations, (c) how novice and expert representations differ, (d) how DEEP and T-MITOCAR representations differ.

Step 2: Present problem scenario 1 to a experts and novices and ask individuals in each group to (a) develop a DEEP concept map to represent how they think about the problem and the solution, (b) use that representation to create a T-MITOCAR representation.

Step 1: Identify 2 challenging mathematics problems that require skills in identifying relevant variables, developing a differential equation, and in solving that equation.

T_0 T_n

Fig. 2 Research overview

Secondly, this study aims to investigate if the tool used to elicit the problem conceptualization affects what is elicited by comparing and contrasting DEEP and T-MITOCAR since some researchers (Pirnay-Dummer et al. 2008, 2010) argue that natural language elicitation is closer to the mental model and similar to the established technique of talk-aloud protocol analysis; however, other researchers (Spector and Koszalka 2004) suggest that what is important for inexperienced learners is to gain an early appreciation for the entire system. Since a holistic overview of all factors influences the problem, the DEEP methodology is better suited for problem conceptualization. That is the reason why this research design first uses DEEP and then T-MITOCAR.

Thirdly, this study aims to investigate whether DEEP or T-MITOCAR, or both, can be used to create an expert reference model that can be used for formative feedback in a meaningful way to personalize and individualize feedback and also serve to promote expert-like understanding.

Participants

The EMM study was conducted at Sabanci University in Turkey in 2009–2010 Spring Semester with the second year undergraduate students of the course "Differential Equations". The participants included 22 college students and 4 experienced academicians in mathematics who taught "Differential Equations" at Sabanci University. Participation was voluntary, and participants were compensated for their participation. The study was approved by the university's institutional review board (IRB) and participant identifying information was kept confidential in compliance with IRB requirements. Demographic information of the participants is summarized in Table 1.

Table 1 Demographic information		Experts (4)	Students (22)
Gender			
	Male	3 (75 %)	21 (95 %)
	Female	1 (25 %)	1 (5 %)
Age			
	15–19	0 (0 %)	4 (18 %)
	20–24	0 (0 %)	18 (82 %)
	25–49	2 (50 %)	0 (0 %)
	50–	2 (50 %)	0 (0 %)
Work experience (year)			
	$0 \leq$ year < 1	0 (0 %)	22 (100 %)
	$1 \leq$ year < 5	0 (0 %)	0 (0 %)
	$5 \leq$ year < 10	2 (50 %)	0 (0 %)
	Over 10	2 (50 %)	0 (0 %)
Faculty			
	FASS	0 (0 %)	4 (18 %)
	FENS	4 (100 %)	18 (81 %)

There were three male and one female experts, two of whom were over 50 years old and had field experience of more than 20 years, whereas two were under 35 years old and had 5 and 10 years of field experience, respectively. One of the experts was a Bulgarian mathematician and the rest were Turkish. These four experts were from the Faculty of Engineering and Natural Sciences (FENS) since faculty members of mathematics program work at FENS, not at the Faculty of Arts and Social Science (FASS). There were 22 student participants with a mean age of 21; 4 students were from FASS and 18 students were from FENS. Among the students, 21 were male and one was female. These 22 college students from the course "Differential Equations" voluntarily participated in this study. All these students took the prerequisite courses as a requirement of the curriculum at Sabanci University even though they were registered to a different faculty. All student participants were Turkish. Males were certainly overrepresented in the sample.

Data collection

The data was collected from participants (four instructors and 22 university students) individually and one at a time. Participants responded to two representative problem scenarios developed or chosen by domain experts. The experiments took approximately one and half an hour for the first problem and around 1 h for the second problem since the participants had gotten used to the process of data collection when the second problem was provided.

The procedure (see Fig. 2) was that participants were first provided a brief introduction to the study. Then they were presented with the first problem scenario (see Appendix 1) and asked to individually develop a DEEP representation in the form of an annotated concept map of the key concepts (variables) involved and their relationships. Participants were then asked to use T-MITOCAR to describe the problem in terms of key concepts and relationships. EMM in HIMATT involves providing the participants with a complex problem and eliciting their thoughts on how they would approach the development of a solution—this is called conceptualizing the problem space (Gogus et al. 2009). The

participants are not asked to solve the problem; rather, they are asked to conceptualize the problem through four tasks: (a) identify and briefly describe the key factors influencing the situation (b) identify and describe how these factors are interconnected (c) indicate additional information that would be required to actually resolve the problem situation, and (d) identify assumptions made in response to the previous three questions.

Figure 1 represents the general research approach. The reason why DEEP was used firstly was based on the assumption that it would facilitate a more holistic or global way of thinking about all aspects of the problem since DEEP is based on the use of causal influence diagrams used by system dynamicists when creating the initial of a large, complex system (Spector and Koszalka 2004). This assumption has not been empirically tested and is recommended for investigation in future studies. After a short break, participants were then asked to repeat this process for the second problem scenario. For each problem scenario, an effort was made to create an expert reference model based on the experts' representations in DEEP and again in T-MITOCAR. All data were then analyzed to address the four research questions.

Data analysis and outcomes of measures

There were three main categories of analysis: surface features (e.g. the number of nodes and links, average number of words to describe each node and link), structural features (e.g. key node clusters, connectedness of the concept maps measured by the percentage of orphan nodes lacking connection back to other nodes), and semantic features (e.g. did experts and novices say the same kinds of things about similar nodes). By considering these three main categories of analysis, the six measures developed in HIMATT were used to analyze the DEEP concept maps and the T-MITOCAR networks based on the written text (Ifenthaler 2009, 2010; Pirnay-Dummer 2007; Pirnay-Dummer 2010; Pirnay-Dummer and Ifenthaler 2010):

- *Surface matching* measure surface complexity of the graphs (Ifenthaler 2009, 2010).
- *Graphical matching* is an indicator for the range of conceptual knowledge (Ifenthaler 2009, 2010).
- *Concept matching* determines differences in language use between the models (Ifenthaler 2009; Pirnay-Dummer 2007).
- *Structural matching* compares the complete structures of two graphs (Ifenthaler 2009; Pirnay-Dummer 2007).
- *Gamma (node density and connectedness) matching* describes the quotient of terms per vertex within a graph (Ifenthaler 2009; Pirnay-Dummer 2007).
- *Propositional matching* compares only identical propositions. It is a good measure to quantify semantic similarity between two graphs (Ifenthaler 2009, 2010).

The measures were calculated automatically within seconds and then displayed as pairwise sets including the six core measures described above (see Fig. 3) and the researcher could download a spreadsheet containing all measures for further statistical analysis (Ifenthaler 2009, 2010; Pirnay-Dummer 2007). These six measures were used to analyze each concept map and determine similarities and differences between the concept maps of learners and the expert reference concept map by getting through the similarity measures that are described in the HIMATT manual (Ifenthaler 2009, 2010; Pirnay-Dummer 2007). Previous HIMATT studies have generally found the concept similarity measure and gamma to be the most useful for tracking expertise (Kim 2008, 2010; Lee 2008; McKeown 2008). Concept similarity (Pirnay-Dummer 2007) means similarities in language use between the models. Gamma, the density of vertices, (Pirnay-Dummer 2007)

measures the connectedness of nodes, 0 means no interconnections and 1 means maximum connects such as all connected. According to previous research (e.g. Spector and Koszalka 2004; Kim 2008), while novice representations show few interconnections, expert representations shows significant interconnectedness.

In addition to these measures of HIMATT, the *feedback function* of the SMD Technology was implemented in HIMATT (Ifenthaler 2009, 2010). When analyzing and comparing the models in this study, this feedback function was used in the form of the cutaway model-based feedback which are considered as graphical re-representations constructed from a set of concepts and *vertices (concepts or corners)* whose relationships are represented by *edges (links or lines)* (Ifenthaler 2010; Ifenthaler et al. 2009). The *feedback function* of the SMD Technology in HIMATT automatically generates a standardized reference re-representation (e.g. the expert's solution), participant re-representation (e.g. the learner's solution) and cutaway re-representation (Ifenthaler 2010). By comparing the participant re-representation (e.g. the learner's solution) with the reference re-representation (e.g. the expert's solution), the cutaway re-representation was created by HIMATT to give feedback to the learner according to the expert's solution (Ifenthaler 2010) after the experiment.

Findings from research questions

In the HIMATT subject environment, the participants create a concept map within the DEEP module by adding nodes and links to the concept map and annotating the nodes with additional information. The reference mental model of experts (see Fig. 4) was created using the DEEP tool in the HIMATT Subject Environment. This reference model is a common model that is created by the subject matter expert and co-researcher in this study, through comparing four experts' models drawn by the examination of the mental models of 4 experts (see Appendix 1). The solution of the problem is given in Appendix 2.

Do novice participants exhibit common patterns of thoughts when they conceptualize complex mathematical problems?

Since participants conceptualize their thoughts on complex mathematical problems in two ways: (1) using the DEEP tool to construct a concept map with written explanation (2) using the T-MITOCAR tool to express written texts, the research questions are answered

**Fig. 3** Compare function including all six HIMATT core measures
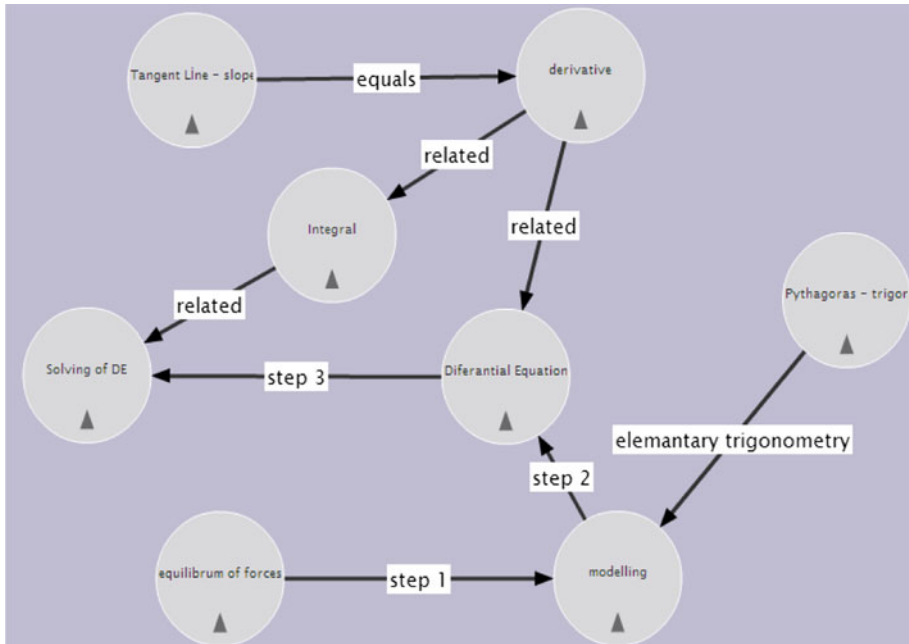
Fig. 4 The common representation of a mental model of experts created using DEEP

by using data from both the DEEP tool and the T-MITOCAR tool. When the mental model measurements of 22 novices were analyzed, there was a significant common mental model in both DEEP and T-MITOCAR tools. Figure 5 shows the common mental model of novices created by using DEEP. This model includes the most frequent used nodes and the links at the students' models that are analyzed by the subject matter expert. The participants constructed the concept maps by organizing information about the problem scenario and identifying the problem solution procedures. Therefore, the concept map in Fig. 5 shows the common concepts, facts, and procedures from the models of each participant. This model demonstrates that the students indicate main concepts and the basic strategy that is appropriate to reach the problem solution.

When the measurement was conducted through the DEEP tool, the results of a single sample t test showed that $t(21) = 10.99$, $p < .001$. When the measurement was conducted through T-MITOCAR tool, the results of a single sample t test showed that $t(21) = 12.32$, $p < .001$. The results indicated that novices had a common mental model in both the conceptual and the grammatical context. By performing single sample t tests in every area of HIMATT's six measures for the data from the DEEP tool, novices showed significant commonalities in the areas of surface measure $[t(21) = 18.66, p < .001]$, graphical matching $[t(21) = 19.2, p < .001]$, structural matching $[t(21) = 34.6, p < .001]$, and node density $[t(21) = 25, p < .001]$. It was determined from these results that novices had significant commonalities in every area except concept matching. The results are summarized in Table 2. There were no results for the proportional matching.

By performing single sample t test in every area of HIMATT's six measures for the data from the T-MITOCAR tool, commonality was found in novices' surface measure $[t(21) = 6.429, p < .001]$, graphical matching $[t(21) = 13.087, p < .001]$, and structural
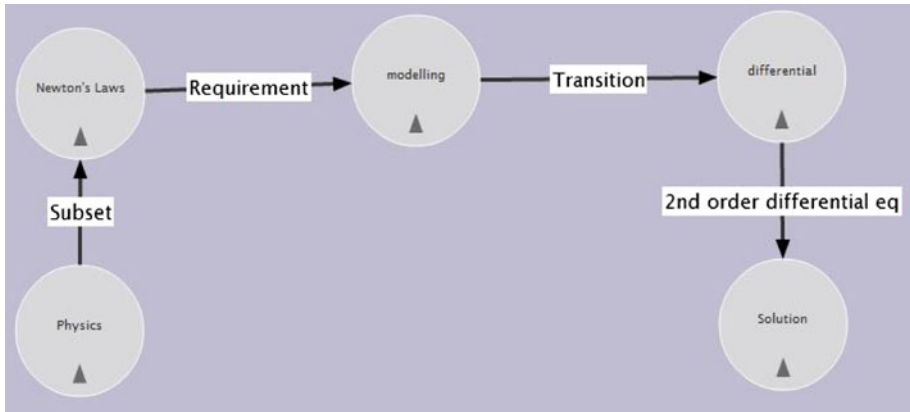
Fig. 5 The common representation of a mental model of novices created using DEEP

Table 2 Single sample t test results of novice participants using the DEEP tool

DEEP	t	df	Mean square difference	Significance level
Surface matching	18.66	21	.73	<.001
Graphical matching	19.2	21	.75	<.001
Conceptual matching	1.81	21	.02	.084
Structural matching	34.6	21	.85	<.001
Node density (Gamma)	25	21	.79	<.001

matching [$t(21) = 10.99$, $p < .001$], node density [$t(21) = 6.501$, $p < .001$] and proposition matching [$t(21) = 2.366$, $p < .05$]. It was observed that novices had commonalities in every measurement area. Results are summarized in Table 3.

Do novices conceptualize complex mathematical problems differently from experts?

When the DEEP models of experts and novices were compared qualitatively, there were some common nodes and explanations for relations. However, it can be seen that the experts' models (see Fig. 6) were more detailed than those of the novices (see Fig. 7). Appendix 3 shows two sample DEEP models from an expert participant and a novice participant from the two participants.

The *feedback function* of the SMD Technology in HIMATT allows viewing the experts' models (e.g. Fig. 6), the students' models (e.g. Fig. 7) and a cutaway model (see Fig. 8) by comparing the students' models with the experts' models in order to give feedback to the learner according to the expert's solution. In Fig. 8, circles indicate similar concepts and ellipses indicate no similarity and/or missing concepts.

When mental models of experts and novices were compared quantitatively using DEEP, no significant commonality was found. Using the six matching measurements of DEEP as the dependent variable and the expertise level as the independent variable, a one-way ANOVA indicated that there was no significant commonality between novices and experts; $F(1,24) = 2.89$, $p = .102$. These measurements are summarized in Table 4. There were no results for the proportional matching.

Table 3 Single sample *t* test results of novice participants using the T-MITOCAR tool
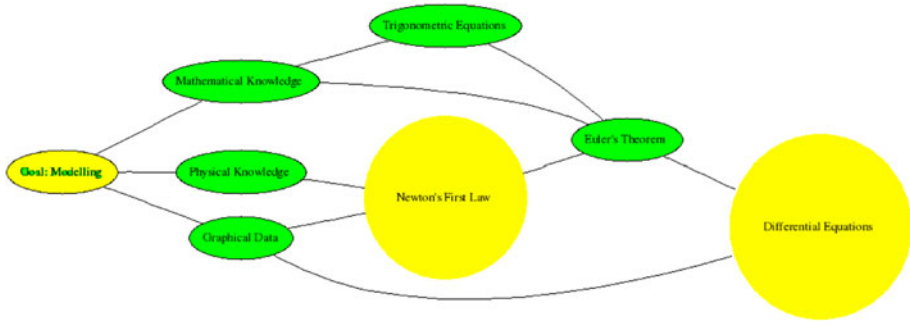
T-MITOCAR	*t*	*df*	Mean square difference	Significance level
Surface matching	6.43	21	.45	<.001
Graphical matching	13.09	21	.7	<.001
Conceptual matching	7.06	21	.13	.084
Structural matching	10.99	21	.42	<.001
Node density (Gamma)	6.5	21	.39	<.001
Proposition matching	2.37	21	.01	<.001
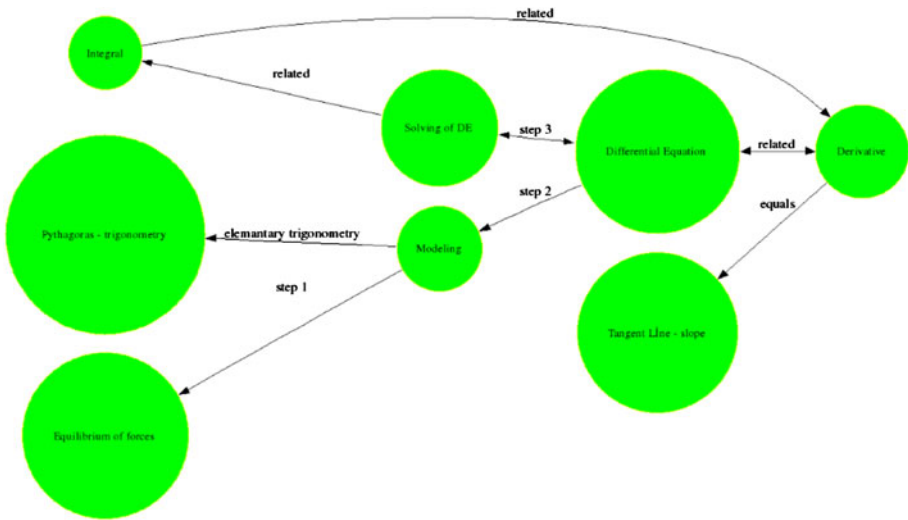
Fig. 6 The expert model derived from the DEEP tool
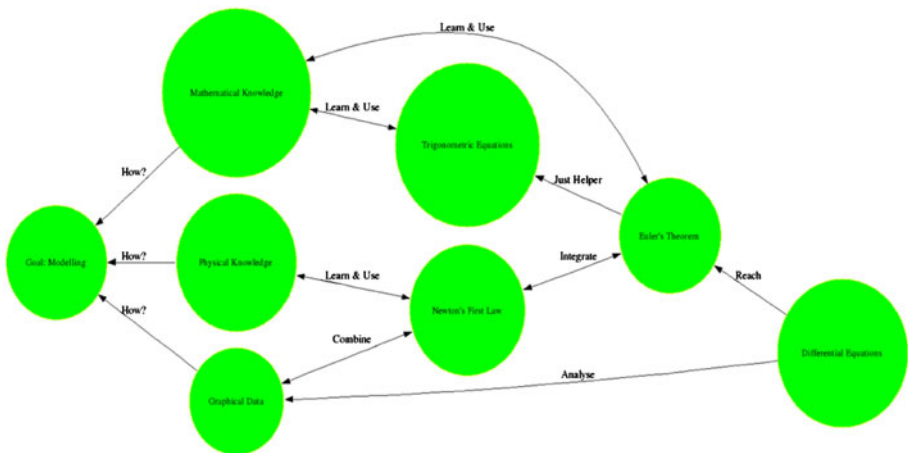
Fig. 7 The novice model derived from the DEEP tool

Fig. 8 The model display Cutaway data for comparison of the model in Fig. 7 with the model in Fig. 6

Table 4 A one-way ANOVA comparing novices and experts using DEEP

DEEP	*F*	Mean squares	Significance level
Surface matching	2.39	.084	.135
Graphical matching	1	.038	.327
Conceptual matching	.59	.001	.452
Structural matching	.36	.004	.556
Node density (Gamma)	.08	.002	.784

Similarly, when the models of experts and novices were compared qualitatively using DEEP, the experts' models (see Fig. 9) were more detailed and the nodes were more connected than the novices' models (see Fig. 10). The *feedback function* of the SMD Technology in HIMATT allows viewing the experts' model (a re-representations from the participant's T-MITOCAR data, e.g. Fig. 9), the students' models (a re-representations from the participant's T-MITOCAR data, e.g. Fig. 10) and a cutaway model (see Fig. 11) by comparing the students' models with the experts' models in order to give feedback to the learner according to the expert's solution. In Fig. 11, circles indicate similar concepts and ellipses indicate no similarity.

Using the six matching measurements of T-MITOCAR as the dependent variable and the expertise level as the independent variable, the one-way ANOVA indicated that there was no significant commonality between novices and experts; $F(1,24) = .004$, $p = .949$. When six of the measures were examined one by one, it was observed that there was no significant commonality in any area. Measurements conducted by T-MITOCAR are summarized in Table 5.

These results show that there is a difference between the way experts and novices conceptualize the complex mathematical problems.

What differences in DEEP and T-MITOCAR patterns and responses exist according to the measures of HIMATT?

HIMATT allows recording individual responses during data collection and comparing these responses with each other and/or with a reference model during data analysis to find
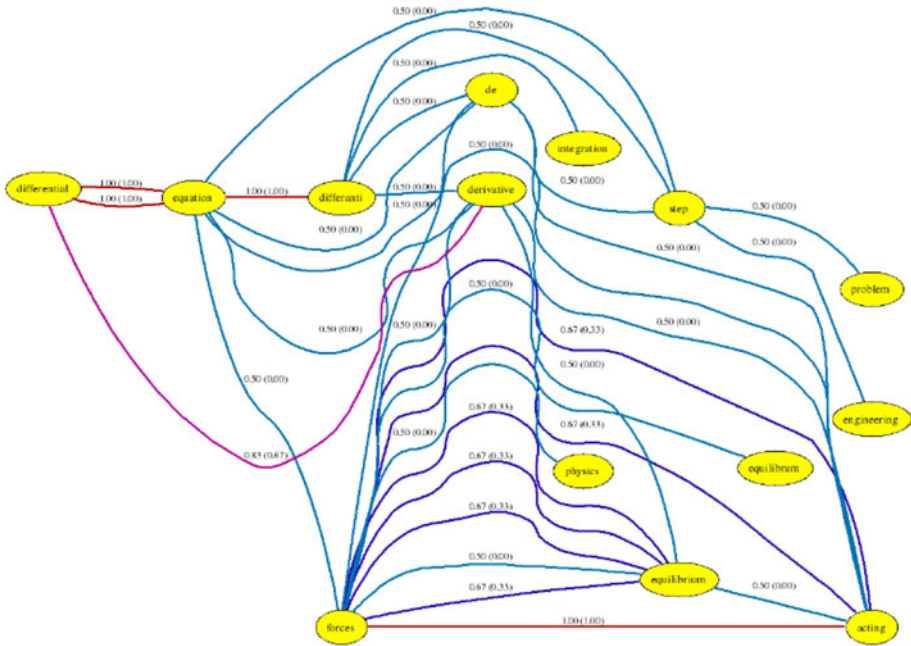
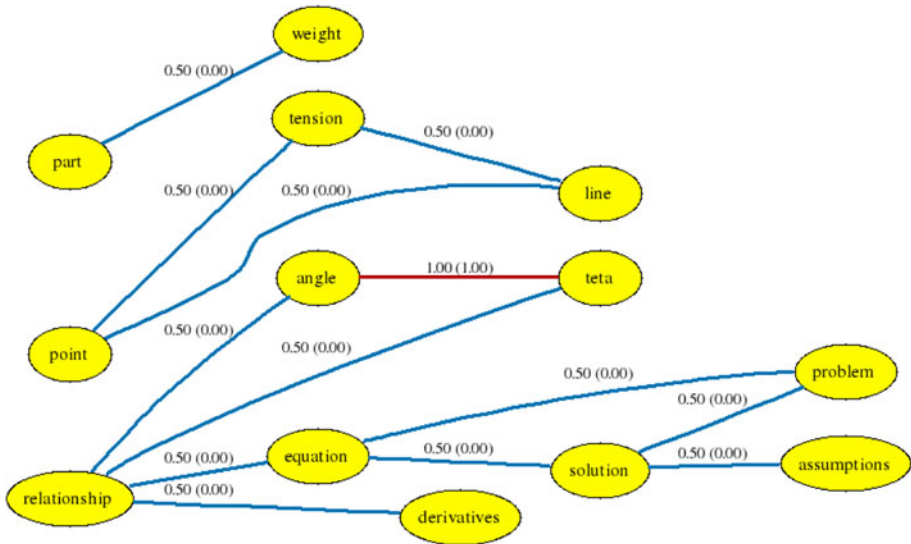**Fig. 9** The expert model derived from the T-MITOCAR tool

Fig. 10 The novice model derived from the T-MITOCAR tool

patterns that indicate similarities between individuals' responses. In addition to the three main research questions of the study, a comparison between DEEP and T-MITOCAR was conducted by performing repeated sample *t* test in the six measurement areas. According to
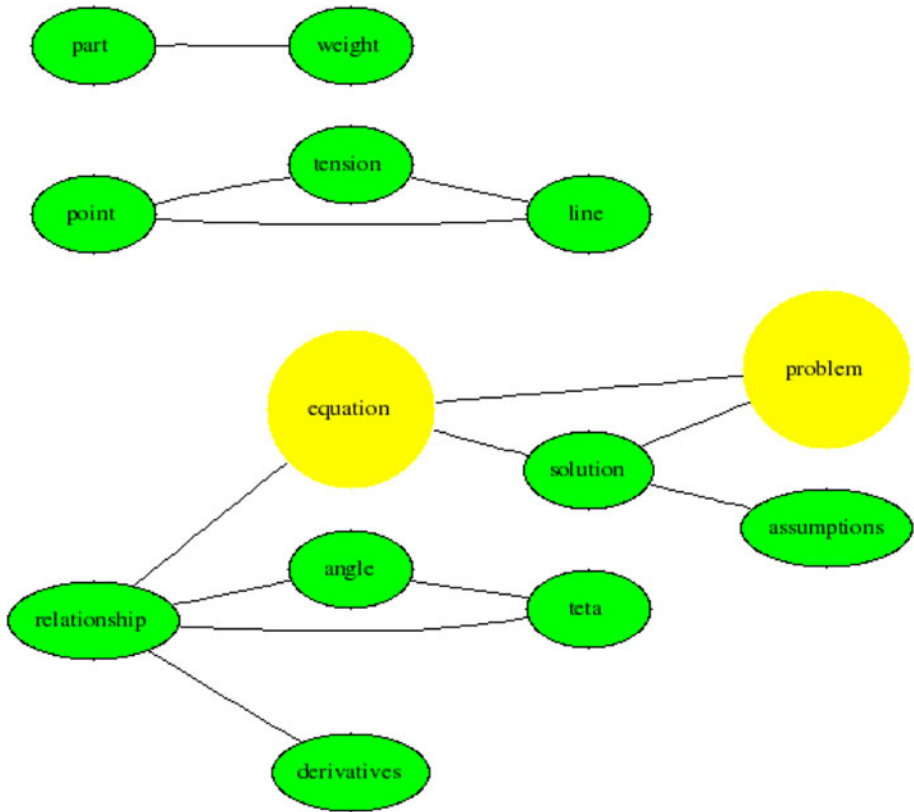
Fig. 11 The model display Cutaway data for comparison of the model in Fig. 10 with the model in Fig. 9

Table 5 A one way ANOVA comparing novices and experts using the T-MITOCAR

T-MITOCAR	F	Mean squares	Significance level
Surface matching	.02	.003	.88
Graphical matching	.81	.06	.378
Conceptual matching	.27	.002	.609
Structural matching	.2	.01	.658
Node density (Gamma)	.1	.01	.75
Proposition matching	.01	.00	.911

the findings, only the graphical matching measurement was similar in the results given by DEEP and T-MITOCAR; [$t(25) = .72, p = .476$]. In other areas, DEEP and T-MITOCAR had significantly different measurements; surface measure [$t(25) = 3.31, p < .05$], concept matching [$t(25) = −7.28, p < .001$], structural matching [$t(25) = 12.55, p < .001$], node

Table 6 Repeated sample t-test comparing DEEP and T-MITOCAR measurements

DEEP versus T-MITOCAR	t	df	Mean	Significance level
DEEP—T-MITOCAR Surface matching	3.31	25	.26	.003
DEEP—T-MITOCAR Graphical matching	.72	25	.05	.476
DEEP—T-MITOCAR Conceptual matching	−7.28	25	−.11	<.001
DEEP—T-MITOCAR Structural matching	12.55	25	.44	<.001
DEEP—T-MITOCAR Node density (Gamma)	5.25	25	.39	<.001
DEEP—T-MITOCAR Proposition matching	−2.61	25	−.01	.015

density [$t(25) = 3.31$, $p < .05$] and preposition matching [$t(25) = -2.61$, $p < .05$] as shown in Table 6.

Except for graphical matching, these results support the assumption that DEEP measures conceptual graphical representations and T-MITOCAR measures descriptive text representations.

Discussions and conclusions

This study tested HIMATT in an as yet untested domain—mathematics—to see if and to what extent annotated concept maps can distinguish expert responses from those of less experienced persons. The general HIMATT does hold up in mathematics even though math problems are well-structured in most cases but can also be complex and challenging, and seem amenable to treatment with annotated concept maps and the HIMATT approach.

This study primarily determined (a) how participants conceptualize complex mathematical problems and (b) whether there exists recognizable patterns in expert and student conceptualizations that can facilitate the development of expertise. The use of concept maps and text summaries as means of capturing students' mental models (external representations) were proposed as viable means to measure students' understanding and conceptualization of problem scenarios (Ifenthaler 2010; Johnson et al. 2006). The analysis and comparisons of the externalized representations consisted of six measures suggested by the HIMATT toolset (Pirnay-Dummer and Ifenthaler 2010). The automated quantitative analysis generated structural measures (*surface*, *graphical*, *gamma*, and *structural matching*) and semantic measures (*concept* and *propositional matching*) to obtain students' representations and compare the problem conceptualizations both graphically and textually (Ifenthaler 2010).

A common mental model was extracted by comparing the mental models of participants. The common mental model based on expert representations was used as a reference model to analyze individual models for purposes of formative feedback (e.g. finding expert nodes not included in a novice representation). The expert reference model was used to develop a relative measure of the level of expertise and estimate the level of learning. Except for concept matching, all of the five measurement areas included commonalities between the expert and the novice models. This finding is consistent with the notion that the lack of common concepts between an expert representation and a novice representation is a good indicator of relative levels of expertise (see Spector and Koszalka 2004). Differences in language use were found when the concepts in the models were compared. This

discrepancy occurs because in mathematics, it may be difficult to describe mathematical relationships in words rather than in formulas.

This study investigated if the tool used to elicit the problem conceptualization affects what is elicited by comparing and contrasting DEEP and T-MITOCAR. This study design asked participants firstly to use the DEEP tool and then T-MITOCAR tool. Findings show that the tool used here does seem to influence the conceptualization elicited although it is not clear which might be better used first since both seem to be useful in a practical way.

Findings of the research questions 1, 2 and 3 suggest that both DEEP and T-MITOCAR can be used to create an expert reference model that can be used for formative feedback in a meaningful way to personalize and individualize feedback and also serve to promote expert-like understanding. A practical question can be asked: is it better to support identifying gaps or to target formative feedback? Findings suggest that either can be used to support formative feedback, even though there are some differences between the two representations. Findings show that using DEEP is good for gaining a sense of the entire problem space in the case of visually-oriented learners whereas T-MITOCAR allows collecting the direct response to the problem as a text document. This research suggests benefiting from the advantages of both DEEP and T-MITOCAR. The potential for the tools, DEEP and T-MITOCAR, "to provide a learner with near real-time feedback comparing his or her problem conceptualization with that of an expert or another learner exists now and is of obvious value in promoting the development of productive mental models to solve complex problems" (Spector 2010, p. 11).

When the mental model measurement was conducted by using T-MITOCAR, commonality was found only in graphical matching. This finding may be due to the fact that T-MITOCAR creates a graph from text, although prior studies with T-MITOCAR have found concept measures to be useful along with the gamma measure. In any case, the gamma measure does provide a basis for determining a relative level of expertise and may help in identifying gaps in a novice representation. The results of this study suggest that experts have a common mental model in conceptual and graphical representations but not in grammatical or propositional similarity. Grammatical similarity is associated with the propositional similarity measure that is used in this study. However, the results of the propositional similarity suggest that the measurement by the propositional matching is useless in HIMATT since the requirements for a propositional match are too severe since it must be an exact match. The gamma measure is associated with connectedness but not graphical similarity while graphical similarity requires having the same nodes connected to the same nodes in two different representations done by pair wise comparison when there are common nodes (concepts) in two representations.

According to the previous mental model studies, a mental model progression starts with a model that contains the ideas that are most simple, representative and fundamental (van Merriënboer 1997). And then, subsequent models add complexity to a part of the former models and become their elaborations. In line with the descriptions of mental models by van Merriënboer (1997), mental models of experts are seen as extremely complex schemata in which the nodes are either concepts or principles and some types of relationships are used. However, students' mental models were simple mental models that included isolated concepts, principles, and relationship between concepts (van Merriënboer 1997).

While experts gave more importance to the causal relationships of key concepts, novices used a wider range of information. This difference is in line with the findings of Glaser (1996) and Perez et al. (1995) indicating that experts use well-established principles in their field of expertise while novices are usually limited to descriptions and surface explanations as they lack knowledge of general principles. Novices could not explain many

relationships between and among concepts due to gaps in their knowledge and unawareness of general principles. This characterization is in line with the findings of Shin et al. (2003), which have emphasized that subject knowledge is an effective factor in successfully solving complex structured problems. In this study, learners were presented with real life problems so that the patterns of their causal thinking could be evaluated (Jonassen 2000).

In addition, previous research emphasizes that concept maps and causal interaction diagrams have a critical role in the evaluation of complex problems (e.g., Jonassen et al. 1993; McKeown 2008; Seel 2004; Shute et al. 2000; Spector and Koszalka 2004). The use of concept maps is important because it reveals the external representations of the individual's thinking, knowledge and ability to solve a complex mathematical problem.

This evaluation study of *mental models,* using a title of the EMM methodology, requires analyzing and comparing individuals' mental models and capturing shared models of groups. There are two dimensions of this evaluation: (1) comparing individuals' *mental models* in one group and capturing shared models of groups (2) comparing *mental models* of two groups that have two various levels of expertise and analyzing these compared *mental models*. In further studies, one more dimension can be added to the EMM methodology and this amended methodology can be used in more detailed research studies related to the assessment of learning outcomes and performance evaluation. Thus, the EMM methodology can be used in three dimensions: (1) comparing individuals' *mental models* in one group and capturing shared models of groups (2) comparing *mental models* of two groups that have two various levels of expertise and analyzing these compared *mental models* in terms of learning outcomes and performance evaluation (3) analyzing *mental models* of individuals and groups to investigate the changes of progress over time.

A limitation of this study is having a relatively small subject group. It is important to work with larger samples when a more facile interface is eventually developed for the tools. Further important research studies should be conducted for a large-scale use of the EMM methodology. The EMM methodology can be used to provide real-time and dynamic feedback to learners in learning situations involving challenging and complex problems. Also, the use of the EMM methodology allows determining the impact on learning effectiveness and efficiency.

Appendix 1

The first problem scenario: the hanging chain (Boyce and DiPrima 2005)

Imagine a flexible steel chain, attached firmly at equal height at both ends, hanging under its own weight (see Fig. 12). What shape will it describe as it hangs? This is a classical
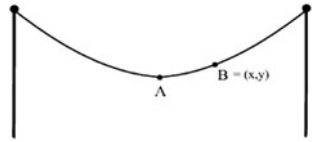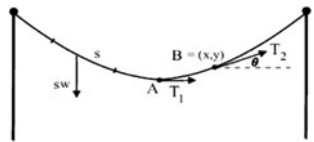
Fig. 12

Fig. 13

Fig. 14

problem of mechanical engineering, and its analytical solution involves calculus, elementary physics, and differential equations. We describe it here.

We analyze a portion of the chain between points A and B, as shown in Fig. 13, where A is the lowest point of the chain and B = (x, y) is a variable point. We let T1 be the horizontal tension at A; T2 be the component of tension tangent to the chain at B; w be the weight of the chain per unit of length. Here T1, T2, w are numbers. Figure 14 exhibits these quantities.

Notice that if s is the length of the chain between two given points, then sw is the downward force of gravity on this portion of the chain; this is indicated in the figure. We use the symbol θ to denote the angle that the tangent to the chain at B makes with the horizontal.

Problem

Think of the hanging chain as the graph of a function: y (the height of the wire from the ground) is a function of x. Find an equation of $y(x)$ in terms of $y(0)$ (the height at the point A), *T1* and w.

Problem instruction

Your task is to reflect on the problem scenario and then indicate how you conceptualize the problem and how you would go about solving the problem. Here, the main purpose is NOT to solve the problem. The primary interest is in understanding what you believe in relevant to developing an acceptable solution.

Appendix 2

Solution of the problem 1

Concept map was drawn (see Fig. 1). Description of each concept and links were written as below:

Concepts

Equilibrium of forces, modeling, obtaining differential equation (DE, Pythagoras, trigonometry, derivative, tangent line, slope, initial value, solving of the differential equation (DE), slope of tangent, tangent line, slope.

   Descriptions of concepts

1. Equilibrium of forces: This is an engineering problem; one assumes that the cable has the shape of a curve $y = f(x)$ and identifies forces acting at a particular point of the cable. In this step engineering common sense is needed. To derive the DE that governs this phenomenon one has to use vector calculus and the physical notion of equilibrium of the three acting forces. Horizontal and vertical components of force acting on a body are equal. Newton's first law, tension, physics can be used instead of equilibrium of forces.
2. Modeling: Expressing a real world problem in terms of mathematics. Writing a differential equation.
3. Differential Equation (DE): After obtaining differential equation, one identifies the type of the DE and solves it. In this case by the method of separation of the variables.
4. Solving of DE: Integration is used to obtain the solution of differential equation.
5. Derivative: differential equation consists of a function and its derivatives. There will be a need to express the arc length variable by x and y, and this is achieved by differentiating the equation obtained from the equilibrium of the acting forces with respect to x and using the formula for the derivative of arc length in terms of dy/dx.
6. Tangent line—slope: The slope of the tangent line gives the derivative of the function.
7. Integral: derivative or differentiation is reverse operation of integration or integral.
8. Pythagoras—trigonometry: There will be a need to express the arc length variable by x and y. Elementary trigonometry is used for modeling.

   Description of links

1. Step 1: This is an engineering problem; one assumes that the cable has the shape of a curve $y = f(x)$ and identifies forces acting at a particular point of the cable. In this step engineering common sense is needed. To derive the DE that governs this phenomenon one has to use vector calculus and the physical notion of equilibrium of the three acting forces. Horizontal and vertical components of force acting on a body are equal. Newton's first law, tension, physics can be used instead of equilibrium of forces.
2. Step 2: Expressing a real world problem in terms of mathematics. Writing a differential equation. Obtaining DE.
3. Step 3: After obtaining differential equation, one identifies the type of the DE and solve it. In this case by the method of separation of the variables.
4. Integration is used to obtain the solution of differential equation.

5. Differential equation consists of a function and its derivatives. There will be a need to express the arc length variable by x and y, and this is achieved by differentiating the equation obtained from the equilibrium of the acting forces with respect to x and using the formula for the derivative of arc length in terms of dy/dx.

6. Derivative or differentiation is reverse operation of integration or integral.

7. Elementary trigonometry is used for modeling.

8. Slope of tangent line equals to derivative or differentiation.

Appendix 3

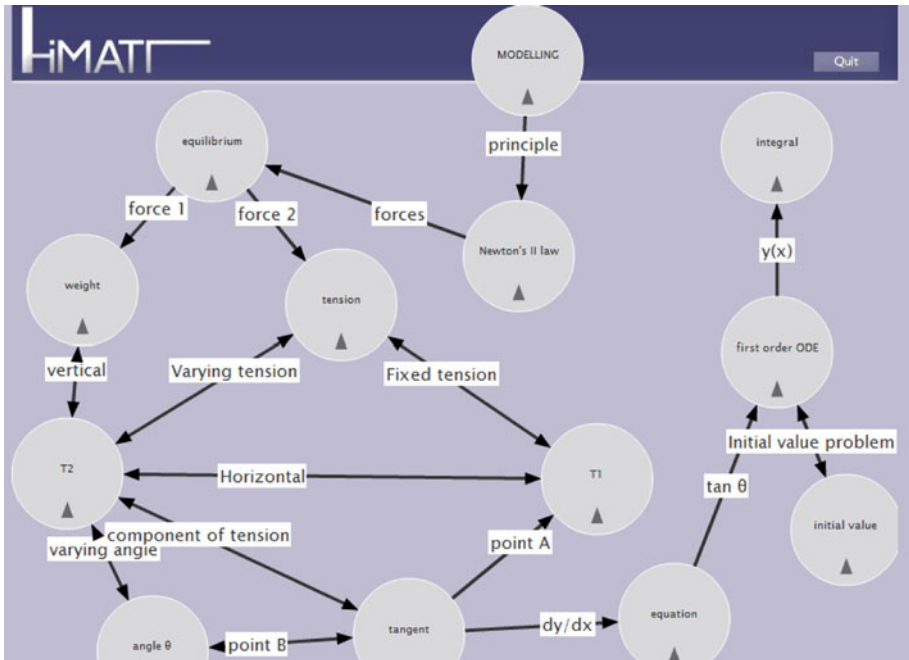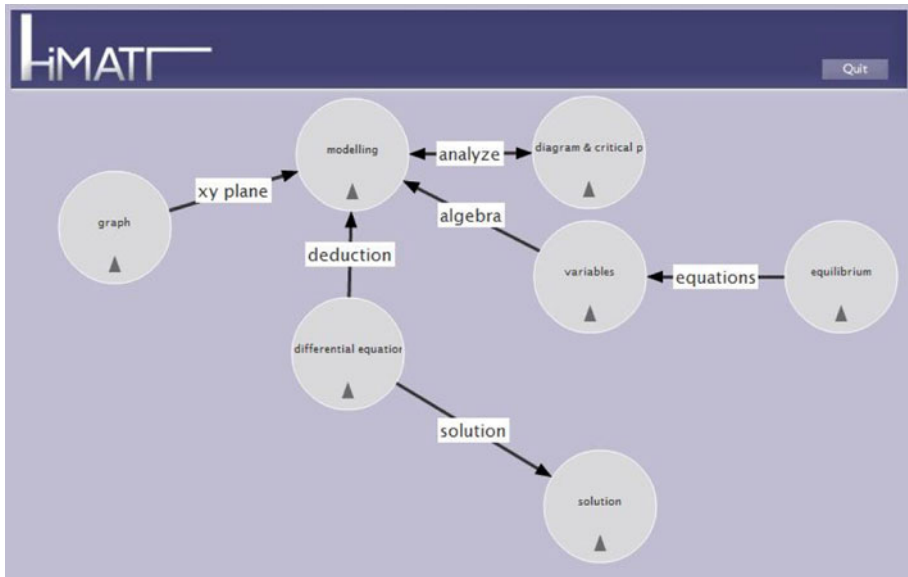Sample DEEP models from an expert participant and a novice participant

See Figs. 15 and 16.

Fig. 15 Sample for one of the experts' DEEP model

Fig. 16 Sample for one of the novices' DEEP model

References

Alexander, C. (1963). *Notes on the synthesis of form*. Cambridge, MA: Harvard University Press.

Andrews, G., & Halford, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*(2), 153–219.

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins Publishers Ltd.

Boyce, W. E., & DiPrima, R. C. (2005). *Elementary differential equations and boundary value problems* (8th ed.). Hoboken, NJ: Wiley.

Dabbagh, N. H., Jonassen, D. H., Yueh, H. P., & Samouilova, M. (2000). Assessing a problem-based learning approach to an introductory instructional design course: A case study. *Performance Improvement Quarterly, 13*(3), 60–83.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (revised ed.). Cambridge, MA: Bradford Books/MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge, MA: Cambridge University Press.

Fisher, K. M. (2000). SemNet® semantic networking. In K. M. Fisher, J. H. Wandersee, & D. Moody (Eds.), *Mapping biology knowledge* (pp. 143–166). Dordrecht, The Netherlands: Kluwer.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 303–311). Mahwah, NJ: Lawrence Erlbaum.

Gogus, A. (2009). Assessment of learning in complex domains. *The First International Congress of Educational Technology, "Trends and Issues in Educational Research"*, 1-3 May 2009, Canakkale, Turkey.

Gogus, A. (2012a). Evaluation of mental models: Using highly Interactive Model-Based Assessment Tools and Technologies (HIMATT) in mathematics domain. *Technology Instruction, Cognition and Learning (TICL), 9*(1), 31–50.

Gogus, A. (2012b). Model-based assessment of conceptual representations. *Journal of Instructional Psychology* (accepted).

Gogus, A., & Gogus, N. G. (2009, November). Evaluation of Mental Models (EMM) in mathematics domain. *Proceedings of IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2009),* 19-22, 2009, Italy.

Gogus, A., Koszalka, T. A., & Spector, J. M. (2009). Assessing conceptual representations of ill-structured problems. *Technology Instruction, Cognition and Learning, 7*(1), 1–20.

Herl, H. E., O'Neil, H. F, Jr, Chung, G. L., & Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior, 15*(3–4), 315–333.

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. (Dissertation, Freiburg: Universitäts, 2006).

Ifenthaler, D. (2009, November). *Bridging the gap between expert-novice differences: The model-based feedback approach.* IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2009), Rome, Italy.

Ifenthaler, D. (2010). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development, 58*(1), 81–97. doi:10.1007/s11423-008-9087-4.

Ifenthaler, D., Masduki, I., & Seel, N. M. (2009). The mystery of cognitive structure and how we can detect it. Tracking the development of cognitive structures over time. *Instructional Science*. doi: 10.1007/s11251-009-9097-6.

Ifenthaler, D., Pirnay-Dummer, P., & Seel, N. M. (2007). The role of cognitive learning strategies and intellectual abilities in mental model building processes. *Technology, Instruction, Cognition and Learning, 5*, 353–366.

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.

Jacobson, M. J. (2000). Problem solving about complex systems: Difference between experts and novices. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth international conference of the learning sciences* (pp. 14–21). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, T.E., O'Connor, D.L., Pirnay-Dummer, P.N., Ifenthaler, D., Spector, J.M., & Seel, N. (2006). *Comparative study of mental model research methods: relationships among ACSMM, SMD, MITOCAR & DEEP methodologies.* Paper presented at the Second International Conference on Concept Mapping, San Jose, Costa Rica.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development, 45*(1), 65–95.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development, 48*(4), 63–85.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for assessing, conveying, and acquiring structural knowledge.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kim, H. (2008). *An investigation of the effects of model-centered instruction in individual and collaborative contexts: The case of acquiring instructional design expertise.* Unpublished dissertation, Educational Psychology & Learning Systems, Florida State University, Tallahassee, FL.

Kim, A. (2010). *The effects of initial mental model construction on mental model progression, performance, intrinsic motivation, and metacognition in learning financial literacy.* Unpublished dissertation, Educational Psychology & Learning Systems, Florida State University, Tallahassee, FL.

Kim, M. K. (2012). *Assessment technologies for adaptive instruction: Diagnosis of stage-sequential learning progress in problem solving contexts.* Unpublished dissertation, Educational Psychology & Instructional Technology, University of Georgia, Athens, GA.

Lakoff, G., & Johnson, M. (1981). *Metaphors we live by*. Chicago, IL: Chicago University Press.

Lee, J. (2008). *Effects of model-centered instruction and levels of learner expertise on ill-structured problem solving.* Unpublished dissertation, Educational Psychology & Learning Systems, Florida State University, Tallahassee, FL.

Liu, X., & Hichey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. *International Journal of Science Education, 18*(8), 921–937.

Markham, K. M., Mintzes, J. J., & Jones, M. G. (1994). The concept map as a research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching, 31*(1), 91–101.

McClure, J., Sonak, B., & Suen, H. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching, 36*, 475–492.

McKeown, J. O. (2008). *Using annotated concept map assessments as predictors of performance and understanding of complex problems for teacher technology integration*. Unpublished dissertation, Educational Psychology and Learning Systems, Florida State University, Tallahassee, FL.

Milrad, M., Spector, J. M., & Davidsen, P. I. (2003). Model Facilitated Learning. In S. Naidu (Ed.), *Learning and teaching with technology: Principles and practices* (pp. 13–27). London: Kogan Page.

Payne, S. J. (1991). A descriptive study of mental models. *Behaviour and Information Technology, 10*(1), 3–21.

Perez, R. S., Johnson, J. F., & Emery, C. D. (1995). Instructional design expertise: a cognitive model of design. *Instructional Science, 23*(5–6), 321–349.

Pirnay-Dummer, P. (2007). Model inspection trace of concepts and relations. A heuristic approach to language-oriented model assessment. *AREA 2007*, Chicago, IL.

Pirnay-Dummer, P. (2010). Complete structure comparison. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 235–258). New York: Springer.

Pirnay-Dummer, P., & Ifenthaler, D. (2010). Automated knowledge visualization and assessment. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge* (pp. 77–115). New York: Springer.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2008, October). *Highly integrated model assessment technology and tools*. Paper presented at the CELDA International Conference, Freiburg, Germany.

Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development, 58*(1), 3–18.

Polk, T. A., & Seifert, C. M. (2002). *Cognitive modelling*. Cambridge, MA: MIT Press.

Ruiz-Primo, M. A., & Shavelson, R. J. (1997). Concept-map based assessment: On possible sources of sampling variability. *ERIC Document Reproduction Service No. ED422403*.

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.

Seel, N. M. (2003). Model centered learning and instruction. *Technology, Instruction, Cognition and Learning, 1*(1), 59–85.

Seel, N. M. (2004). Model-centered learning environments: Theory, instructional design, and effects. In N. M. Seel & S. Dijkstra (Eds.), *Curriculum, plans, and processes in instructional design: International perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Shin, N., Jonassen, D. H., & McGee, S. (2003). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching, 40*(1), 6–33.

Shute, V., Torreano, L., & Willis, R. (2000). DNA: Toward an automated knowledge elicitation and organization tool. In S. P. Lajoie (Ed.), *Computers as cognitive tools, Vol. two: No more walls* (pp. 309–338). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Spector, J. M. (2010, October). *Assessing progress of learning in complex domains*. Invited paper presentation at the 11th International Conference on Education Research, Seoul, South Korea.

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*, 167–183.

Spector, J. M., & Koszalka, T. A. (2004). *The DEEP methodology for assessing learning in complex domains*. Technical Report No. NSF-03-542. Syracuse, NY: Syracuse University, Instructional Design Development, and Evaluation (IDD&E).

Taleb, N. N. (2008). *The black swan: The impact of the highly improbable*. New York: Penguin.

Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open ended concept maps. *Educational Technology Research and Development, 54*(1), 61–78.

van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

Young, I. (2008). *Mental models: Aligning design strategy with human behavior*. New York: Rosenfeld Media.

Dr. Aytac Gogus has been an educational researcher and instructional designer at the Center for Individual and Academic Development, Sabanci University since September 2007. She received her M.S. from Mathematics Department of Middle East Technical University, Ankara, Turkey and her second M.S. in 2001 and her Ph.D. in 2006 from Instructional Design, Development, and Evaluation (IDD&E) Program at Syracuse University in NY, USA.